Queens iGEM & Stony Brook iGEM Collaboration

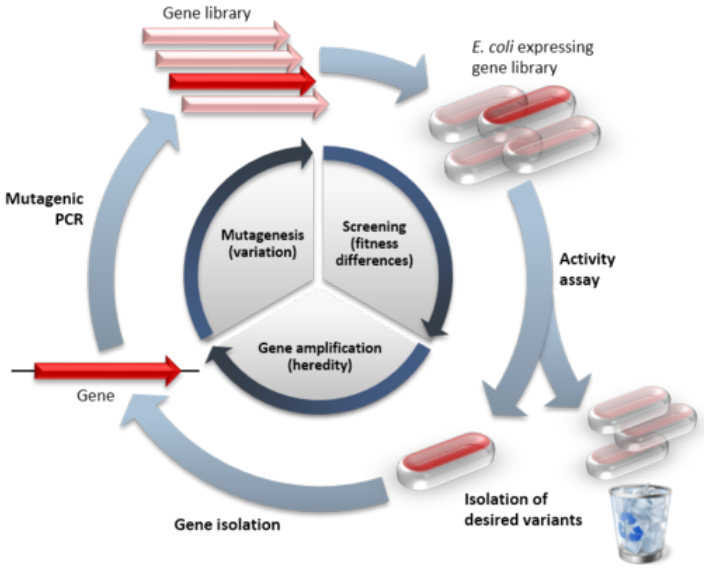
Directed Evolution

Choosing the best mutagenic method

In this pamphlet you will find information about a variety of directed evolution techniques. Both of our team's sought to perform directed evolution, and therefore we collaborated on this guide to help identify features of a variety of techniques. We hope this guide will aid your team with deciding on a directed evolution technique, just as it has helped us.

Introduction

Directed Evolution is a broad term that can refer to any method of introducing genetic mutations and selecting variants which best perform a user defined function. Directed Evolution can take place in vivo (in an organism) or in vitro (outside of the organism, usually in a solution). This can include evolving single proteins and enzymes to increase their efficiency, or even evolving entire organisms to exhibit a desired phenotype.

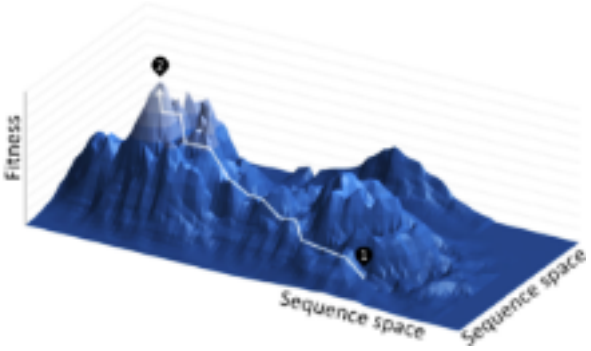


An Iterative Cycle

When evolving individual proteins, it is the molecular biologist's goal to introduce mutations into different copies of the gene. Usually this involves randomly mutating the gene involved in encoding the protein and generating multiple versions of the gene. This is called "making a library of variants". Then the variants are compared in a biochemical assay. The variants which perform best at the chosen assay are then selected and undergo further mutation until a desired outcome has been reached.

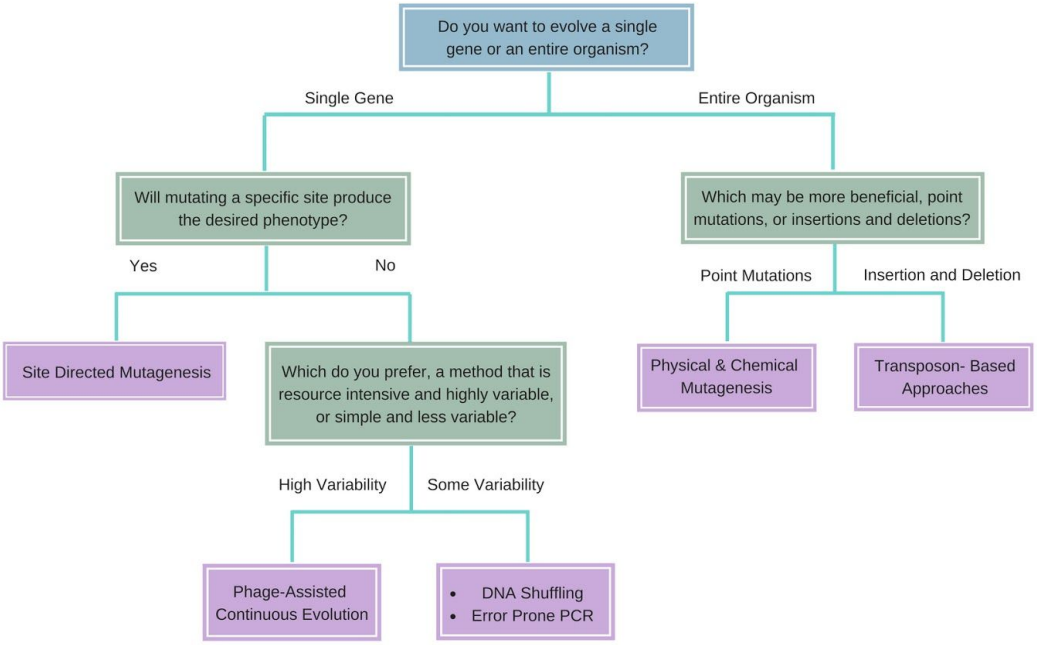
Mountain Analogy

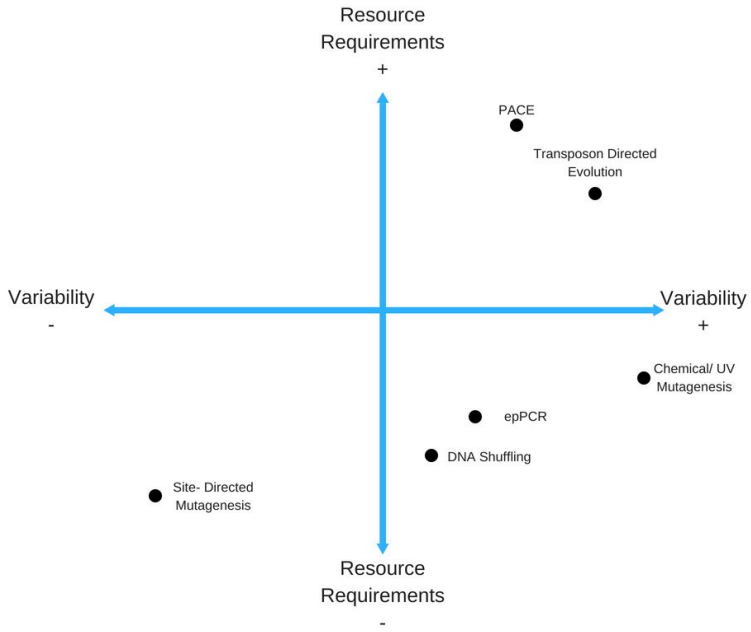
Directed evolution is often portrayed as climbing a mountain. Reaching the peak is the goal, and there exists a number of different paths that can be taken to get to the top. Instead of elevation, the molecular biologist is focused on producing the optimal function of a protein or yield of an enzyme. The x-y space represents all possible combinations of a gene, and the path is the iterations of mutations required to get to the optimal gene sequence.



Methods

There are a variety of different methods to perform directed evolution. Our flow chart and graph outline some of the most common methods, which can be found in detail below.





Gene Specific:

Error Prone PCR:

Error prone PCR is a common method for creating a combined library of variants based on a single gene. Error prone PCR is most commonly performed under conditions that reduce the accurate replication of DNA by *Taq* DNA polymerase during the synthesis of new DNA in-vitro. In a normal PCR, the replication of DNA by the polymerase is fairly accurate, however in error prone PCR the composition of the reaction buffer is altered so the polymerase makes mistakes in the base pairing during DNA synthesis.[11] This results in errors or mutations in the newly synthesized complementary DNA. The frequency of errors can be regulated by carefully controlling the buffer composition.[17] The average number of mutations per DNA template increases as a function of the number of doubling times in your reaction. In experiments that use error prone PCR for directed evolution, the frequency of mutation is normally controlled around 1-3 substitutions per kb of DNA.[11] *Taq* DNA polymerase is used in this setting, as it does not contain a proofreading function and therefore doesn't automatically fix mistakes made.

For example, the addition of $MnCl_2$, increased $MgCl_2$, and altered ratios of dNTPs will further decrease the fidelity of the Taq resulting in increased error rate.

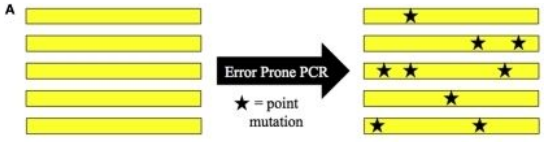


Figure 1. Creating a library of genes with varying point mutations.[18]

Advantages:

You may not want your entire sequence to be mutated. You can control the boundaries of your sequence that is mutated based on your 5' and 3' primers. You can vary the mutation rate by altering the reaction buffer composition.

DNA shuffling

DNA shuffling is a technique that can be applied alone, or in conjunction with Error Prone PCR to increase genetic variability. DNA shuffling can be used to rapidly generate beneficial

mutations through directed evolution *in vitro.* This method can create a large DNA library size through the combination of independently isolated mutations into a single offspring. DNase (an enzyme used to cut DNA) or restriction enzymes can be used to fragment the sequences. Restriction enzymes will cut in similar places along the parent DNA sequence, and the resulting fragmented sequences are joined by DNA ligase.[19] DNase is used to randomly break apart the parent genes of the desired segment of DNA into pieces of 50-100 bp.[6] The segments are then amplified using PCR without the use of primers. Through this method, DNA fragments with proficient homology of the parent genome will anneal to each other. These segments are then extended by DNA polymerase. The sequences undergoes several rounds of PCR extension, until the mutated DNA is the same size of the parent gene. The annealed sequences are amplified using primers. At this point, restriction sites used in cloning or ligation can be added to the ends of the sequence. The generated hybrid DNA sequences form a library of mutants which are tested through a phenotype assays to be tested for unique properties.

Here's how the procedure works: DNase (a deoxyribonuclease, an enzyme used to cut DNA) randomly fragments a set of parent genes into DNA segments that are 50-100 bp long.[6] Then, PCR takes place without primers to ensure that DNA fragments from different parents but with sufficiently similar sequences, anneal to each other. This then leads to extension by DNA polymerase. Multiple rounds of PCR extension are done, until the "new" DNA strand is as long as the parent gene.[19] Once the length is achieved, amplification is done using PCR with primers. The significance of the primers is that sequences encompassing restriction enzymes necessary for ligation of cloning can now be added.

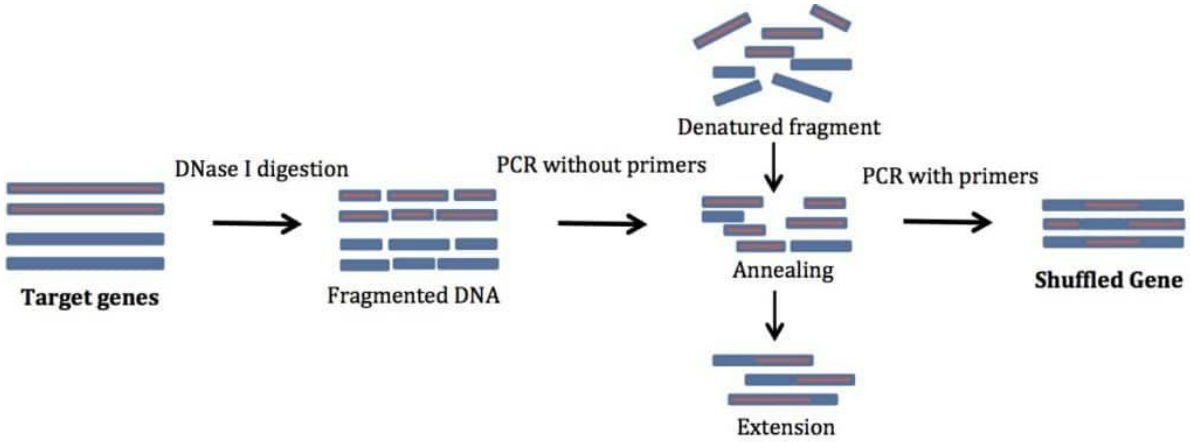


Figure 2. Overview of DNA shuffling procedure.[3]

Finally, the shuffling is done in 1 of 2 ways: using restriction enzymes or DNase 1. Restriction enzymes that cut in similar places are used to digest members of the same gene family. DNA ligase is used to join the fragments, thus producing large numbers of the hybrid that can be used for further testing.

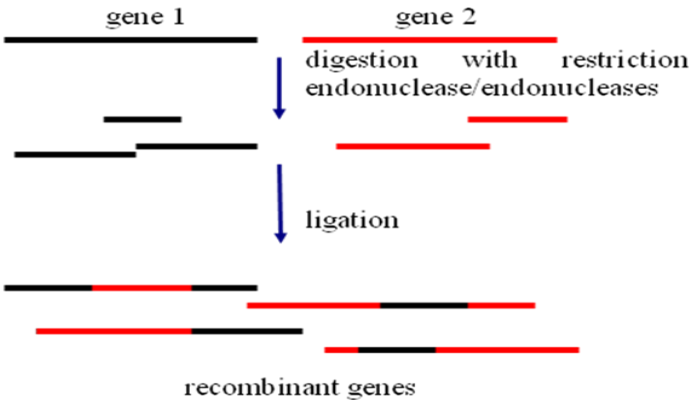


Figure 3. [16]

Similarly, with DNase 1, several members of the gene family are fragmented, and then PCR is run. During PCR, different members of the family are cross-primed. For example, homological DNA fragments will anneal to each other. The hybrids formed from this process are then used to generate a library of mutants that are tested for unique properties.

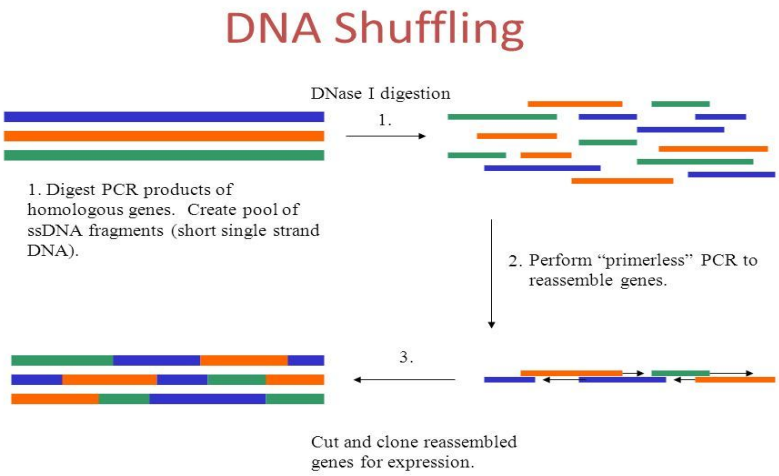


Figure 4. Process of digestion and priming in PCR.[12]

Site Directed Mutagenesis

Site Directed Mutagenesis is another type of directed evolution that is used to make specific and intentional changes to the double stranded plasmid DNA of a gene and any gene products. This process is done *in vitro* and uses custom designed primers to confer a desired mutation. Some applications of this technique are to investigate the structure and biological activity of DNA, RNA and protein molecules, and for protein engineering.[7]

So how is this process done? There are several ways: by traditional PCR, inverse PCR, primer extension, and using CRISPR technology. With traditional PCR, primers are designed to include the desired mutation, which can be a base substitution, addition, or deletion. During PCR, the mutation is incorporated into the amplicon, thereby replacing the original sequence.[15] A potential drawback of this method is that these mutations can only be incorporated into regions of sequence complementarity to the primers, and not regions between.[20]

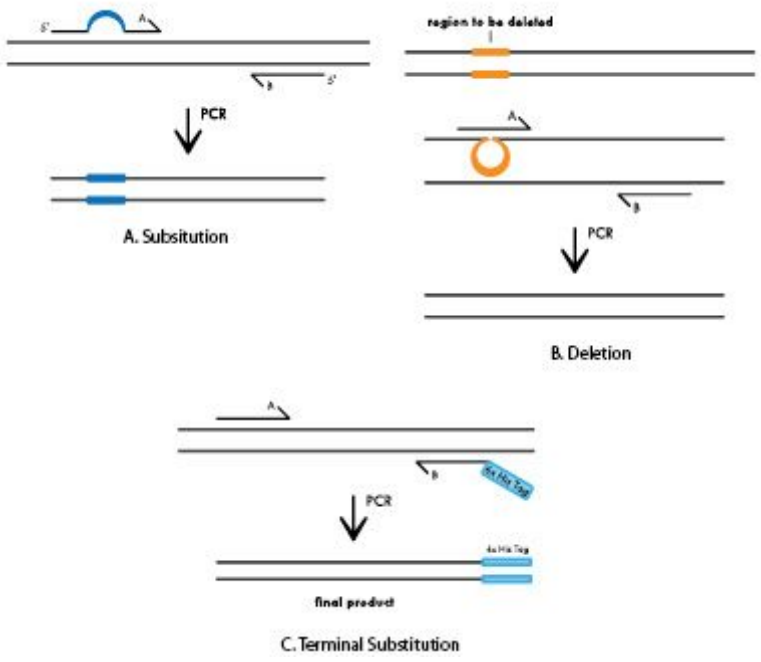


Figure 5.[15] With inverse PCR, primers oriented in the reverse direction allow for amplification of a stretch of unknown sequence. Here, primers incorporating the desired mutation are used to amplify a circular plasmid to delete, change or insert a desired sequence.

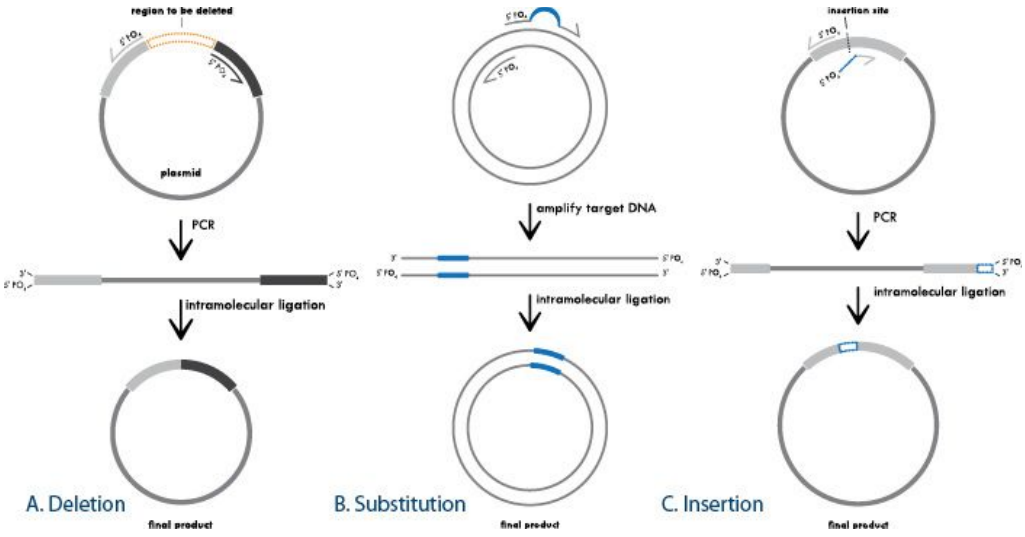


Figure 6.[15] With primer extension, mutagenic primers are incorporated in independent, nested PCR before they're combined in the final stage.

The procedure involves the use of flanking primers that are complementary to the ends of the target sequence, and 2 internal primers with complementary ends that hold the desired mutation, and eventually hybridize to the region that will change.[15]

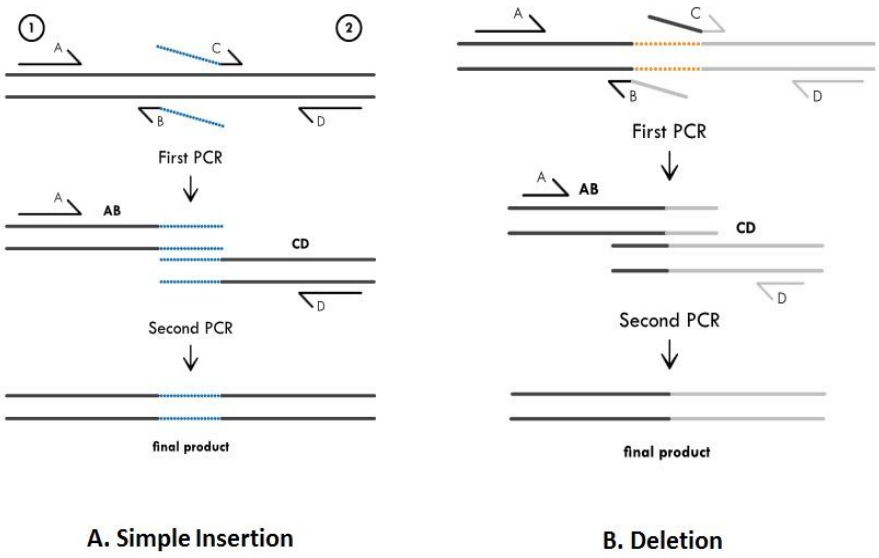


Figure 6.[15]

CRISPR gene editing is a process that is useful for introducing point mutations in DNA sets. To put things simply, the cas9 enzyme is attached to a guide (gRNA) that searches the DNA segment for a complementary section.[10] Once it's found, the cas9 enzyme cuts this section of DNA, and it is here that the desired mutation is added to the sequence.[8]

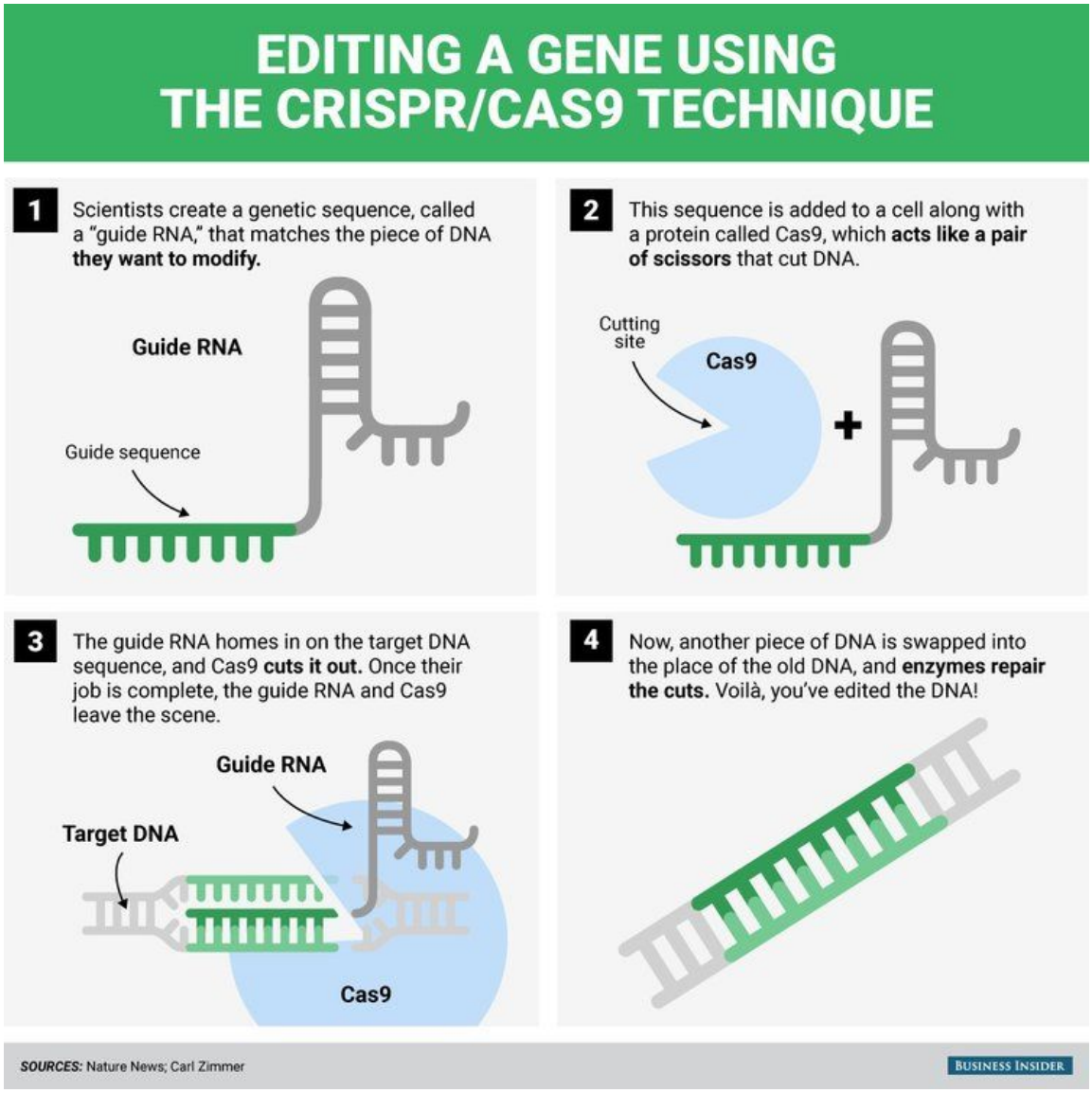


Figure 7. Editing a gene using the CRISPR/CAS 9 technique.[14]

Phage-Assisted Continuous Evolution (PACE) :

Phage-Assisted Continuous Evolution or PACE for short, takes advantage of the short lives of bacteriophages. Bacteriophages are viruses that infect bacteria and produce new viruses in ~10 minutes (23). The bacteriophage that infects E. Coli is called M13. Like all viruses, M13 cannot reproduce on its own. It needs to hijack E. Coli's cellular machinery to produce its proteins and replicate its DNA. One protein that is vital to M13 life cycle is pIII (from gene III). Without this protein, M13 production is reduced by 100,000,000 fold because the protein is essential for the pilus (the part which injects viral DNA into the E. Coli) of the virus (21).

The key innovation behind PACE is to remove gene III from M13's genome and place it in an "accessory plasmid" in the host cell (the E. Coli). In M13, gene III is replaced with a gene of interest (we'll refer to this as the "GOI") to the molecular biologist. The gene of interest (GOI) is the gene the biologist wants to evolve. At the same time, the molecular biologists engineers the E. Coli to have two interesting features:

1) Gene III (which is now in a plasmid in the E. Coli called the accessory plasmid) is linked to the desired activity of the GOI in the virus using some kind of promoter sequence that initiates transcription of gene III when the desired activity is present. Therefore, if the E. Coli detects greater activity from the GOI, more pIII protein can be produced (from gene III) and the M13 can replicate more efficiently.
2) The E. Coli has a mutagenic plasmid which is induced by simple sugar called arabinose. This mutagenic plasmid increases the mutation rate of all DNA being replicated (although we only care about the viral DNA) by producing error-prone polymerases that don't proofread the DNA . These mutagenic plasmids increase the mutation rate by a factor of 100. Because the plasmid is only induced by arabinose, the molecular biologist can prevent the E. Coli DNA from being mutated and only activate the mutation-mechanism once it enters the lagoon (more on this below).

The idea behind PACE is fairly simple once it's broken down; however, mechanically implementing it is a bit tricky. During PACE, the E. Coli should themselves never mutate and only act as a host for the virus and GOI. As such, PACE requires a constant influx (addition) of E. Coli with the right plasmids and a constant outflux of old E. Coli and inefficient viruses. If

there was not a constant influx, the E. Coli would have time to replicate and their DNA might mutate. If there was not a constant outflux, inefficient viruses containing faulty GOI's could stay in the media indefinitely. By controlling the influx and outflux using pumps, the efficiency of the viruses can be regulated. The mixture of the E. Coli and viruses is often called the lagoon.

Because of the constant, automated influx of new host cells, PACE requires little user input once operational and allows for the continuous evolution of a more optimized GOI, while cells and viruses that have severe mutations or non-functional GOI's will be flushed out in the gene flow. The result is an optimized and evolved gene that maximizes production of a protein, at a much faster and more efficient rate than before, with minimal human intervention.

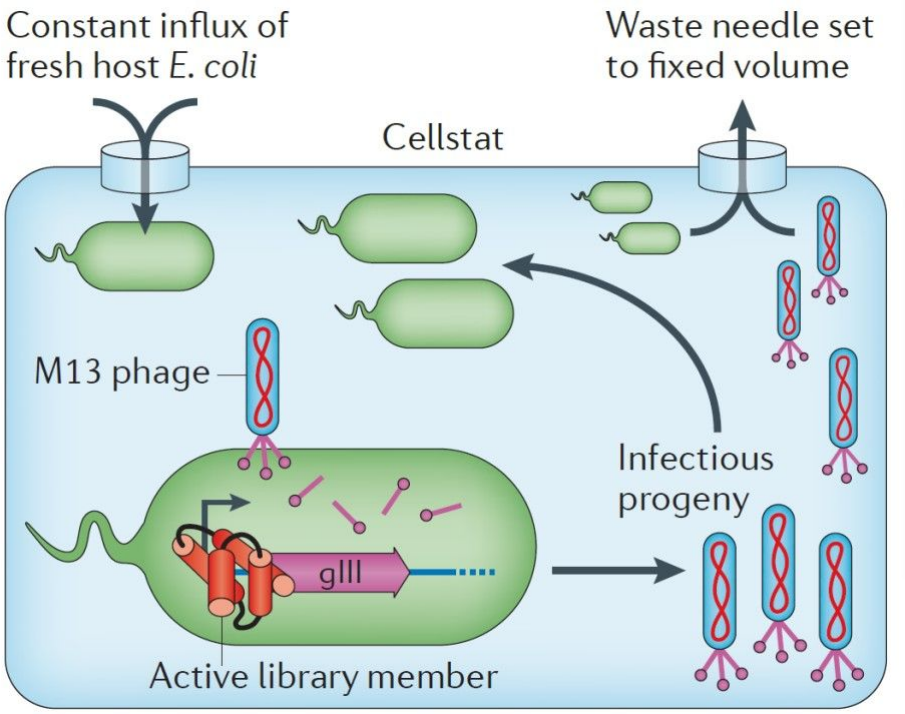


Figure 8a. Infection cycle of E. coli by M13 phage in the lagoon.[13]

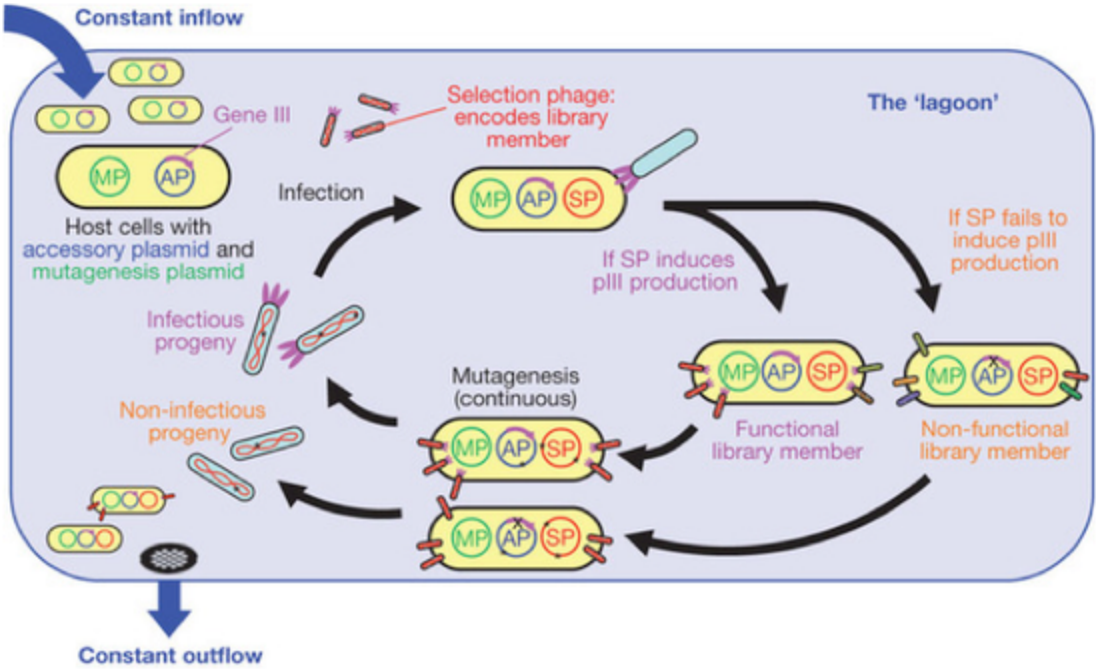


Figure 8b. More detailed Infection cycle of E. coli by M13 phage in the lagoon.[23]

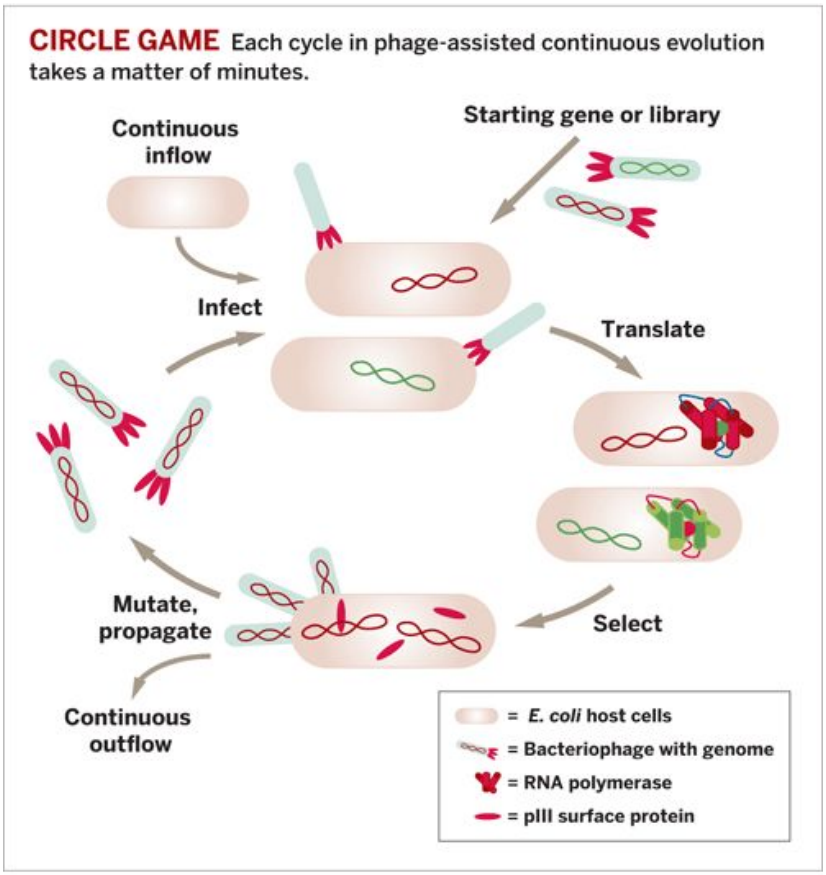


Figure 9. Overview of a single cycle in PACE.[2]

Genome Wide Evolution:

Some scientists refer to evolving the entire organism when they think of “directed evolution”. Genome wide directed evolution can refer to the rational/partial construction of an entire genome, or the practice of increasing mutation rate under a specific selection pressure. An example is this paper, which “evolved” cyanobacteria to tolerate various biofuels just by exposing them to higher concentrations in each round.[2] In a sense, humans have been doing this for thousands of years with crops and agriculture through selective breeding and artificial selection.

Random mutagenesis:

Today scientists do similar things but with modern tools through random mutagenesis techniques such as the use of physical and chemical agents to randomly damage DNA, creating mutations. Common agents for this process include methyl methanesulfonate (MMS), deaminating compounds such as nitrous acid, base analogues such as 2-aminopurine, and ultraviolet irradiation

Chemical: 2- Aminopurine, EMS

Chemical mutagens, while convenient to use compared to UV mutagenesis, also come with the risk of being carcinogens to humans. One such agent is ethyl methanesulfonate (EMS), which can act as a base-pair substitute and replace guanine, causing point mutations. Additionally, thymine, not cytosine, normally binds to EMS, which causes a transition mutation (GC to AT).[24]

2-aminopurine is an analog of adenine. However, it also can replace cytosine. As such, mutations arise when AT bonds are replaced by CG bonds (called mismatch). Because of the large presence of mismatches, the DNA repair mechanism breaks down and frameshift mutations increase in frequency.[25]

UV:

In UV-based random mutagenesis, ultraviolet light (light between 10-400 nm) is used to increase the rate of mutation in organisms. In our example, Cyanobacteria. The advantage of this method is that it can be used “in vivo” and is relatively simple, involving no reagents or equipment other than UV lights and a cover to protect the scientist.

Here’s how the mechanism works: UV light excites bonds in the DNA and causes what are called “pyrimidine dimers”. Essentially, the carbon in nucleic acids, such as Thymine, gets enough energy to form double bonds, changing the structure of DNA, and creating a “molecular lesion”.[4]

This instability can lead to mutations. Additionally, most organisms have an “SOS response” where proteins detect lesions in the DNA and attempt to eliminate errors. These SOS responses are usually driven by ultraviolet light; therefore, for optimum mutation rate, UV exposure should be limited in time.[1]

For example, with our cyanobacteria, it turns out that the optimal dosage is 10-50 Joules of UV light per m^2. Our approach is to shine a UV lamp at them for a brief interval (a few seconds to minutes), then cover them from light <520 nm (blue and violet light) to prevent the SOS response (also known as photolyase).[1]

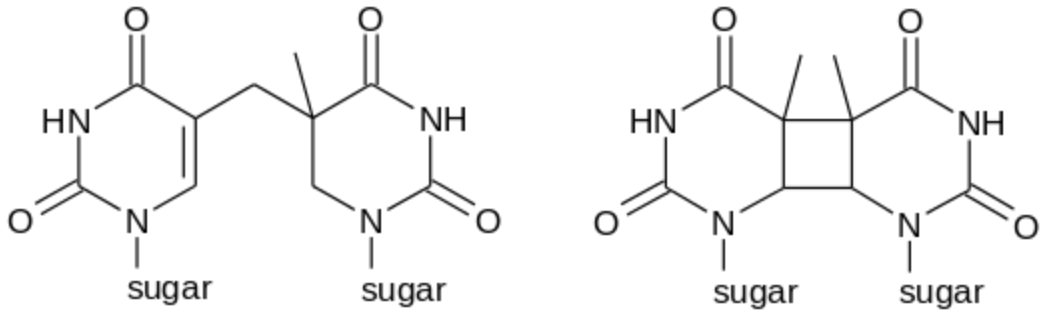


Figure 10.

Transposon-Based Approaches:

Transposon-based approaches for directed evolution (DE) require large resources; however, they generate more diverse variations than gene point mutations. MuDel, an engineered transposon, is ideal for DE, due to its low specificity and high rate of successful integration into the genome of the host.[9] Using transposon-based approaches, changes to the protein backbone can be made; hence, it allows structural changes in proteins to be studied. Here we will briefly describe three transposon-based methods: trinucleotide deletions, trinucleotide exchange, and domain insertions.The three methods share their first step: transposing with MuDel to create a MuDel insertion library (Fig.1).

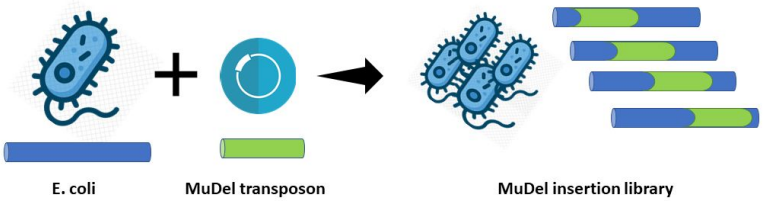


Figure 11. Production of MuDel library, produced by the transposition of MuDel into *E. coli*.

MuDel contains a recognition sequence for MlyI, a restriction endonuclease that cleaves 4bp outside of its recognition site. Trinucleotide deletions are made by following the MuDel transposition with a MlyI digestion, leading to a trinucleotide deletion. The reason the digestion results in a trinucleotide deletion, despite the enzyme cutting 4bp out of each end, is due to the Duplication that occurs at the insertion site of MuDel (Fig.2).

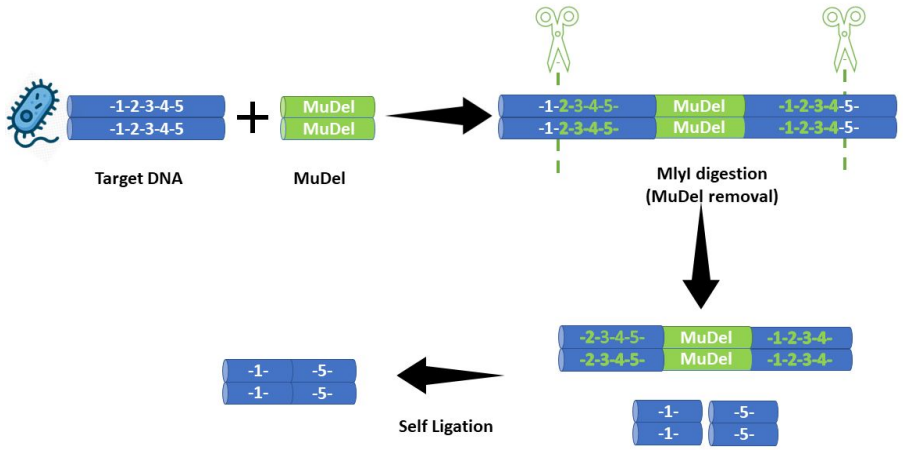


Figure 12. Trinucleotide deletion process, highlighting the duplication of the target gene upon MuDel insertion, which ultimately results in the trinucleotide deletion.

Trinucleotide deletions open the opportunity for trinucleotide exchange, which involves following the MlyI digestion with a DNA cassette (SubSeq). The DNA cassette can insert random trinucleotides, or it can be designed to insert certain trinucleotides. Like trinucleotide exchange, domain insertions involve inserting a DNA cassette; however, the DNA cassettes for domain insertion involve entire protein segments. For a more detailed approach on transposon directed evolution see *Directed Evolution Library Creation*.[1]

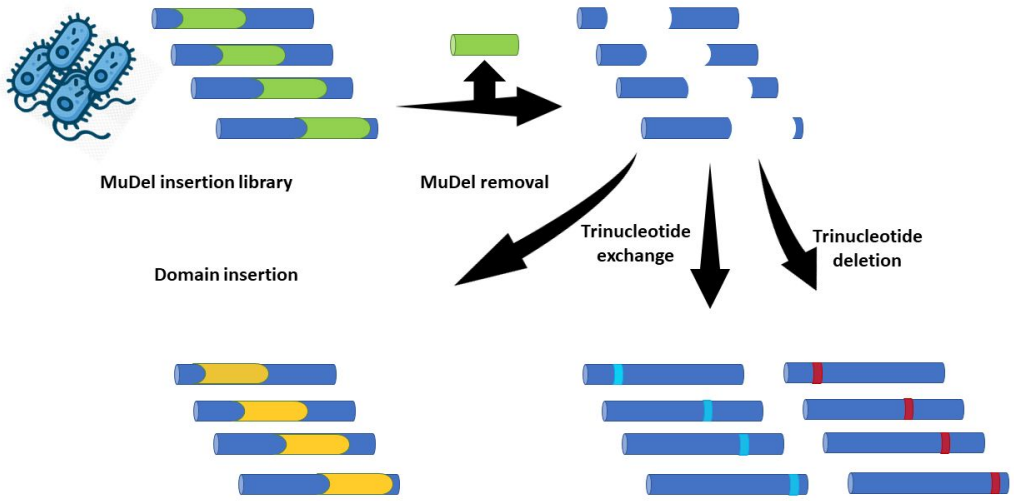


Fig.3 The three follow-up methods for the MuDel library: domain insertions, trinucleotide exchange and trinucleotide deletions.

References:

(1) Aksenov SV. Induction of the SOS Response in Ultraviolet-Irradiated Escherichia coli Analyzed by Dynamics of LexA, RecA and SulA Proteins. *J Biol Phys*. 1999;25(2-3):263-277. doi:10.1023/A:1005163310168

(2) Arnaud CH. Picking Up The Pace Of Evolution. *Chemical & Engineering News Archive*. 2011;89(16):7. doi:10.1021/cen-v089n016.p007.

(3) Basu, D. (2017). DNA Shuffling Like a Pro. Retrieved from https://bitesizebio.com/29721/dna-shuffling-like-pro/

(4) D'Orazio J, Jarrett S, Amaro-Ortiz A, Scott T. UV Radiation and the Skin. *Int J Mol Sci*. 2013;14(6):12222-12248. doi:10.3390/ijms140612222

(5) Esvelt, K. M., Carlson, J. C., & Liu, D. R. (2011). A system for the continuous directed evolution of biomolecules. *Nature*, *472*(7344), 499–503. https://doi.org/10.1038/nature09929

(6) He HH, Meyer CA, Hu SS, et al. Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification. *Nature Methods*. 2013;11:73-78. doi:10.1038/nmeth.2762

(7) Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., & Pease, L. R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, *77*(1), 51–59. https://doi.org/10.1016/0378-1119(89)90358-2

(8) Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, *157*(6), 1262–1278. https://doi.org/10.1016/j.cell.2014.05.010

(9) Jones, D. D.; Arpino, J. A. J.; Baldwin, A. J.; Edmundson, M. C. Transposon-Based Approaches for Generating Novel Molecular Diversity during Directed Evolution. In Methods in molecular biology (Clifton, N.J.) (Vol. 1179, pp. 159–172). *https://doi-org.proxy.queensu.ca/10.1007/978-1-4939-1053-3_11*

(10) Koike-Yusa H, Li Y, Tan E-P, Velasco-Herrera MDC, Yusa K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nature Biotechnology*. 2014;32(3):267-273. doi:10.1038/nbt.2800

(11) McCullum, E. O., Williams, B. A. R., Zhang, J., & Chaput, J. C. (2010). Random Mutagenesis by Error-Prone PCR. In Methods in molecular biology (Clifton, N.J.) (Vol. 634, pp. 103–109). https://doi.org/10.1007/978-1-60761-652-8_7

(12) McDowell, H. (2011). Protein Engineering and Directed Evolution. Retrieved from https://slideplayer.com/slide/7451995/
(13) Phage-Assisted Continuous Evolution - Creative Biostructure. https://www.creative-biostructure.com/phage-assisted-continuous-evolution-411.htm. Accessed August 26, 2018.
(14) Roach, D., Lewis, T. (2015). CRISPR, the gene-editing tech that's making headlines, explained in one graphic. Retrieved from https://www.businessinsider.com/crispr-gene-editing-explained-2015-12
(15) Sabel, J., Brookman-Amissah, N. (2017). Methods for site-directed mutagenesis. Retrieved from https://www.idtdna.com/pages/education/decoded/article/methods-for-site-directed-mutagenesis
(16) Tőzsér, J., Emri, T., Csősz, E. (2011). Protein Biotechnology. Retrieved from https://www.tankonyvtar.hu/hu/tartalom/tamop425/0011_1A_Proteinbiotech_en_book/ch11.html
(17) Vanhercke T, Ampe C, Tirry L, Denolf P. Reducing mutational bias in random protein libraries. *Analytical Biochemistry*. 2005;339(1):9-14. doi:10.1016/j.ab.2004.11.032
(18) Weinstein, J. R. (2013). New Methods in Engineering Adeno-Associated Virus (AAV) for Improved Gene Delivery. Retrieved from http://digitalassets.lib.berkeley.edu/etd/ucb/text/Weinstein_berkeley_0028E_13479.pdf
(19) Zhao, H., & Arnold, F. H. (1997). Optimization of DNA shuffling for high fidelity recombination. Nucleic Acids Research, 25(6), 1307–1308. https://doi.org/10.1093/nar/25.6.1307
(20) Zoller MJ (1991). New molecular biology methods for protein engineering. Curr Opin Biotechnol, 2(4): 526–531.
(21) Esvelt, Kevin M. “PACE.” *Sculpting Evolution: Phage-Assisted Continuous Evolution*, MIT Media Lab, www.sculptingevolution.org/pace.
(22) Smeal, Steven W., et al. “Simulation of the M13 Life Cycle II: Investigation of the Control Mechanisms of M13 Infection and Establishment of the Carrier State.” *Virology*, vol. 500, 2017, pp. 275–284., doi:10.1016/j.virol.2016.08.015.
(23) Esvelt, Kevin M., et al. “A System for the Continuous Directed Evolution of Biomolecules.” *Nature*, vol. 472, no. 7344, 2011, pp. 499–503., doi:10.1038/nature09929.
(24) Pitsikas, Photini, et al. “Mechanism of 2-Aminopurine-Stimulated Mutagenesis in Escherichia Coli.” *Mutation Research/Fundamental and Molecular Mechanisms of*

Mutagenesis, vol. 550, no. 1-2, Apr. 2004, pp. 25–32., doi:10.1016/j.mrfmmm.2004.01.008.

(25) Sega, Gary A. “A Review of the Genetic Effects of Ethyl Methanesulfonate.” Mutation Research/Reviews in Genetic Toxicology, vol. 134, no. 2-3, 1984, pp. 113–142., doi:10.1016/0165-1110(84)90007-1.