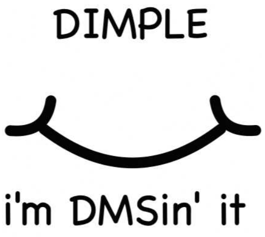
Aug 25, 2023 Version 2

DIMPLE library generation and assembly protocol v1.1 V.2



DOI
dx.doi.org/10.17504/protocols.io.rm7vzy7k8lx1/v2

Christian Macdonald[1], David Nedrud[2], Patrick Rockefeller Grimes[1], Donovan Trinidad[1], James Fraser[1], Willow Coyote-Maestas[1]

[1]University of California, San Francisco; [2]Bio-Techne

Atlas of Variant Effects ...

Christian Macdonald
University of California, San Francisco

DOI: https://dx.doi.org/10.17504/protocols.io.rm7vzy7k8lx1/v2

Protocol Citation: Christian Macdonald, David Nedrud, Patrick Rockefeller Grimes, Donovan Trinidad, James Fraser, Willow Coyote-Maestas 2023. DIMPLE library generation and assembly protocol v1.1. **protocols.io** **https://dx.doi.org/10.17504/protocols.io.rm7vzy7k8lx1/v2**Version created by **Christian Macdonald**

Protocol status: Working
We use this protocol and it's working

Created: October 24, 2022

Last Modified: August 25, 2023

Protocol Integer ID: 71768

Keywords: dimple protocol, dimple library generation, qcing mutagenic library, mutagenic library, descriptions of qc, protocol, qc

Disclaimer

Abstract

This is a protocol for generating and QCing mutagenic libraries using the DIMPLE protocol. This version is updated to include expanded descriptions of QC and to clarify certain steps.

Materials

Enzymes and cells

- PrimeSTAR GXL DNA Polymerase **Catalog** #R050A
- MegaX DH10B T1R Electrocomptent Cells **Life Technologies** **Catalog** #C6400-03
- NEB Golden Gate Assembly Mix **New England Biolabs** **Catalog #** E1601S (BsaI)

Kits (NEB kits given only as examples - any comparable kit should work)

- Monarch PCR and DNA Cleanup Kit - 50 preps **New England Biolabs** **Catalog** #T1030S
- Monarch® DNA Gel Extraction Kit **New England Biolabs** **Catalog** #T1020
- Monarch Plasmid Miniprep Kit - 250 preps **New England Biolabs** **Catalog** #T1010L

Media and chemicals

- SOC Outgrowth Medium - 100 ml **New England Biolabs** **Catalog** #B9020S
- PCR Water (nuclease free) **Catalog** #PCPW

Equipment and consumables

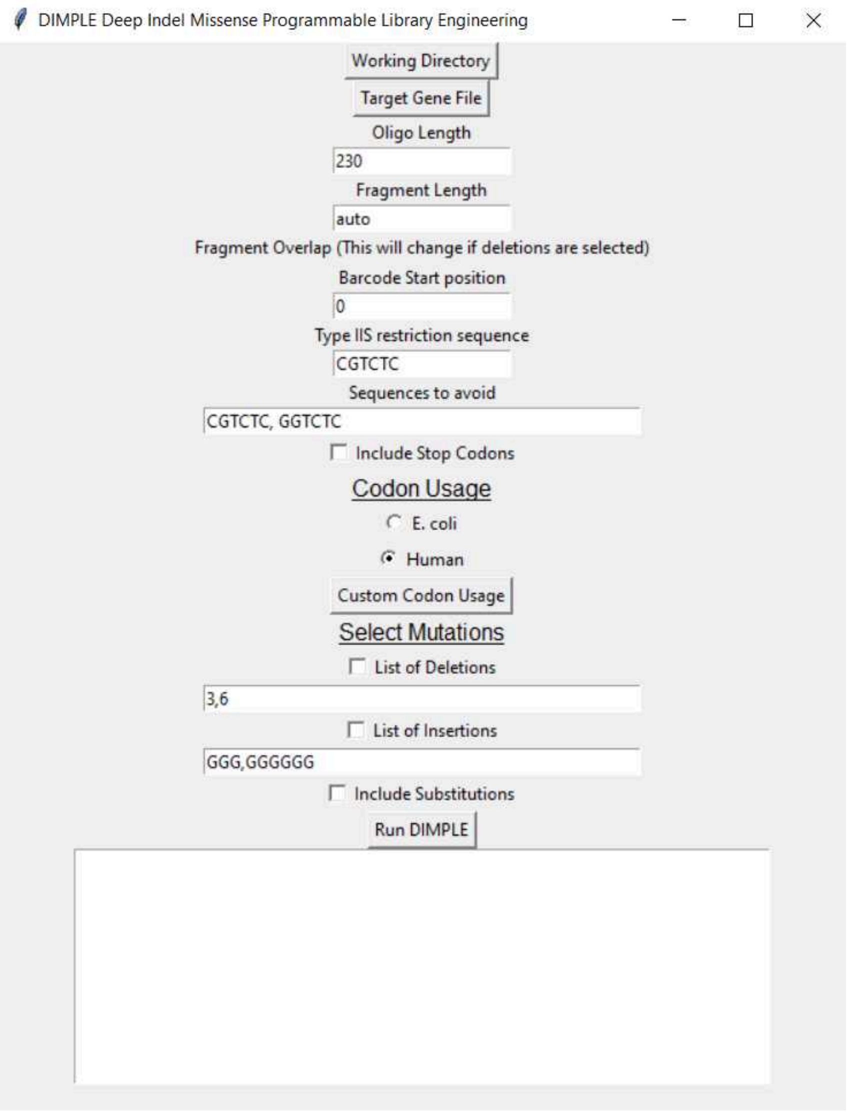
- Thermocycler
- Electroporator
- Shaker
- Horizontal electrophoresis system
- Benchtop centrifuge
- OD meter

- Electroporation cuvettes (0.1 cm)
- Cuvettes for OD measurement
- Selection agar plates
- Agarose
- Optional: large BioAssay plates for plasmid purification

Troubleshooting

Preparation

1 **Use DIMPLE to generate mutagenic oligos and primers.**



Oligo Length
230
Fragment Length
auto
Fragment Overlap (This will change if deletions are selected)
Barcode Start position
0
Type IIS restriction sequence
CGTCTC
Sequences to avoid
CGTCTC, GGTCTC
☐ Include Stop Codons
Codon Usage
○ E. coli
● Human
Select Mutations
☐ List of Deletions
3,6
☐ List of Insertions
GGG,GGGGGG
☐ Include Substitutions

Snapshot of the default DIMPLE GUI

2 **Important notes:** DIMPLE breaks a gene up into sub-library fragments and generates mutagenic insert oligo pools, where each oligo contains barcodes, Type IIS restriction cutsites, and a sub-region of the gene. Be sure to review your library generation vector and gene sequences and look for pre-existing Type IIS restriction sites. Use site-directed mutagenesis to remove unwanted off-target sites. Also note that, by default, DIMPLE

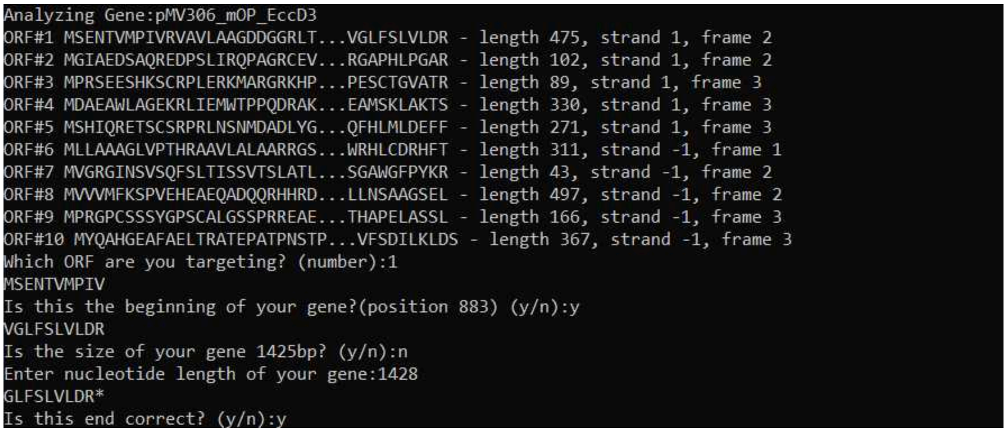
expects restriction enzymes with 6 bp recognition sites and 4 bp overhang lengths (as with BsaI and BsmBI).

3 **Input wild type gene sequence:** Input your gene sequence, including backbone, in fasta format. Place all genes to be mutagenized in the same fasta file. You can define which positions your genes start and end in the fasta header. Note that this uses 0 indexing: for example, if your first nucleotide is 884 on snapgene, use 883 in the fasta header).

- **For example:** >geneA start:35 end:250

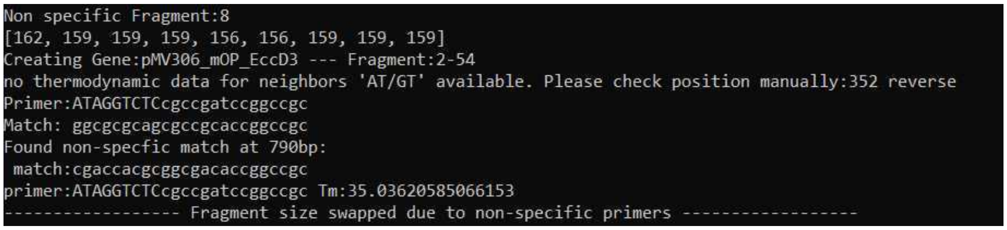
```
>TSHR start:657 end:2885
gtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggact
GACCAAGGTGTACTCCACCGACATATTCTTTATTTTGGAGATCACCGATAACCCTTATATGACAAGTATTCCTGTCAATGCTTTTCAGGGGTTGTGCAACGAAACCCTGACGCTCAAGTTGTACAATAACGGGTTCACC
ATCGACTGGCAGACCGGACCAGGCTGCAACACAGCCGGATTTTTCACCGTGTTCGCTTCTGAGCTGTCCGTGTACACTCTGACTGTCATTACATTGGAACGGTGGTATGCAATCACATTCGCCATGCGACTCGACCGGA
tcagcgaacatgatcatcgcgtcatccaggccagcgccaacgccgcggaatttctgaatctcggaagcgtactcggcgttccgctcgccgagatcgacggcgatctgttgatcaagatcctgccgcatctcgatcccac
accaataggccgaaatcggcaaaatcccttataaatcaaaagaatagaccgagatagggttgagtgttgttccagtttggaacaagagtccactattaaagaacgtggactccaacgtcaaagggcgaaaaaccgtcta
ccggctgtcagcgcaggggcgcccggttctttttgtcaagaccgacctgtccggtgccctgaatgaactgcaagacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgtt
ggggccaatacgcccgcgtttcttccttttccccaccccaccccccaagttcgggtgaaggcccagggctcgcagccaacgtcggggcggcaggccctgccatagcctcaggttactcatatatactttagattgattt
>ABCG2 start:15 end:1979
AGACACGTCTCACCATGTCTTCCAGTAATGTCGAAGTTTTTATCCCAGTGTCACAAGGAAACACCAATGGCTTCCCCGCGACAGCTTCCAATGACCTGAAGGCATTTACTGAAGGAGCTGTGTTAAGTTTTCATAACAT
CAACTCCTCCTTCTACAAAGAAACAAAAGCTGAATTACATCAACTTTCCGGGGGTGAGAAGAAGAAGATCACAGTCTTCAAGGAGATCAGCTACACCACCTCCTTCTGTCATCAACTCAGATGGGTTTCCAAGCGT
ATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCAGAGGGCACAATCCTATTCCGCGCTATCCGACAATCTCCAAGACATTAGGTGGAGTTCAGTTCGGCGTATGGCATATGTCGCTGGAAAGAACATGTGAG
AGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCG
ACGACGGCCAGTCATGCATAATCCGCACGCATCTGGAATAAGGAAGTGCCATTCCGCCTGACCT
```

*If you do not define your start & end position, the software will analyze different ORFs and ask you to define these positions.



```
Analyzing Gene:pMV306_mOP_EccD3
ORF#1 MSENTVMPIVRVAVLAAGDDGGRLT...VGLFSLVLDR - length 475, strand 1, frame 2
ORF#2 MGIAEDSAQREDPSLIRQPAGRCEV...RGAPHLPGAR - length 102, strand 1, frame 2
ORF#3 MPRSEESHKSCRPLERKMARGRKHP...PESCTGVATR - length 89, strand 1, frame 3
ORF#4 MDAEAWLAGEKRLIEMWTPPQDRAK...EAMSKLAKTS - length 330, strand 1, frame 3
ORF#5 MSHIQRETSCSRPRLNSNMDADLYG...QFHLMLDEFF - length 271, strand 1, frame 3
ORF#6 MLLAAAGLVPTHRAAVLALAARRGS...WRHLCDRHFT - length 311, strand -1, frame 1
ORF#7 MVGRGINSVSQFSLTISSVTSLATL...SGAWGFPYKR - length 43, strand -1, frame 2
ORF#8 MVVVMFKSPVEHEAEQADQQRHHRD...LLNSAAGSEL - length 497, strand -1, frame 2
ORF#9 MPRGPCSSSYGPSCALGSSPRREAE...THAPELASSL - length 166, strand -1, frame 3
ORF#10 MYQAHGEAFAELTRATEPATPNSTP...VFSDILKLDS - length 367, strand -1, frame 3
Which ORF are you targeting? (number):1
MSENTVMPIV
Is this the beginning of your gene?(position 883) (y/n):y
VGLFSLVLDR
Is the size of your gene 1425bp? (y/n):n
Enter nucleotide length of your gene:1428
GLFSLVLDR*
Is this end correct? (y/n):y
```

It is important to include the plasmid backbone so that the software can avoid making primers that nonspecifically recognize a region outside of your gene.



```
Non specific Fragment:8
[162, 159, 159, 159, 156, 156, 159, 159, 159]
Creating Gene:pMV306_mOP_EccD3 --- Fragment:2-54
no thermodynamic data for neighbors 'AT/GT' available. Please check position manually:352 reverse
Primer:ATAGGTCTCcgccgatccggccgc
Match: ggcgcgcagcgccgcaccggccgc
Found non-specfic match at 790bp:
 match:cgaccacgcggcgacaccggccgc
primer:ATAGGTCTCcgccgatccggccgc Tm:35.03620585066153
------------------ Fragment size swapped due to non-specific primers ------------------
```

4 Using the DIMPLE GUI

4.1 *Working Directory:* When you open the DIMPLE GUI, first designate your working directory. This is the folder you wish your mutagenic oligo & primer outputs to be saved in.

4.2 *Target Gene File:* Upload your gene fasta text file with the Target Gene File button.

4.3 *Oligo length:* Designate the length of your oligo. This will include the barcodes, Type IIS restriction cutsite, and a region of your target gene that will be mutagenized. So, target gene +50bp.

4.4 *Fragment length:* You can adjust the length of your fragments, we recommend leaving the fragment length set to auto.

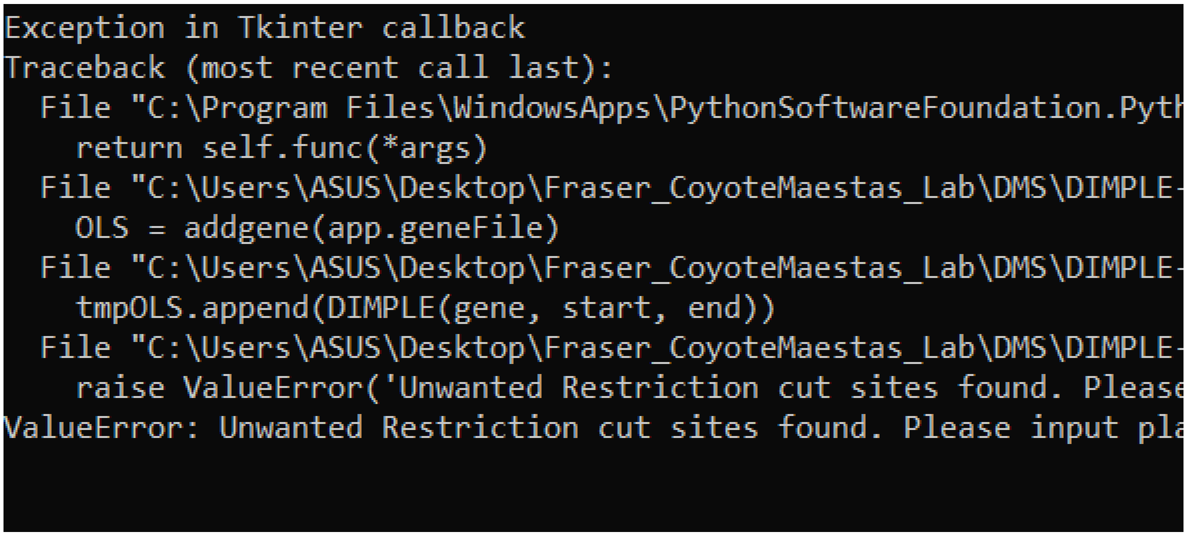
- DIMPLE will automatically break up your gene into roughly the same fragment sizes, and will determine which lengths work best to avoid matching overhangs.

4.5 *Fragment overlap:* The number of base pairs shared between sub-regions. We have seen that setting overlap to 0 base pairs leads to errors. The overlap is set to 4 base pairs by default.

4.6 *Barcode start position:* The software selects from a set pool of barcodes when designing oligos. You can define which number barcode in the list this starts from.

4.7 *Type IIS restriction sequence:* You can select which Type IIS restriction cutsite you'd like to append to your inserts and backbone. You can choose between the BsmBI (CGTCTC) or BsaI (GGTCTC) sequences.

4.8 *Sequences to avoid:* This will allow the code to design fragments without the cutsite you selected in the *Type IIS restriction sequence* section. If there is an error when running, this implies you have a cutsite in your gene or vector that needs to be removed.



```
Exception in Tkinter callback
Traceback (most recent call last):
  File "C:\Program Files\WindowsApps\PythonSoftwareFoundation.Pyth
    return self.func(*args)
  File "C:\Users\ASUS\Desktop\Fraser_CoyoteMaestas_Lab\DMS\DIMPLE-
    OLS = addgene(app.geneFile)
  File "C:\Users\ASUS\Desktop\Fraser_CoyoteMaestas_Lab\DMS\DIMPLE-
    tmpOLS.append(DIMPLE(gene, start, end))
  File "C:\Users\ASUS\Desktop\Fraser_CoyoteMaestas_Lab\DMS\DIMPLE-
    raise ValueError('Unwanted Restriction cut sites found. Please
ValueError: Unwanted Restriction cut sites found. Please input pla
```

4.9 *Codon usage:* Your oligos can be codon optimized to your organism of choice. The default options are *E. coli* or *Human*, but you can also upload a codon usage table for any organism by pressing "Custom codon usage."

4.10 *Select mutations:*

Deletions: If you'd like to generate a library with deletions across your gene, select "List of deletions". Enter how many base pairs you'd like to be deleted across your library.

- **For example,** if you'd like deletions that are one codon long, enter "3" for 3 nucleotides.
- **For example,** If you'd like to include both deletions that are one codon long AND two codons long in the same pool, enter "3,6" for both 3, and 6 nucleotide long deletions.

Insertions: If you'd like to generate a library with different amino acids inserted across your gene, select "List of Insertions".

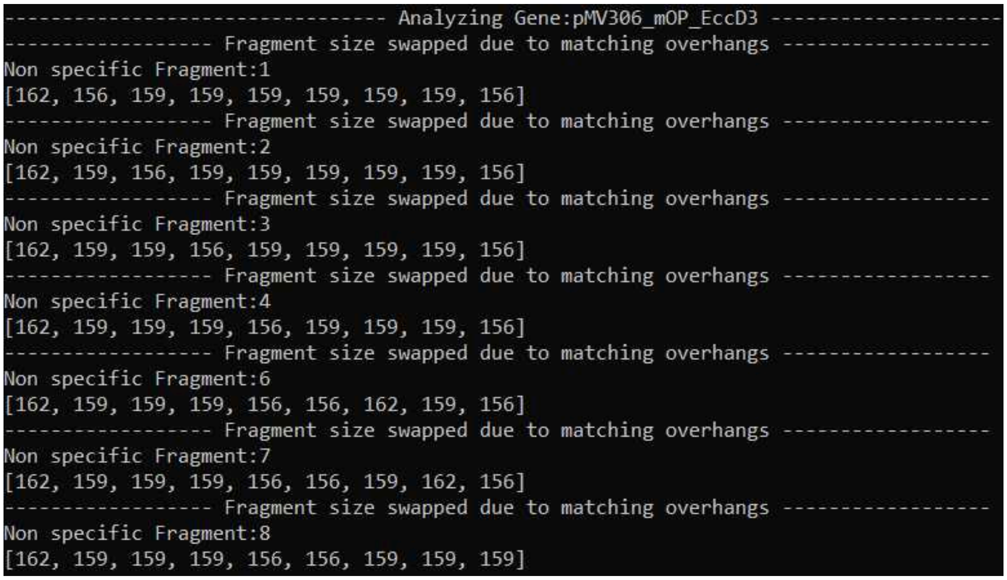
- **For example,** If you'd like to insert glycines across your gene, include "GGG" in the box.
- **For example,** If you'd like to insert glycine-serine across your gene, include "GGGTGC" in the box.
- **For example,** if you'd like to insert glycine AND glycine-serine across your gene, include "GGG,GGGTGC" in the box.

Substitutions: If you'd like to generate a library where an amino acid at each position is swapped with every other possible amino acid, select "Include Substitutions."

4.11 *Run DIMPLE:* Hit "Run DIMPLE" when you are ready to generate your pool.

5 **Examples of what the running code and outputs look like:**

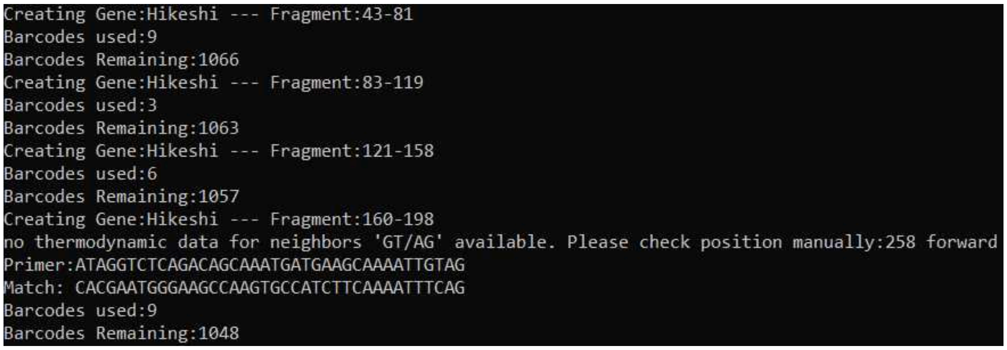
The code will first iterate the ideal fragment length sizes across each gene of interest.



```
--------------------------------- Analyzing Gene:pMV306_mOP_EccD3 --------------------
------------------ Fragment size swapped due to matching overhangs ------------------
Non specific Fragment:1
[162, 156, 159, 159, 159, 159, 159, 159, 156]
------------------ Fragment size swapped due to matching overhangs ------------------
Non specific Fragment:2
[162, 159, 156, 159, 159, 159, 159, 159, 156]
------------------ Fragment size swapped due to matching overhangs ------------------
Non specific Fragment:3
[162, 159, 159, 156, 159, 159, 159, 159, 156]
------------------ Fragment size swapped due to matching overhangs ------------------
Non specific Fragment:4
[162, 159, 159, 159, 156, 159, 159, 159, 156]
------------------ Fragment size swapped due to matching overhangs ------------------
Non specific Fragment:6
[162, 159, 159, 159, 156, 156, 162, 159, 156]
------------------ Fragment size swapped due to matching overhangs ------------------
Non specific Fragment:7
[162, 159, 159, 159, 156, 156, 159, 162, 156]
------------------ Fragment size swapped due to matching overhangs ------------------
Non specific Fragment:8
[162, 159, 159, 159, 156, 156, 159, 159, 159]
```

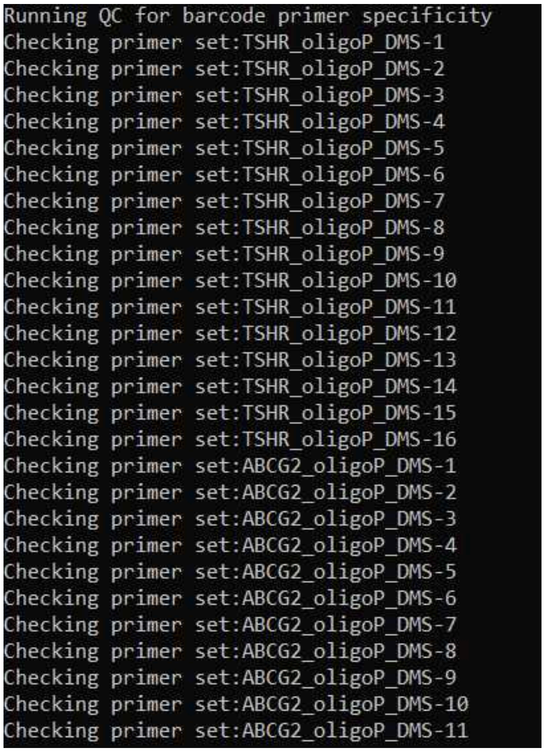
5.1 Next, barcodes are assigned to each fragment. The code will update how many barcodes it attempted to use to generate each fragment pool, and will remove these from the remaining barcode pool.

This will continue for each gene included in your fasta file.



```
Creating Gene:Hikeshi --- Fragment:43-81
Barcodes used:9
Barcodes Remaining:1066
Creating Gene:Hikeshi --- Fragment:83-119
Barcodes used:3
Barcodes Remaining:1063
Creating Gene:Hikeshi --- Fragment:121-158
Barcodes used:6
Barcodes Remaining:1057
Creating Gene:Hikeshi --- Fragment:160-198
no thermodynamic data for neighbors 'GT/AG' available. Please check position manually:258 forward
Primer:ATAGGTCTCAGACAGCAAATGATGAAGCAAAATTGTAG
Match: CACGAATGGGAAGCCAAGTGCCATCTTCAAAATTTCAG
Barcodes used:9
Barcodes Remaining:1048
```

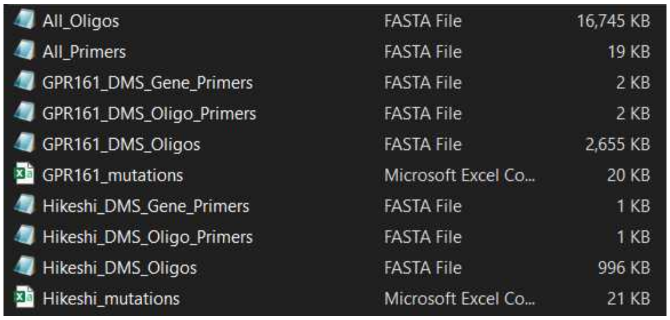
5.2 There is a final QC check for each primer set generated.



```
Running QC for barcode primer specificity
Checking primer set:TSHR_oligoP_DMS-1
Checking primer set:TSHR_oligoP_DMS-2
Checking primer set:TSHR_oligoP_DMS-3
Checking primer set:TSHR_oligoP_DMS-4
Checking primer set:TSHR_oligoP_DMS-5
Checking primer set:TSHR_oligoP_DMS-6
Checking primer set:TSHR_oligoP_DMS-7
Checking primer set:TSHR_oligoP_DMS-8
Checking primer set:TSHR_oligoP_DMS-9
Checking primer set:TSHR_oligoP_DMS-10
Checking primer set:TSHR_oligoP_DMS-11
Checking primer set:TSHR_oligoP_DMS-12
Checking primer set:TSHR_oligoP_DMS-13
Checking primer set:TSHR_oligoP_DMS-14
Checking primer set:TSHR_oligoP_DMS-15
Checking primer set:TSHR_oligoP_DMS-16
Checking primer set:ABCG2_oligoP_DMS-1
Checking primer set:ABCG2_oligoP_DMS-2
Checking primer set:ABCG2_oligoP_DMS-3
Checking primer set:ABCG2_oligoP_DMS-4
Checking primer set:ABCG2_oligoP_DMS-5
Checking primer set:ABCG2_oligoP_DMS-6
Checking primer set:ABCG2_oligoP_DMS-7
Checking primer set:ABCG2_oligoP_DMS-8
Checking primer set:ABCG2_oligoP_DMS-9
Checking primer set:ABCG2_oligoP_DMS-10
Checking primer set:ABCG2_oligoP_DMS-11
```

5.3 All outputs are saved in your working directory.

If you are using DIMPLE to mutagenize several genes, there will be separate files for the primers and mutagenic inserts for each gene, as well as a master list of **all** oligos and primers.



All_Oligos	FASTA File	16,745 KB
All_Primers	FASTA File	19 KB
GPR161_DMS_Gene_Primers	FASTA File	2 KB
GPR161_DMS_Oligo_Primers	FASTA File	2 KB
GPR161_DMS_Oligos	FASTA File	2,655 KB
GPR161_mutations	Microsoft Excel Co...	20 KB
Hikeshi_DMS_Gene_Primers	FASTA File	1 KB
Hikeshi_DMS_Oligo_Primers	FASTA File	1 KB
Hikeshi_DMS_Oligos	FASTA File	996 KB
Hikeshi_mutations	Microsoft Excel Co...	21 KB

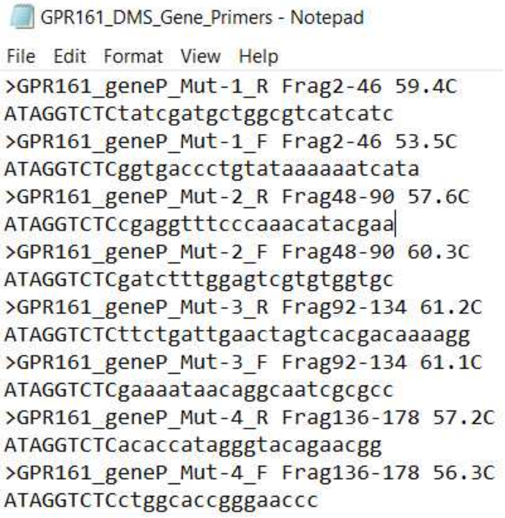
For each gene, DIMPLE generates a list of:

- **EXAMPLE_DMS_Oligos.** This is the list of sub-library oligo pools, where each oligo contains barcodes, Type IIS restriction cutsites, and a sub-region of the gene.

```
>GPR161_DMS-1_Ser2Cys
GAATAGGTGGCAAAACGCTGAGATAGATGCGGTCTCCcgatatgTGCctgaactcaagtc
tgtcctgccgaaaagaactctctaacctcacagaggaggaagggggagaaggcggcgtga
taattacgcagttcattgccatcatcgtgataaccatattcgtatgtttgggaaacctcg
tcattgtggtgaGGAGACCACTAAACGTGACGGGTGCGCGTAAAGTGAGC
>GPR161_DMS-1_Ser2Asp
GAATAGGTGGCAAAACGCTGAGATAGATGCGGTCTCCcgatatgGACctgaactcaagtc
tgtcctgccgaaaagaactctctaacctcacagaggaggaagggggagaaggcggcgtga
taattacgcagttcattgccatcatcgtgataaccatattcgtatgtttgggaaacctcg
tcattgtggtgaGGAGACCACTAAACGTGACGGGTGCGCGTAAAGTGAGC
>GPR161_DMS-1_Ser2Ser
GAATAGGTGGCAAAACGCTGAGATAGATGCGGTCTCCcgatatgTCCctgaactcaagtc
tgtcctgccgaaaagaactctctaacctcacagaggaggaagggggagaaggcggcgtga
taattacgcagttcattgccatcatcgtgataaccatattcgtatgtttgggaaacctcg
tcattgtggtgaGGAGACCACTAAACGTGACGGGTGCGCGTAAAGTGAGC
```

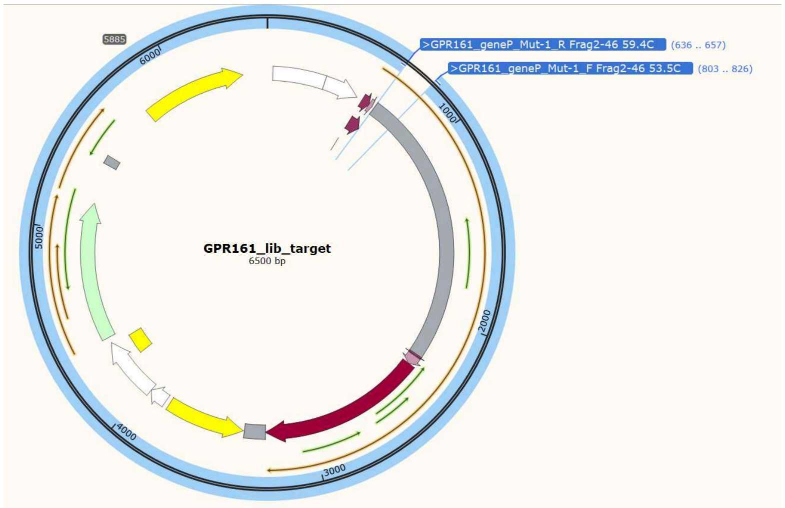
Screenshot of mutagenic inserts generated by DIMPLE

- **EXAMPLE_DMS_Gene_Primers.** These are primers used to add cutsites to and amplify the backbone each mutagenic insert will be ligated into.



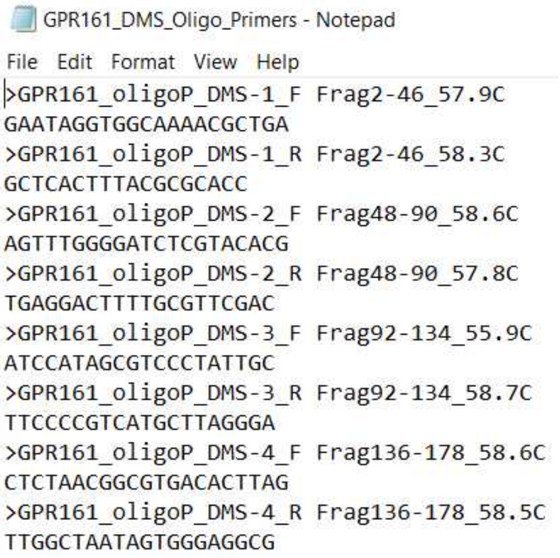
```
>GPR161_geneP_Mut-1_R Frag2-46 59.4C
ATAGGTCTCtatcgatgctggcgtcatcatc
>GPR161_geneP_Mut-1_F Frag2-46 53.5C
ATAGGTCTCggtgaccctgtataaaaaatcata
>GPR161_geneP_Mut-2_R Frag48-90 57.6C
ATAGGTCTCcgaggtttcccaaacatacgaa
>GPR161_geneP_Mut-2_F Frag48-90 60.3C
ATAGGTCTCgatctttggagtcgtgtggtgc
>GPR161_geneP_Mut-3_R Frag92-134 61.2C
ATAGGTCTCttctgattgaactagtcacgacaaaagg
>GPR161_geneP_Mut-3_F Frag92-134 61.1C
ATAGGTCTCgaaaataacaggcaatcgcgcc
>GPR161_geneP_Mut-4_R Frag136-178 57.2C
ATAGGTCTCacaccatagggtacagaacgg
>GPR161_geneP_Mut-4_F Frag136-178 56.3C
ATAGGTCTCctggcaccgggaaccc
```

Screenshot of backbone primers generated by DIMPLE. Each primer name also lists its melting temperature.



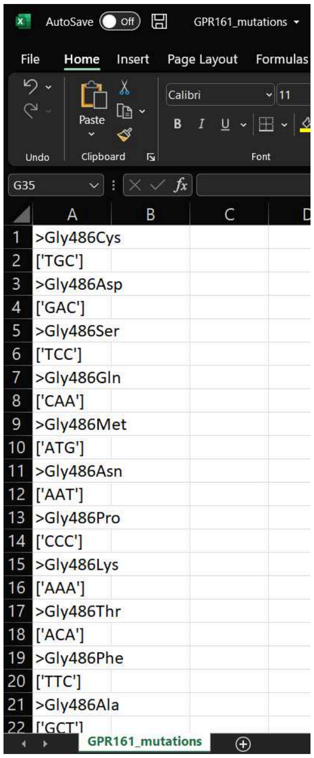
Each primer pair linearizes the backbone

- **EXAMPLE_DMS_Oligo_Primers.** These are primers used to amplify the mutagenic inserts from your oligo pool.



```
>GPR161_oligoP_DMS-1_F Frag2-46_57.9C
GAATAGGTGGCAAAACGCTGA
>GPR161_oligoP_DMS-1_R Frag2-46_58.3C
GCTCACTTTACGCGCACC
>GPR161_oligoP_DMS-2_F Frag48-90_58.6C
AGTTTGGGGATCTCGTACACG
>GPR161_oligoP_DMS-2_R Frag48-90_57.8C
TGAGGACTTTTGCGTTCGAC
>GPR161_oligoP_DMS-3_F Frag92-134_55.9C
ATCCATAGCGTCCCTATTGC
>GPR161_oligoP_DMS-3_R Frag92-134_58.7C
TTCCCCGTCATGCTTAGGGA
>GPR161_oligoP_DMS-4_F Frag136-178_58.6C
CTCTAACGGCGTGACACTTAG
>GPR161_oligoP_DMS-4_R Frag136-178_58.5C
TTGGCTAATAGTGGGAGGCG
```

- **EXAMPLE_mutations.** This is the list of mutations that will be present in your final library.



	A	B	C
1	>Gly486Cys		
2	['TGC']		
3	>Gly486Asp		
4	['GAC']		
5	>Gly486Ser		
6	['TCC']		
7	>Gly486Gln		
8	['CAA']		
9	>Gly486Met		
10	['ATG']		
11	>Gly486Asn		
12	['AAT']		
13	>Gly486Pro		
14	['CCC']		
15	>Gly486Lys		
16	['AAA']		
17	>Gly486Thr		
18	['ACA']		
19	>Gly486Phe		
20	['TTC']		
21	>Gly486Ala		
22	['GCT']		

Library assembly

3d

6 **Prepare oligo pool stock**

Follow any recommendations for oligo pool resuspension: typically, this results in a 10 ng/µL solution, which the following steps assume. Lower or higher concentrations may require alterations.

7 **PCR amplification of oligos, backbone**

Thaw all components beforehand and follow general directions provided by manufacturer.

Amplification of mutagenic inserts and backbones should ideally be performed in parallel, assuming one has two thermocyclers. Alternatively, the backbone amplification can be performed first, then the the insert amplification can be done during the gel purification of the backbone.

8 **PCR amplification of backbone.**

Prepare a master mix with PrimeSTAR GXL polymerase:

	A	B	C	D	E
	Component	**Total amount in master mix (µL)**	Amount/reaction (µL)	Comment	Number of regions
	dNTP	**20**	4		5
	5X buffer	**50**	10		
	Template (vector)	**5**	1	~10 ng	
	Enzyme	**5**	1		
	Nuclease-free H2O	**160**	32		
	Primers		2	Fwd & reverse combined at 20 µM (each)	
	Total	**240**	50		

Mix by vortexing and spin down.

8.1 Transfer 48 µL of master mix into separate tubes for each reaction.

8.2 Add reaction-specific primers to each tube: 2 µL (paired at 20 µM, each) . Mix and spin down.

8.3 Place on thermocycler and amplify:

	A	B	C
	Step	T (° C)	Time (sec)
	1	98	10
	2	55	15
	3	68	60 per kb

	A	B	C
	4	Repeat 1-3 15-24 times	See note
	5	10	Hold

Note: the number of cycles should be optimized to minimize PCR cycles. Increasing cycles introduces the possibility of error. An initial comparison of 16, 18, and 22 total cycles can be used to find a minimum number of cycles which yields sufficient DNA for the assembly reaction (see **step 11** below) and no more. Increasing the template concentration (to 50 or 100 ng) could be preferable to increasing PCR cycles if large numbers are required for amplification.

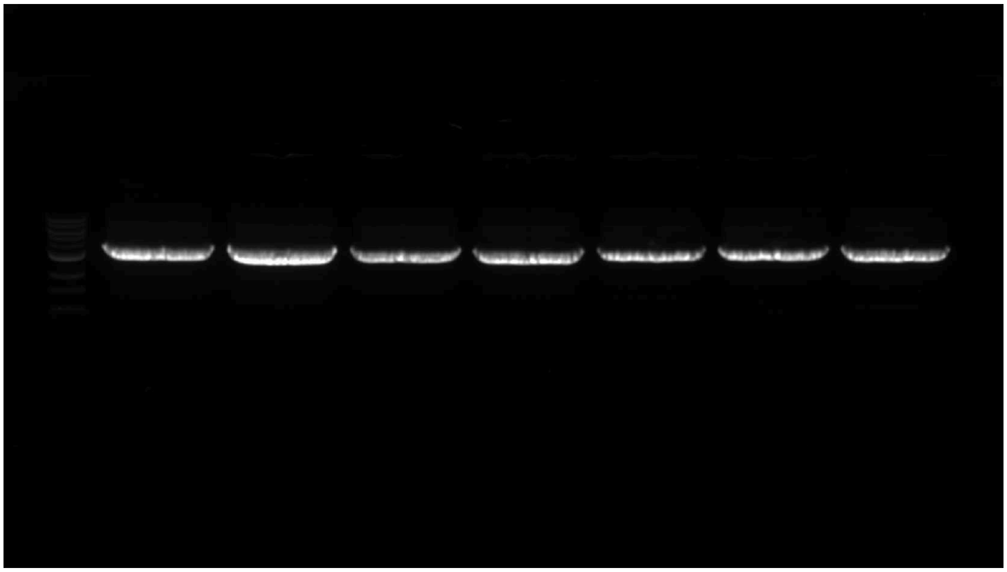
8.4 **Purify amplified product by gel extraction.**

Prepare an agarose gel for gel purification (0.5% - 0.75%). Ideally, it should be of sufficient size to load all samples.

8.5 Load and run backbone products.

8.6 Use a scalpel or razor blade to (carefully!) cut out each product. Using a gel extraction kit, purify the product and elute in 10 µL elution buffer .

8.7 **Important QC step** - evaluate the quality of the amplification with the gel. Each lane should have a single, crisp band at the expected size. Multiple bands, or no bands, might indicate improper annealing and may require Tm or extension time optimization. Blurry and smeary bands might indicate problems with the electrophoresis or amplification conditions. It is essential to get specific, clean amplification of the backbone, and libraries should not be generated with messy reactions.



Example visualization of expected backbone amplification reactions on SYBR Safe-stained agarose gel. Each lane has a single, crisp band at the expected position that is well-separated from other lanes.

9 **PCR amplification of mutagenic inserts.**

Prepare a master mix with PrimeSTAR GXL polymerase:

	A	B	C	D	E
	Component	**Total amount in master mix (µL)**	Amount/reaction (µL)	Comment	Regions
	5X buffer	**50**	10		5
	dNTP	**20**	4		
	Oligo pool	**5**	1	10 ng pool per reaction, assuming a stock at 10 ng/µL	
	Enzyme (PrimeSTAR GXL)	**5**	1		
	Primers		2	Fwd & rev combined, 20 µM (each)	
	Nuclease-free H2O	**160**	32		
	Total	**240**	50		

	A	B	C	D	E

Mix by vortexing.

9.1 Transfer 48 µL insert master mix into separate tubes for each reaction.

9.2 Add reaction-specific primers to each tube: 2 µL (paired at 20 µM, each) . Mix and spin down.

9.3 Place on thermocycler and amplify:

	A	B	C
	Step	T (° C)	Time (sec)
	1	98	10
	2	55	15
	3	68	60
	4	Repeat 1-3 15-24 times	See note
	5	10	Hold

Note: the number of cycles should be optimized to minimize PCR cycles. Increasing cycles introduces both PCR bias and error. An initial comparison of 16, 18, and 22 total cycles can be used to find a minimum number of cycles which yields sufficient DNA for assembly step and no more. For the example libraries here, this is ~100 ng per oligo pool, and

9.4 Use a PCR cleanup kit to purify each product. Elute in 10 µL elution buffer .

10 **Important QC step** - run a gel with each PCR product and visualize it. Ideally, each reaction should ideally give a single strong band at the expected size. Failure to do so may require changing PCR conditions.

11 **Golden gate assembly**

Prepare a master mix for the assembly. Use the table below to calculate: adjust number of regions and amount of necessary backbone and insert for desired amounts per reaction. We have found 3 µL backbone and 1 µL oligo (insert) are usually good.

	A	B	C	D	E
	Component	**Total amount (µL)**	Amount/reaction (µL)	Notes	Number of regions
	10X buffer	**20**	4		**5**
	Enzyme	**15**	3		
	Backbone		3	300 ng	
	Insert		1	2:1 molar ratio: for example libraries here, 100-200 ng	
	Nuclease-free H2O	**145**	29		
	Total	**200**	40		

Vortex to mix.

11.1 Transfer appropriate master mix into separate tubes for each reaction.

Example: if each reaction has 3 µL backbone and 1 µL insert, then transfer **36 µL** to each tube.

11.2 To each tube, add the appropriate vector/insert pair:

For this example:

3 µL backbone (purified)

1 µL insert (purified)

Mix well and spin down.

11.3 Place on thermocycler and run the following program:

	A	B	C
	Step	T (° C)	Time (min)
	1	37	5

	A	B	C
	2	16	10
	3	Repeat 1-2 34 times (35 total cycles)	
	4	60	5
	5	10	Hold

Important: This assumes the assembly is performed with BsaI. If using a different enzyme (such as BsmBI), check the recommended temperatures from the vendor.

11.4 Use a PCR cleanup kit to purify each product. **Elute in** 10 µL nuclease-free H2O .

Important: using elution buffer or TE may cause electroporation to fail!

11.5 **Important QC step**: before electroporating libraries, transform 1 µL of each library into chemically competent cells and plate. If no or very few colonies are observed the next day, this could suggest a failure during assembly. Further check the assembly by running a small portion on a gel.

Transformation and recovery of sublibraries

2d

12 Using a high-efficiency electrocompetent strain of *E. coli*, such as MegaX DH10B, prepare a transformation for each assembly: 3 µL assembly product with 20 µL electrocompetent cells in a 0.1 cm cuvette. Follow the specific instructions for voltage and outgrowth media corresponding to your cells and electroporator. Outgrow cells for 1 hour before proceeding.

13 **Essential QC step:** Count total number of transformants to ensure adequate coverage.

Prepare 5 10-fold serial dilutions: for each reaction, add 900 µL H_2O to six 1.5 mL microcentrifuge tubes. Add 100 µL of the outgrowth mixture to the first, then mix well and add 100 µL of this to the second tube. Repeat for the rest of the tubes.

13.1 Plate 100 µL of the last three dilutions (1:1000, 1:10000, 1:100000) on LB-agar with appropriate antibiotic. Grow overnight and count colonies the next day. Calculate the total number of transformants and determine the variant coverage per reaction. At least 50-fold coverage is **essential**.

13.2 Example calculations

For two oligo subpool reactions, the following plate counts were observed.

	A	B	C	D
	Reaction	1:1000 colonies	1:10000 colonies	1:100000 colonies
	Subpool 1	36	5	0
	Subpool 2	Many	18	2

The total number of implied transformants in the starting outgrowth media was calculated with the following formula:

$$N = C_n \times 10^n \times 100$$

where C_n is the total number of colonies on a particular plate, 10^n accounts for the serial dilutions, and 100 accounts for the fact that only 100 µL out of the 1 mL outgrowth media is being sampled, and only 1/10 of each dilution is being plated. If a large number of colonies are seen on a lower dilution, we typically ignore that one, but a lack of colonies is counted as a zero. Here, we would calculate the following implied transformants:

	A	B	C	D	E
	Reaction	Transformants	Transformants	Transformants	Average
	Subpool 1	3.6e6	5e6	0	1e6
	Subpool 2		2e7	2e7	2e7

Next, the total transformants per variants in each pool is calculated, using the average number of transformants.

	A	B	C
	Reaction	Variants in pool	Coverage
	Subpool 1	10000	300
	Subpool 2	1300	14000

Both these subpools would be good to proceed with.

13.3 **Important QC step:** pick single colonies from the QC plates and use clonal sequencing to check for mutations (e.g., Sanger sequencing or whole-plasmid sequencing). We typically pick 5 colonies per subpool and expect the majority to contain a desired mutation. Too much WT could indicate problems during the backbone amplification, and too many off-target mutations could suggest a failure to amplify or purify properly. In both cases, these would be extremely deleterious to the quality of the final library.

14 Add the remaining outgrowth culture to 30 mL LB supplemented with appropriate antibiotic for further outgrowth in either a Falcon tube or glass culture tube. Shake until the culture reaches OD 0.6-0.7.

Note: it is important to harvest the cells in the exponential phase. This takes 4-6 hours in our experience, but may vary (including between samples in the same day).

15 Harvest DNA from cultures using a miniprep kit, splitting each reaction across 3 columns. Elute each column in 40 µL elution buffer and combine elutions corresponding to a single sublibrary.

Sublibrary pooling and subcloning

2d

16 Measure each sublibrary concentration with a fluorometric method, such as Qubit. With the lowest concentration as the target concentration, prepare a dilution of each sublibrary in H_2O. Mix 2 µL of this dilution from each sublibrary to create an equimolar library pool.

17 Subclone library to transfer from library vector to selection vector. Depending on construct design, this may require digestion/ligation or a Golden Gate reaction.

18 Use PCR cleanup kit to purify reaction, and elute in 10 µL H_2O.

19 Using a high-efficiency electrocompetent strain of *E. coli*, such as MegaX DH10B, transform 3 µL subcloned product with 50 µL electrocompetent cells cells in a 0.1 cm cuvette. Follow the specific instructions for voltage and outgrowth media corresponding to your cells and electroporator. Outgrow cells for 1 hour before proceeding.

20 Repeat transformation and isolation following steps for subpool cloning and quantification as above, except perform 6 serial dilutions and plate samples at 1:10000, 1:100000, and 1:1000000 dilutions and use variant counts for the total library. Again, at

least 200-fold coverage is **essential** for maintaining unbiased libraries. Pick colonies and submit for clonal sequencing as well: we typically submit ~10 samples for our final pooled library.