UNIVERSITY OF CALIFORNIA SAN DIEGO

Optimizing Molecular Cloning Strategies for Naturally Diverse HIV-1 *env*

A Thesis submitted in partial satisfaction of the requirements for the degree Master of Science

in

Biology

by

Brian Andrew Hanst

Committee in charge:

Professor Benjamin Murrell, Chair
Professor Matthew Daugherty, Co-Chair
Professor Michael David

2018

The Thesis of Brian Andrew Hanst is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

Co-Chair

_____

Chair

University of California San Diego

2018

DEDICATION

This body of work is for my parents. I could not do anything without their support. Thank you to my mother, father, abuela, and my brother Kevin. I am greatly indebted to Ben Murrell and Thomas Vollbrecht for their patience, understanding, and mentorship in this long process.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT OF THE THESIS

Optimizing Molecular Cloning Strategies for Naturally Diverse HIV-1 *env*

by

Brian Andrew Hanst

Master of Science in Biology

University of California San Diego, 2018

Professor Ben Murrell, Chair
Professor Matthew Daugherty, Co-Chair

Broadly Neutralizing Antibodies (bNAbs) can bind and neutralize multiple strains of HIV-1 at evolutionarily and structurally conserved sites of the Env surface protein. bNAbs develop in some HIV-infected individuals during infection, and they have been

shown to prevent infection upon SHIV challenge in the Macaque model. This makes bNAb elicitation a primary objective in the effort to create a vaccine for HIV-1.

The origin of bNAbs during natural infection is a result of co-evolution between the HIV-1 Env population and the broad antibody lineage, and studies of longitudinal donor samples can help us understand how these bNAbs arise. These studies are typically quite time consuming, and, in this thesis, I will explore preliminary investigations into high-throughput strategies for characterizing the co-evolution of HIV *env* and antibody lineages.

By creating libraries of diverse *env* sequences, a range of high-throughput experiments become available. This thesis describes the efforts to create libraries of diverse *env* for different projects: we would aim to conduct one project that uses the libraries for transduction of HEK293T cells for cell sorting to determine sequence similarities among *envs* that bind the bNAb, as well as enriched features. The other project uses libraries to grow virus and observe the ability of the virus to escape certain bNAbs. Both projects require cloning Env libraries from primary samples into large plasmid vectors, and this cloning is the focus of my thesis.

CHAPTER 1: HIV-1 *env* Molecular Cloning and Broadly Neutralizing Antibodies

1.1: INTRODUCTION

Human Immunodeficiency Virus-1 (HIV-1) infection affects 36.9 million people in the world as of 2017 (World Health Organization, July 2018 Report). HIV-1 is a retrovirus, consisting of a positive-sense RNA genome that reverse transcribes and integrates its genetic material into the genome of the cell it infects. Antiretroviral therapy is an effective method of treatment, but it is not a cure for HIV. A major obstacle to eradication of the virus from a patient is the latent reservoir, meaning that the integrated virus remains transcriptionally silent in the genome, so viral genes are not expressed (Siliciano and Greene 2011). This means the latent reservoir is not affected by antiretroviral treatments. The latent reservoir makes the realization of a cure difficult, but it does not exclude the idea of a vaccine to prevent infection. Vaccination development is a popular effort because of recent advances in knowledge of humoral immunity against HIV-1. Donors can produce neutralizing antibodies against the Env surface protein, and some of these antibodies neutralize multiple subtypes of HIV-1. These are termed "broadly neutralizing antibodies" (bNAbs) (Burton, *et al.*, 2012). The key to a possible vaccine for HIV-1 rests with targeting the viral envelope (Env) glycoprotein.

Env is the major viral surface-exposed protein, and thus it is the primary target for the humoral immune system. Therefore, Env is the focus of antibody-based vaccine efforts. Env mediates entry through direct interaction with the CD4 receptor on CD4+ T-cells and a coreceptor, typically membrane-spanning chemokine receptor CXCR4 or CCR5, depending on the white blood cell type (Moir, Chun, and Fauci, 2011; Gift, *et al*., 2017). Env forms a trimer on the surface of the virus, and each of the Env protomers

consists of two subunits: The surface subunit, gp120, protrudes from the lipid envelope and mediates binding to the primary receptor and co-receptor of the target cell. The transmembrane subunit, gp41, mediates fusion of the viral and host cell membranes through an N-terminal fusion peptide (Wyatt, 1998). Upon gp120 binding to primary and coreceptor, gp120 undergoes conformational changes, triggering gp41 to insert the fusion peptide into the target cell membrane. The gp41 subunit then bridges the gap between the membranes and brings them into physical contact leading to membrane fusion and subsequent viral entry (Wyatt, 1998). Env binding and membrane fusion with CD4+ T-cells is the start of the viral replication cycle. If binding of Env to either the CD4 receptor or the CXCR4/CCR5 coreceptor is disrupted, infection will not proceed (Miyauchi, *et al.*, 2009).

The structure of Env provides multiple targets that antibodies can bind, on either gp120 or gp41 (Burton, *et al.*, 2012). 10-25% of HIV+ patients produce bNAbs (Doria-Rose, et al., 2010). These antibodies recognize epitopes (the binding sites on the antigen) on the Env protein that are often functionally or structurally conserved, meaning the antibodies can neutralize several strains of virus from genetically diverse subtypes (i.e., breadth) (Doria-Rose, *et al*., 2010; Simek, *et al*., 2009). BNAbs are thus thought to be instructive for vaccine design against HIV-1. Current vaccine design efforts prioritize bNAb development over strain-specific antibody development because natural viral diversity renders strain-specific responses ineffective at protecting against infection (Burton, *et al.*, 2012). Immunogens that induce the development of bNAbs would therefore be a vaccine candidate since passive immunization has been shown to protect against infection in non-human models (Burton, *et al*., 2012; Haynes and Montefiori,

2006; Burton, *et al.*, 2009). The epitope may also span both regions of the protein (Kwong and Mascola, 2012). BNAb lineages take time to develop their breadth. Furthermore, the rapid mutation rates of HIV-1 enable the virus to produce genetic mutations that may confer viral escape from neutralization (Landais, *et al.*, 2016). The rapid mutation rates of Env lead to a cycle of antibody maturation and viral escape in an infected patient, which ultimately has shown that even with bNAb development, there may not be viral suppression (Landais, *et al*., 2016; Doria-Rose, *et al*., 2015).

The strong possibility of viral escape from a bNAb does not mean that designing a vaccine around eliciting a broadly neutralizing response is fruitless. Studies have shown that non-human primates with passive, intravenously administered bNAbs are able to protect against natural infection after SHIV challenge (Burton, *et al.*, 2009).

BNAbs must be able to tolerate the large degree of HIV-1 sequence variability within Env. Besides sequence variability, there are certain structural qualities of Env that the bNAb must circumvent in order to bind. For example, bNAbs must be able to bind despite extensive glycosylation on the Env surface that shields the epitope (Burton, *et al*., 2012). It is important to identify which bNAbs are able to provide the greatest amount of coverage. Factors that allow bNAb classification include epitope, rate of maturation, and degree of somatic hypermutation (which is the extent of mutation within the variable region of the bNAb genes, and a crucial step in the development of a bNAb to maturity) (Walker, *et al*., 2011). Another problem with bNAbs is that it can take years (e.g. up to 4-5 years) to develop bNAbs naturally within a patient, and this development time poses a challenge to making a vaccine accessible to the population at large (Bonsignori, *et al.*,

2016). Therefore, further characterization of these bNAbs and additional understanding of their development is crucial to immunogen design for a vaccine.

An experimental approach for examining the behavior of bNAbs against Env would be to express naturally diverse Envs from patient samples in transduced cells. Previous preparations of Env clones have yielded success in expressing Env on the surfaces of cells (Bruun, *et al*., 2017). Transduced cells with Env on the surface allow for the possibility of sequencing the *env* that has integrated into the cell genome. The ability to sequence would be helpful in identifying similarities between Envs that bind. However, the most common efforts built towards characterizing individual antibodies rely on cloning libraries that are not diverse enough to represent the variability of the Env from an individual patient. High-throughput characterization assays that will identify a bNAb's original development are greatly aided by full sequence variability of the Env population that is being used to test the bNAb. The greater that the sequence variability can represent the natural diversity of the donor Env, the higher the probability that the motif that elicited the development of the bNAb can be identified.

The preparations needed for examining diverse Env populations for bNAb interaction can be cumbersome. The efficiency of the cloning process that generates the *env* gene libraries needs to be high. Ideally, when the libraries are transduced into cells for protein expression, each cell expresses an individual variant. Additionally, the *envs* in this project come from donors that have elicited bNAbs. In these donors, it is not readily known which of the circulating Env proteins caused the rise of the bNAbs that the patient's B-cells produced, and the traditional means of determining this are time consuming and labor-intensive. Neutralization assays that indicate the concentration of

the bNAb necessary to neutralize a given percentage of the virus, for example, antibody concentration to neutralize 50% of the virus (IC_{50}), can be performed for single Env protein variants at a time (Martinez-Navio and Desrosiers, 2012). This works well, but if the development of a specific bNAb was caused by a rare Env variant, then it would take countless assays until the specific variant would be identified.

The cloning method described in this thesis will maintain sequence variability of the *envs* so that the libraries can be used for two ongoing projects. The first project that we aim to use our cloning for is an attempt at sorting HEK293T cells transduced with our donor *env* in such a way that ensures individually unique variants on each cell surface. The libraries of diverse *env* genes derived from a donor with samples at various timepoints will be cloned into pLentiIII-HA (ABM, Inc.). These constructs will be transformed into bacterial cells, wherein single colonies will contain DNA for individual *env* variants. The constructs containing *env* will be extracted from these individual colonies, and we selected some of the DNA extraction products to Sanger sequence so that we could verify the presence of *env* within the plasmids. The resulting plasmid will be co-transfected with a co-plasmid to produce a lentivirus. Once we obtained this lentivirus, we planned to transduce HEK293T cells to incorporate the *env* genetic material into the cellular genome, with each cell expressing one variant of Env on its cell surface. Following this, Fluorescence-Activated Cell Sorting (FACS) could be utilized to profile antibody binding to the donor Envs. After the performance of FACS, the sorted cell populations should have pronounced enrichment for a common feature (e.g. an increased frequency of the positively sorted cell population will have a shared trait, such as a specific glycan mutation). Deep sequencing could then be used to compare the

sorted cell populations and unsorted cell populations for differences in the sequences of positively sorted and control population cells. This process has the potential to be much faster and more cost-effective than prior screening protocols, and we have termed this project “Immunogen Fishing”. With our high-throughput cloning approach, once the HEK293T cell library has been established, a wide variety of bNAbs can be tested against these Envs. This means we could potentially identify rare Env variants as the antigens for certain antibody lineages.

The second project is a “High-Throughput Env Neutralization Assay”. To prepare for this project, we cloned *env* genes into Q23Δenv-eGFP, a modified proviral vector originally modified by the Bloom group at Fred Hutchinson Cancer Center (Haddox, *et al.,* 2016). Dr. Bloom’s project focused on assessing the effects of mutations in *env* on viral replication by inducing random amino acid mutations. Bloom’s proviral plasmids each contained random codon mutations in *env*, and these mutations were manufactured in the lab. The Bloom group transfected these mutant plasmids into 293T cells and passaged them in Sup.T1-CCR5 cells. The transfected cells, transfection supernatant, and Sup.T1-CCR5 cells were all deep sequenced to evaluate mutant selection. We decided to use a similar type of workflow to examine escape from neutralization in the context of naturally diverse donor *envs*. Cloning *env* into Bloom’s proviral vector enables us to grow live virus in cells that can be passaged. Ideally, these cells containing virions would be passaged in the presence or absence of a bNAb. Identification of the surviving virions after exposure to the bNAb using Next Generation Sequencing would be able to tell us the viral variants that were able to escape neutralization. This strategy can help us directly identify what members of the given Env population are capable of escape.

We successfully cloned *env* libraries for both potential projects and verified the presence of *env* in our constructs via gel electrophoresis and Sanger sequencing. The molecular cloning of *env* with the two aforementioned vectors led to some interesting results. We saw that pLentiIII-HA lentiviral vector *env* plasmid cloning maintained some natural diversity with the help of the RAD.jl tool (Kumar, *et al.,* 2018). These libraries were used to test long-read sequence analysis algorithms for denoising PacBio sequences (Kumar, *et al.,* 2018). These plasmids were also used for transduction of HEK293T cells to assess cell surface expression and sorting, but cell surface expression failed despite the genetic presence of *env.* Further efforts to troubleshoot the cloning and transduction process are necessary before cells with single variant cell surface *env* expression can be used for our methods. The insertion of *env* into Q23Δenv viral vector cloning was performed to produce live, growing virus for in vitro experiments. These proviral vectors can be used to assess which Envs grow in different cell types, or in the presence of certain neutralizing antibodies.

Our cloning strategy was designed with the intention of preserving the insert without scarring the DNA. We decided to approach this with a seamless cloning strategy, one that used PCR-generated primer overhangs (New England BioLabs). Other decisions concerning our cloning workflow were inspired by a previous study (Stano, *et al.,* 2017) that also used lentiviral transduction to express Env clones on cell surfaces, and which saw enrichment in the sorting results for truncated cytoplasmic tails, and this was perhaps also aided by constitutive expression of Env induced by a cytomegalovirus promoter, which is why we elected to use some of their expression strategy and clone with

constitutive CMV promoters in our pLentiIII-HA vectors. We also used CTT-truncated *env* as part of our cloning strategy.

Our diversity-preserving cloning protocols can contribute to understanding how bNAbs are elicited in certain donors, and in understanding how the Env population responds to these autologous bNAbs. Sorting and neutralization assays from diverse libraries would dramatically increase the rate of information obtained relative to performing experiments at a rate of one Env variant at a time.

CHAPTER 2: MATERIALS, METHODS AND RESULTS

2.1: MATERIALS AND METHODS

The overarching strategy for the cloning process starts with amplifying the *env* insert with PCR, which will attach our primers that eventually serve as our overhangs. We then prepare the backbone vector via digestion with the necessary restriction enzyme(s) for the specific vector. These two components will be ligated together using the NEBuilder DNA Hi Fi Cloning Kit.

PCR

The *env* DNA products that come from PC63 and PC39 are donor-derived Clade C *env* PCR products obtained from the Protocol C cohort (Landais, *et al.*, 2016). PC63 and PC39 are codes for donors that participated in this study. We were provided cDNA products of PC63 and PC39 and amplified these using Advantage 2 PCR Enzyme system (Clontech). We used the following protocol: After thawing short amplicon buffer, dNTPs, and primers, vortexing each, we created Master Mixes with the following contents:

Table 1. Contents of Advantage 2 PCR for PC39/63 *env*.

Component	Amount (µL/tube)	2.0x PC39	2.0x PC63
10X Buffer	2.5	5.0	5.0
50X dNTP	0.5	1.0	1.0
PC39/63 FWD Primer (10µM)	1.0	2.0	2.0
PC39/63 REV Primer (10µM)	1.0	2.0	2.0
Advantage 2 Polymerase	0.5	1.0	1.0
ddH_2O	17.5	38.5	38.5
DNA	0.5		
Reaction Total	25	50	50

After assembling the Master Mixes without the DNA, we added 0.5 µL cDNA product into the respective PCR tubes. We vortexed gently to mix and spun the tubes down, and then we ran the PCR under the following conditions:

Table 2. Reaction Conditions for PC39/63 *env* PCR.

Stage	Cycle	Temperature (℃)	Time	Additional Note
1	1	95	1 minute	
2	26	95	15 seconds	
		64	30 seconds	Decrease by 0.2 ℃ every cycle with plateau at 60 ℃
		68	3 minutes	
3	1	68	10 minutes	
4	1	4	Hold	

Table 3. Sequences of Forward and Reverse Primers used to amplify and insert PC63 *env* gene into Vectors.

>pL_PC39_Fwd	ccctatcagtgatagagagCTAGccgccaccatgagagtgagggRgatactg agg
>pL_PC39_CTT_Rev	ggtactcgagctgcagtctagaggatcctaTGCTAAGAATCCGTTCACTAATC
>pL_PC63_Fwd	ccctatcagtgatagagagCTAGccgccaccatgagagtgaaggggataYR gagg
>pL_PC63_CTT_Rev	GgtactcgagctgcagtctagaggatcctaTGCTAAGAATCCGCTCACTAATC
>Q23deltaEnv20180507Fwd Hutch	Agataggttaattgagagaataagagaaagagcagaagacagtggcaatgagag tgaaggggataCRgagg
>RevPrimerdeltaEnvMurrell Hutch	ctcctggtgctgctggaggggcacgtctcattctttctcttacagcaggccatcc

Backbone Vector Preparation for Immunogen Fishing

After generating our amplicons, we moved forward with constructing our digested backbones. We inserted the resulting amplicons into pLentiIII-HA. For the Immunogen Fishing project, we digested pLentiIII-HA with single cutting restriction enzymes BamHI-HF and NheI (New England Biolabs) for 1 hour at 37°C, treated with Shrimp Alkaline Phosphatase and incubated at 37°C for 30minutes, and then loaded onto a 0.8% agarose gel with TBE and ethidium bromide to run for 90 minutes at 100 V. The gel fragments of the digestion were cut on a UV block and purified using the Zymoclean Gel DNA Recovery Kit according to manufacturer's instructions (Zymo Research).

Seamless-Style Ligation Reaction and Transformation for Immunogen Fishing

The ligation reaction was performed using undigested *env* PCR product that had been run on a gel in the same conditions as above and purified in the same fashion, with digested, gel-purified backbone. The exonuclease component of the NEB DNA Hi Fi Master Mix removes some of the 5' end of the primer part of the molecule to create 3' overhangs that will bind together with the help of a DNA Ligase. The primers used were designed to be the overhangs on the ends of the *env* genes, so that the ligase within the NEBuilder Hi Fi DNA assembly mix could insert the ends of the *env* into the digested backbone (New England Biolabs). We used NEBuilder DNA HiFi Assembly Master Mix (New England Biolabs) to facilitate the ligation reaction and we transformed 5 ng of this cloning product into Stellar Competent *E. coli* (Clontech Laboratories), following manufacturer instructions except for performing the transformation in a 1.5 mL microcentrifuge tube instead of a 14 mL round bottom conical tube so that the transformants could be heat-shocked in the lab's water bath. A transformant volume of

100 µL out of 500 µL total volume was plated on Luria Broth (LB) agar plates containing kanamycin. Individual colonies picked the from the transformant plates were grown in 4mL LB with added kanamycin and plasmid DNA was extracted from the broth using the Monarch Plasmid Miniprep Kit (New England Biolabs). A 0.8% agarose+ethidium bromide+TBE gel was run to verify the expected plasmid size. We performed gel clean-ups of our inserts, backbones, and plasmids all with the Zymoclean Gel DNA Recovery Kit (Zymo Research). We verified the concentrations of our inserts, backbones, and plasmids using the Nanodrop Spectrophotometer (Thermo Fisher Scientific).

It should be noted that we were never able to fully eliminate negative control colonies from their respective plates. For the Immunogen Fishing cloning procedures, which consisted of the NEBuilder reaction of PC63 V36 in pLentiIII-HA, once we realized that the cloning reaction was efficient in terms of yielding colonies more than double the colony number of a negative control plate, we repeated the transformations but increased the number of transformation tubes (individual aliquots of competent *E. coli*) to 12. We did this to increase the number of *env* variants to preserve diversity for more downstream applications for the two projects. Thus, we performed 12 side by side transformations of PC63V36 into 12 aliquots of 50µL Stellar cells. Each of these 12 transformations went through the standard Stellar Competent Cell workflow through the 1-hour shake stage at 37°C, and then these twelve transformants were merged into one 1000 mL Erlenmeyer flask with a total transformant volume of 100 mL once the Luria Broth with Kanamycin was added. We have called this mixture of transformants the pLentiIII_PC63 pooled mixture.

We needed to ensure that we had inserted *env* into the vector, so we took the plasmid products and digested with the same restriction enzymes as the ones used backbone digestion or similar. For example, for the Immunogen Fishing cloning, we occasionally digested with a similar enzyme that was still in the multiple cloning site, such as HindIII. For the digestion of pLentiIII-HA containing *env* in this thesis, we show an image digested with NheI and XbaI instead of BamHI and NheI because BamHI cuts some *envs*, but not enough to appear on a consensus sequence for single cutter enzymes (Fig. 5). We ran the digested plasmid products on a 0.8% agarose + TBE/EtBr gel to look for the correct band size that corresponds with an incorporated insert. Once we knew we had a band size that indicated the presence of *env,* which would be either a 2.2 kb band or a 2.6 kb band to depending on the genetic presence of a cytoplasmic tail, we proceeded with picking individual colonies from our transformation plates to be Mini Prepped. We submitted the individual colony mini prep products that comprised the mixture to Sanger sequencing from Genewiz using standard Sanger Preparatory instructions (Genewiz). Alignments of Sanger sequencing were performed using Aliview and Benchling Alignment Tool using MAFFT Algorithm.

Verification of Ligation Reaction for Immunogen Fishing

Once both our digest results and our Sanger sequencing results dictated that we had *env* present in the mixture, we elected to proceed with Next Generation Sequencing (NGS) using the mixed midi prep as the DNA template. In order to obtain a high enough number of single-colony *env* variants to compare the sequences to each other, we retransformed our pLentiIII_PC63 pooled mixture of 12 transformations in Stellar Competent E. coli and plated a 1:1000 dilution. We increased the dilution factor to

1:1000 to ensure that there would be enough space on the plate to pick individual colonies. From this plate, we selected 96 individual colonies and mini prepped them. The mini preps were then barcoded by Project Scientist Thomas Vollbrecht and library prepped by the UCSD Translational Virology Core. The resulting library prep products were deep sequenced using the PacBio RS-II by the UCSD Institute for Genomic Medicine.

Cloning Workflow for High-Throughput Env Neutralization

For the High-Throughput Env Neutralization Assay project, our group also wanted to grow virus using the *env* library we created for further study. Thus, we cloned PC63 V36 into Q23Δenv (Haddox, *et al.*, 2016), which is a modified HIV vector without *env* that we obtained from Jesse Bloom at Fred Hutchinson Cancer Research Center. We decided to clone PC63 into Q23Δenv because of Jesse Bloom's success with cloning mutated *envs* into the vector with the same restriction site that we used for this project. We also designed our primers according to the basic template of how Bloom's primers were structured, with a preference for equal melting temperature rather than for equal length (Dingens, *et al.*, 2017). We digested the vector first with BsmBI (New England BioLabs) and designed primers using the same overhangs from Jesse Bloom's *env* primers while adjusting the *env*-matching portion of the primers to match the *envs* that we were using to perform the cloning based on our consensus sequences for PC63. Bloom and colleagues did not use PC63 for their cloning (Haddox, *et al.*, 2016*)*

Once we were able to amplify the insert and construct the overhangs using the Advantage 2 PCR kit as described above, we largely repeated our previous workflow to insert the *env* into the vector using the NEBuilder DNA HiFi Assembly Master Mix. We

then performed 12 transformations into Stellar Competent Cells according to manufacturer's instructions except for performing the transformation in a 1.5 mL microcentrifuge tube instead of a 14 mL round bottom conical tube, so that the transformants could be heat-shocked in the lab's water bath. We grew the transformants each in (12) 14 mL round bottom conical tubes for 16 hours using 4 mL LB media with added ampicillin. We extracted DNA from the cultures using the Monarch Plasmid Miniprep Kit from New England Biolabs and then verified the concentration of the plasmid prep products using the Nanodrop Spectrophotometer (Thermo Fisher Scientific). We then mixed 400 ng from each prep product into one Q23_PC63 pooled mixture tube for further downstream work. We retransformed this mixture of mini preps and scaled up the DNA using a Midi Prep of the transformant, and we also retransformed the gel extraction product of this mini prep mixture. These two midi prep products were used to transfect HEK293T cells to see if virus could be produced.

The Generation of Infectious Virus for High-Throughput Env Neutralization Assay Project

We transfected our *env* in Q23Δenv plasmids into HEK293T cells because we wanted to assess the ability of the proviral plasmid to create virus for further assays and cellular studies. We seeded HEK293T cells with 3.5 x10^5 cells per well for 2 wells within a 6-well plate. We transfected 2000 ng of PC63 *env* plasmid into each of these wells using Lipofectamine 3000 as the transfection agent (Thermo Fisher Scientific). We prepared the transfection reagent by diluting Lipofectamine 3000 in Opti-MEM and vortexing for 3 seconds. Our master mix of DNA, P3000 reagent, and Opti-MEM consisted of 100 uL Opti-MEM, 2 ug DNA, and 4 uL P3000. We added the master mix to

the diluted Lipofectamine 3000, incubated for 10-15 minutes at room temperature, and then added 200 uL of transfection mix per well. After swirling the 6 well-plate to mix, we incubated the transfections at 37℃ for 72 hours

Following the transfection incubation period, we collected the virus using a 20% sucrose cushion. We spun the sucrose cushion at 23,500 RCF for 1 hour at 4℃. After resuspending each viral pellet in 200 µL PBS.100 µL of the sample was used for a Tzm-bl infectivity assay. The Tzm-bl assay was performed in triplicate, and cell-only negative controls were performed in quadruplicate on 48-well plates.

In order to stain and read the Tzm-bl assay, we removed the supernatant from the wells, added 70 µL Triton-X 0.05% per well as well as 70 µL BriteLite per well.

Next Generation Sequencing Analysis

The RAD.jl webserver (Kumar, *et al.*, 2018) was used to de-noise the PacBio reads for the pLentiIII_PC63 *env* library. Phylogenetics tree of the pLentiIII_PC63 *env* library were constructed in two ways: one tree was constructed based on analysis of the barcoded sequences, performed by Ben Murrell and colleagues (Fig. 15). Another phylogeny was inferred by the Neighbor joining algorithm (Saitou and Nei, 1987) (Fig. 14), with the sequences for this phylogeny based on the results from uploading the fastq file of the PacBio sequencing run from the UCSD Institute for Genomic Medicine (IGM) into the RAD.jl denoising tool.

2.2: METHODS FIGURES

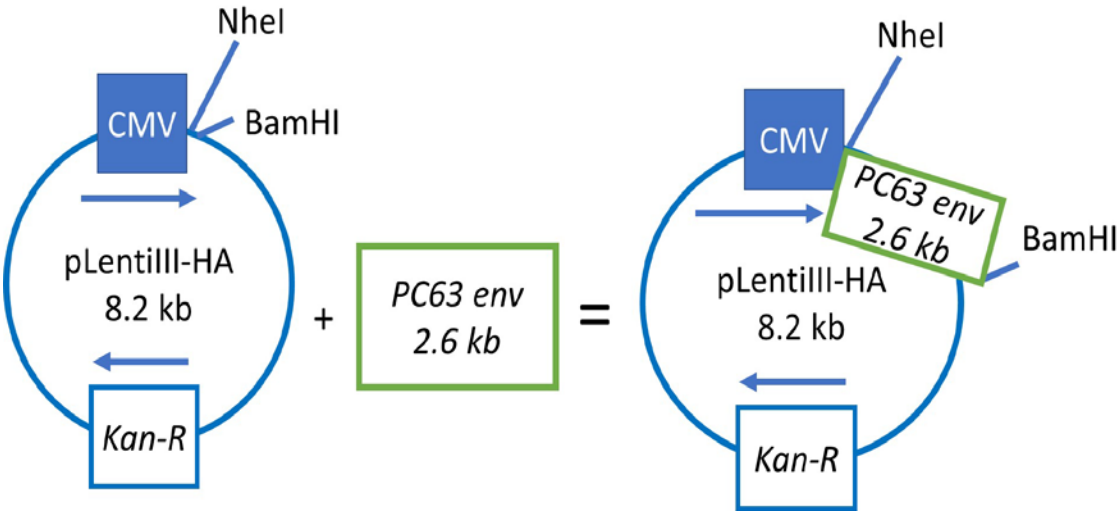

Figure 1. Schematic for construction of *env* plasmids. Primer design was done to amplify *env* starting with the beginning of the gene, right at the first start codon. Amplifying *env* with these primers resulted in high efficiency cloning for PC63 V36 in both pLentiIII-HA and Q23Δenv vectors, but cell surface expression was not successful for the pLentiIII-HA plasmid. Coloring: Blue signifies the vector component, while Green signifies the insert component.

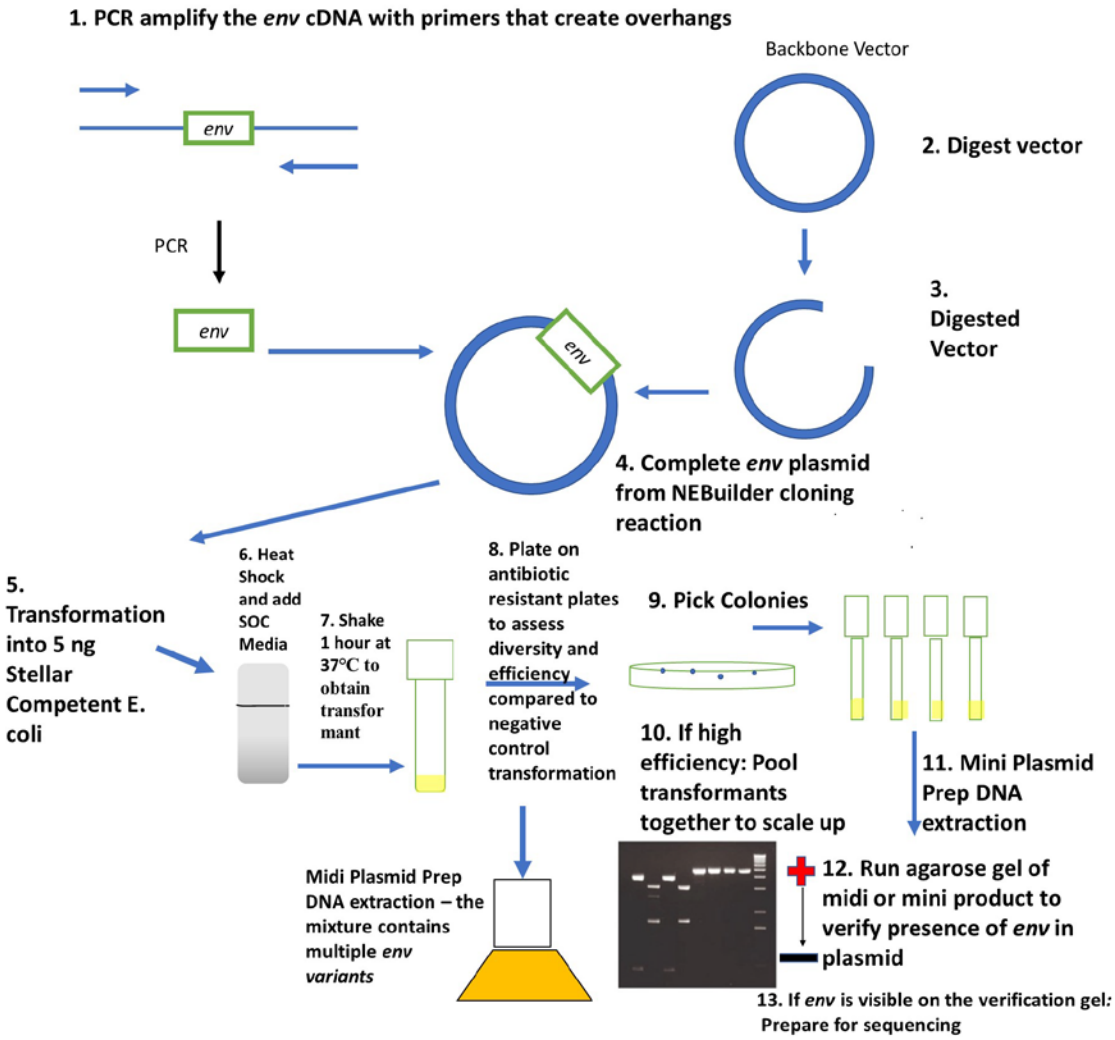

Figure 2. Workflow of *env* plasmid construction with Q23Δenv. The diagram illustrated here signifies the workflow for both vectors used. PCR was performed to amplify the patient-derived *env* genes and attach the primers to be used for the cloning reaction. The *env* insert was not digested, but the Q23Δenv backbone was digested with BsmBI restriction enzyme, which is a double cutter. The insert and backbone were cloned together using the NEBuilder Hi Fi DNA Assembly Cloning Kit. 5 ng of NEBuilder cloning reaction product was transformed into Stellar Competent E. coli using manufacturer's instructions. After the initial growth phase induced by a shake of 1 hr at 225 rpm, cloning efficiency was assessed by plating 20% of the total transformant volume on antibiotic resistant plates. We proceeded with scaling up the transformant and mini prepping individual colonies if the plates containing full plasmid had more than double the number of colonies on the backbone-only negative control plate. We verified the presence of *env* by digesting midi or mini prep products and running the plasmid on a gel to visualize a band indicative of *env* presence. Coloring: Blue signifies the vector component, while Green signifies the insert component.

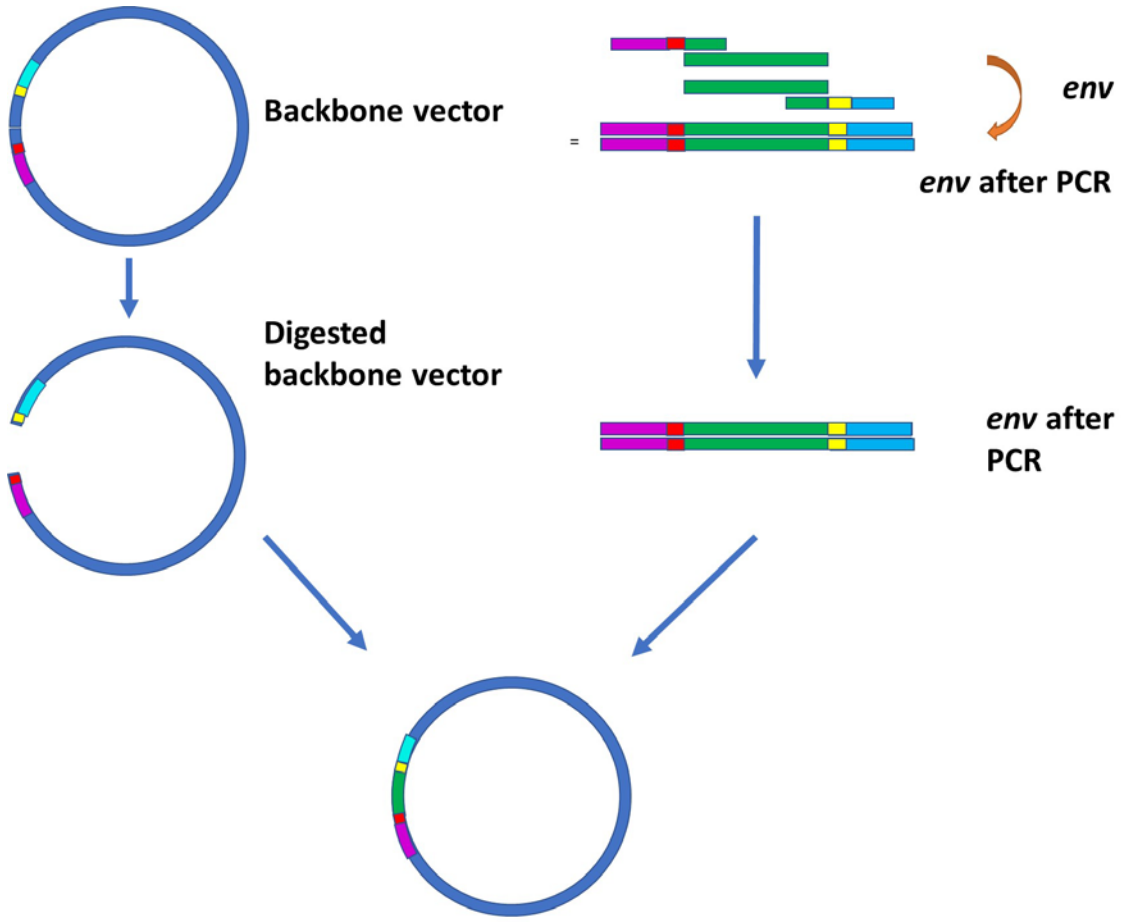

Figure 3. *env* cloning mechanism. Schematic on PCR-induced overhangs for *env* and incorporation into digested vector using NEBuilder Hi Fi DNA cloning kit using Advantage 2 PCR kit with 2SA buffer. This approach was utilized for pLentiIII-HA and for Q23Δenv backbone vectors.

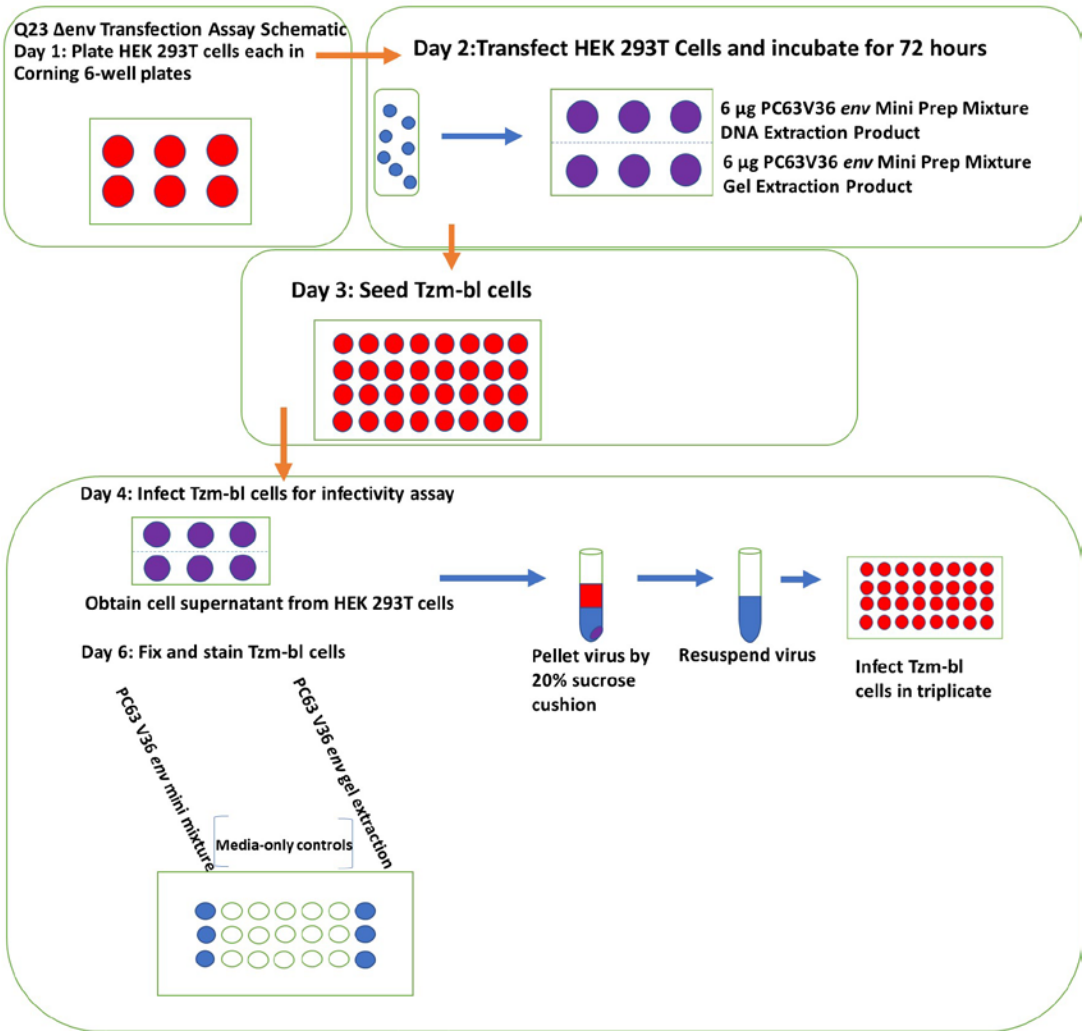

Figure 4. Q23Δenv Transfection Assay Schematic. Day 1: HEK293T cells plated in 6-well plates. Day 2: Proviral Q23Δenv plasmid transfected into HEK 293T cells. Day 3: Tzm-bl cells plated on 48-well plate. Virally-infected supernatant was taken HEK293T cells. The supernatant was used was used to measure infectivity. A standard Luciferase assay was performed in triplicate using Tzm-bl cells.

2.3: RESULTS

Construct Design and Cloning Efficiency

The primer design for the Q23Δenv primers and for the pLentiIII-HA primers were both successful for the PC63 *env* variant, but we did not achieve the same cloning consistency in pLentiIII-HA for the PC39 *env* as we did with PC63 *env* (Fig. 1). However, we did achieve amplification of PC39 and the primers we used were same as PC63 with respect to the overhang portion, but the amplicons did not ligate well. Both donor *env* PCR products were subjected to the same general workflow (Fig. 2). For PC63, we continued our workflow past the colony stage if we knew that the colony numbers of the experimental plates were above double the number of negative control plate colonies. We didn't consistently reach this objective with PC39 cloning. Cloning efficiency, in terms of the number of colonies surpassing the number of the negative control colonies, is the basis by which we considered a cloning reaction to be successful. In this case, for a pLentiIII-HA and PC63 V36 *env* plasmid cloning attempt, we saw 2 individual colonies on an LB-Kanamycin plate for the pLentiIII-HA negative control, which lacked insert and was a cloning reaction performed only with a digested backbone and extra nuclease-free water to equalize the volume with that of an experimental cloning reaction. Because we plated one-fifth the amount of the total transformant, we expect that there were approximately 10 circularized plasmids containing only pLentiIII-HA without *env* in solution. On this same day, from the same experiment, we plated a transformation of PC63 V36 with an individual colony number of 87. This colony number lead to an estimation of 435 circularized plasmids in the total transformant volume, and with approximately 10 of these plasmids lacking insert because of backbone religation, this

would mean approximately 425 plasmids would contain unique *envs*. For the PC39 V36 cloning reaction transformation on this same day, the plate revealed 6 colonies, so we estimated that there were only 30 plasmids in the transformant. While this number was above double the number of negative control colonies, this efficiency was not as successful, suggesting a reduced number of *envs* in bacteria compared to PC63 cloning. To add on to this, for our successful cloning of PC39 V24 into pLentiIII-HA, we obtained a plate of ten colonies for PC39 V24 and 2 colonies on the negative control plate. The cloning reactions for both sets of backbones resulted in colony counts well above negative controls but the ratio was marginally improved for this attempt compared to other PC39 cloning attempts. Examples for the array of colony numbers we obtained were taken using the Colorimetric Setting on a ChemiDoc XRS+, for both retransformation at high dilutions and for initial plasmid cloning attempts (Fig. 12).

For the PC63 V36 and Q23Δenv transformation workflow into Stellar *E. coli*, when we performed the 12 transformations to generate a pooled midi product as a source *env* diversity, we randomly selected 4 of the transformation products to plate as well as a negative control plate (Fig. 2). The four plates had individual colony numbers of 114, 183, 13, and 195, respectively, and the negative control plate had an individual colony number of 9. The only plate that did not meet our standard of efficiency was the plate with 13 colonies.

Our cloning strategy with exonuclease generated overhangs proved to consistently insert *env* into the backbone vectors for both pLentiIII-HA and Q23Δenv (Fig. 3).

Transfection of *env* plasmids and measuring infectivity

The transfection of Q23Δenv plasmid containing PC63 was successful, yielding between 1.80 and 3.45 million Relative Light Units (RLU) upon assessment of the viral infectivity using a Tzm-bl assay (Fig. 4, Fig. 10). These results are at least 3 orders of magnitude higher than the Media-only controls. However, because there was no standard in terms of a certain number of Relative Light Units (RLU) that indicated the presence of a given number of virions, the only thing that we can definitively say about the infectivity of the proviral plasmid DNA that we constructed was that it was capable of generating virus.

Gel Imaging

For our two cloning strategies it was necessary to verify our plasmids by checking the size of the plasmids on 0.8% TBE+EtBr gel. We digested the Plasmid Prep products of the plasmid in question with the same or extremely similar (in the same multiple cloning site) restriction enzymes that were used to generate the backbone. Digesting the plasmid enables us to see the *env* band and the digested backbone band in the same lane. The pooled midi product of PC63 V36 *env* in pLentiIII-HA showed both insert and backbone vector bands on the agarose gel (Fig. 5). The same can be said for PC63 V36 *env* in Q23Δenv (Fig. 6)

In one case, we verified the size of a gel band for a cloning attempt of PC63 V36 *env* into pLentiIII-HA using NheI and XbaI. The BamHI I cut site is in some *env* variants, but not all of them, and so XbaI cut some of the *env*s in two, but the visible smaller bands sum up to the 2.2 kb size of the PC63 V36 env with its cytoplasmic tail removed (Fig. 7). We digested the pLentiIII-HA backbone with BamHI and NheI, showing that the vector size is only 100 bp away because the DNA fragment spliced out is at the bottom of the

gel near the corresponding 0.1 kb region of the ladder. (Fig. 8). NheI, BamHI, and XbaI all single cut in the multiple cloning site of pLentiIII-HA. The size of the piece of DNA spliced out from Q23Δenv as a result of the double cutting BsmBI was much bigger, at 1.3 kb (Fig. 9).

For the insertion of PC39 V24 into pLentiIII-HA, we used HindIII as a restriction enzyme to verify the cloning on a gel image because HindIII cuts pLentiIII-HA five times in different places all along the vector, which assesses the integrity of the entire vector, not just the multiple cloning site (Fig. 11). We saw an additional band in the PC39 V24 *env* clone that was amplified with primers that cut off the cytoplasmic tail. We did not see the same elevated band in the PC39 V24 *env* clone that was amplified with a different primer that maintained the full *env.*

Comparing nucleotide sequences of PC63/PC39 *env* to consensus sequence in Benchling

Sanger sequences of full *env* plasmids were obtained from Genewiz, using standard instructions for sequencing preparation (Genewiz). The nucleic acid sequences were compiled and compared to the corresponding consensus sequence using Aliview and Benchling Alignment Tools, using the MAFFT configuration on Benchling. These results indicated that *env* was successfully inserted into the pLentiIII-HA and Q23Δenv.

Natural Diversity and Next Generation Sequencing

The PC63 V36 in pLentiIII-HA *env* plasmids were used to assess the validity of a newly developed denoising tool for long amplicons, RAD.jl (Kumar, *et al.*, 2018). This tool was created to edit long-read sets of data, such as a collection of reads from PacBio RSII sequencing, to “clean up” the reads, resulting in sequences without the error

typically involved in the sequencing run. The template used to assess the RAD.jl denoising tool was the pLentiIII_PC63 pooled mixture. We retransformed the pooled mixture and picked 96 colonies. To check the validity of the tool, we added identifiable forward and reverse sequencing barcodes, performed by Thomas Vollbrecht of the Guatelli Lab. These DNA sequences with added barcodes were amplified and pooled for sequencing. The barcodes can used to keep track of the sequences that come from each well to obtain accurate deep sequences, Once this sequencing was performed, Dr. Murrell and his team were able to computationally remove the barcodes, and then used the sequences without the barcodes to ask the question: how well does the RAD.jl tool organize and declutter the sequences without the barcodes? Please see (Kumar, *et al.*, 2018) for details on the long-read amplicon denoising tool. A side effect of performing this NGS analysis was that we were able to confirm the diversity of our pLentiIII-HA_PC63 cloning library. We were able to construct a phylogeny from the variants of the 96 wells that were barcoded (Fig. 15). In addition, I was able to examine the diversity of the cloned libraries myself by uploading the IGM fastq file of the sequencing run, obtained from the IGM core through Professor Murrell, and perform a de novo reconstruction of the variants using the RAD.jl webserver. The results revealed that the DNA sequences of the run performed came from 114 templates (Fig. 13). We originally intended to input only 96 templates, which is based on the number of Mini Prep products that we submitted for NGS, but we believe that the sequencing run identified 114 templates because more than one colony was unintentionally selected from the transformation plates during the mini prep stage. The RAD.jl webserver also provides a phylogeny, inferred by the Neighbor Joining algorithm (Saitou and Nei, 1987) based on

the 114 DNA template sequences from the library (Figure 14). These observations indicated the presence of at least 96 distinct variants, so therefore we expect that there are a large number of *env* variants present in the pooled mixture.

2.4: RESULTS FIGURES

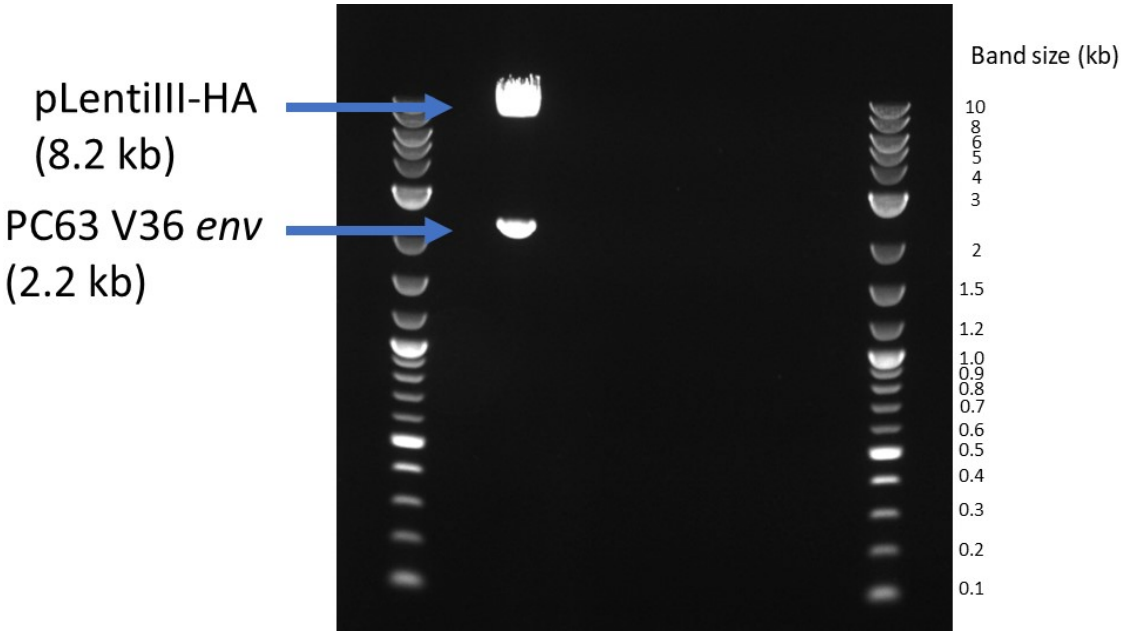

Figure 5. PC63v36 pooled midi digested with XbaI and NheI. 0.8% TBE and Ethidium Bromide Agarose gel confirms the presence of *env* in pLentiIII-HA. 2-log Ladder (New England BioLabs) was run alongside the digested plasmid to visualize. This *env* is only approximately 2.2 kb long because the reverse primer cut off the cytoplasmic tail.

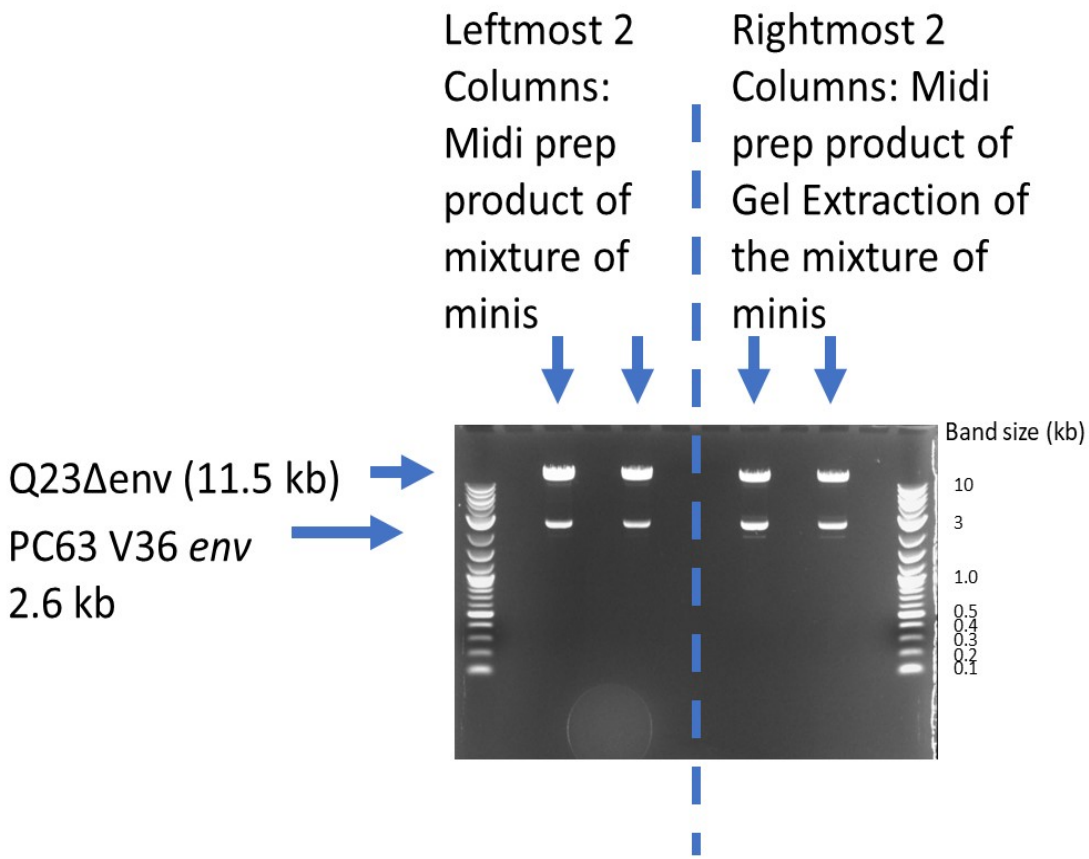

Figure 6. Q23Δenv and PC63 V36 env plasmid digested with BsmBI. Full proviral plasmid confirmed using a 0.8% TBE and Ethidium Bromide Gel, with 2-Log Ladder used to compare band sizes. These *env* bands are slightly bigger than the *env* bands seen in the gels for the pLentiIII-HA plasmids because the reverse primers for cloning into Q23Δenv do not cut off the cytoplasmic tail.

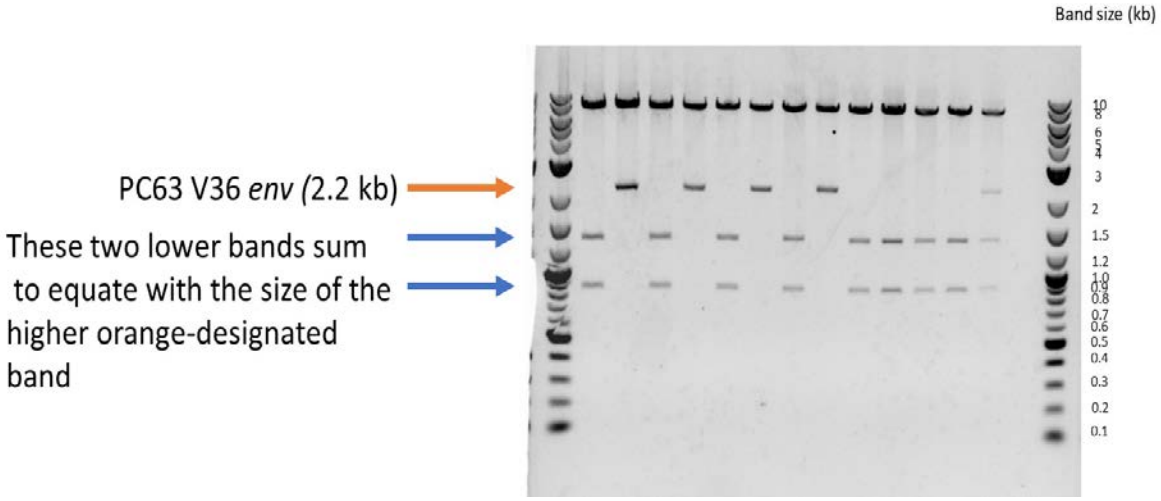

Figure 7. PC63V36 in pLentiIII-HA digested with NheI and BamHI. This 0.8% TBE and Ethidum Bromide in Agarose gel contained the mini prep products of individual colonies as opposed to Midi Prep products as seen in prior gels. For the restriction enzyme digestion of these mini preps, BamHI cuts on some variants of *env*, but not all. This is why some *env* bands have been cut and appear in the gel as two smaller bands. However, adding the approximate sizes of the two smaller bands approximates the size of the *env* band that we expect in the plasmid.

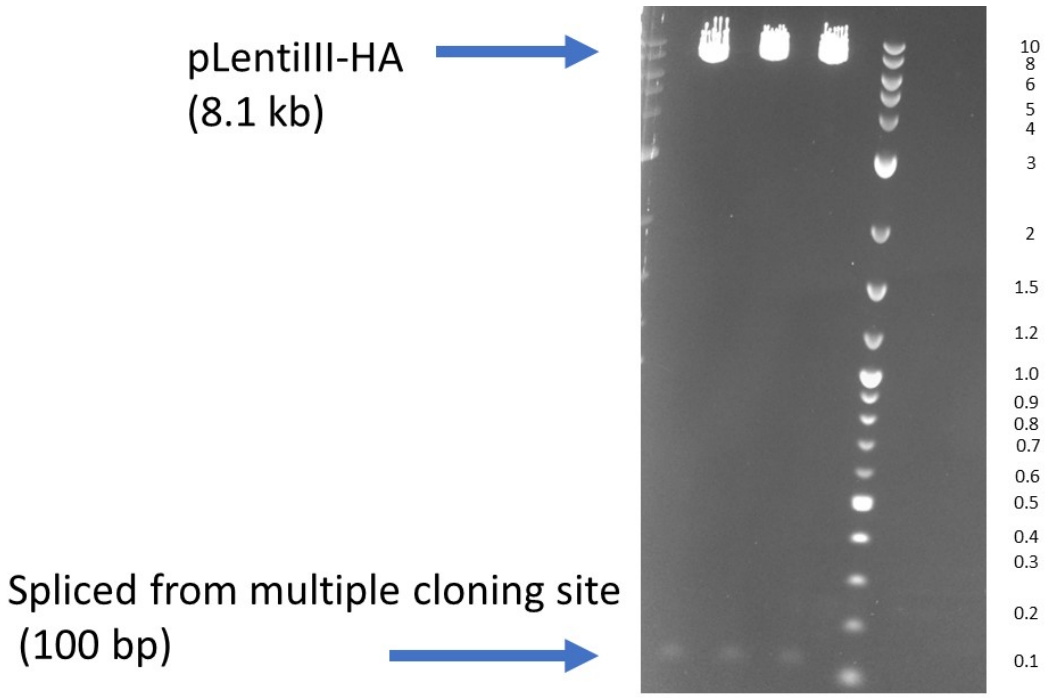

Figure 8. pLentiIII-HA + NheI+BamHI Vector Check. This Agarose gel visualizes the size of the pLentiIII-HA backbone without *env*, the DNA spliced out of the Multiple cloning site of the plasmid can be seen at the bottom of the gel.

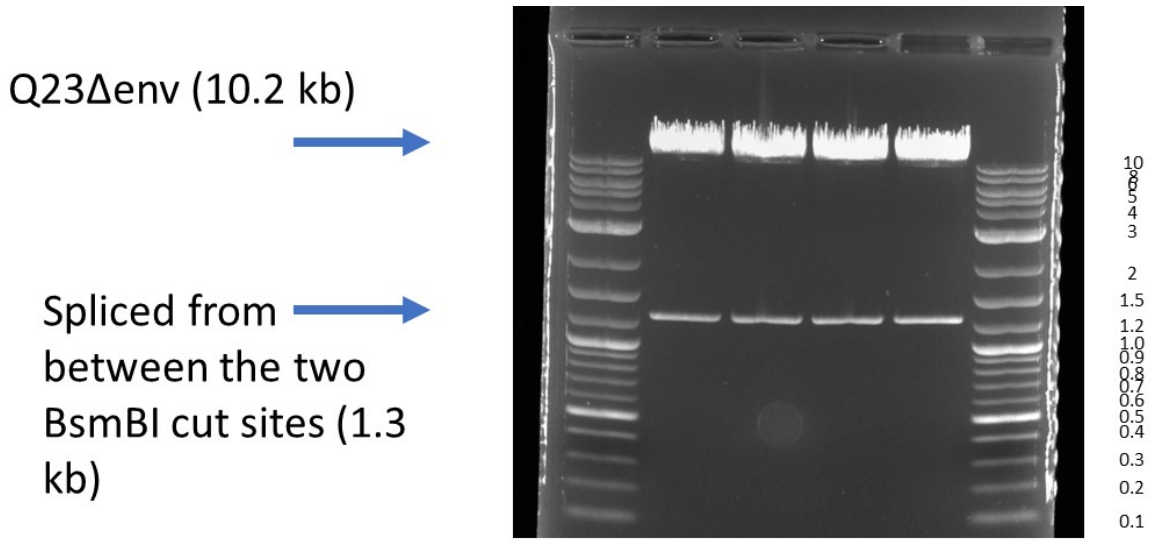

Figure 9. Q23Δenv backbone vector digested with BsmBI. This 0.8% Agarose gel assesses the size of the proviral vector given to our lab group by Adam Dingens. There is a 1.3 kb difference between the two cut sites of the BsmBI double cutter. 2-Log Ladder is provided for band comparison.

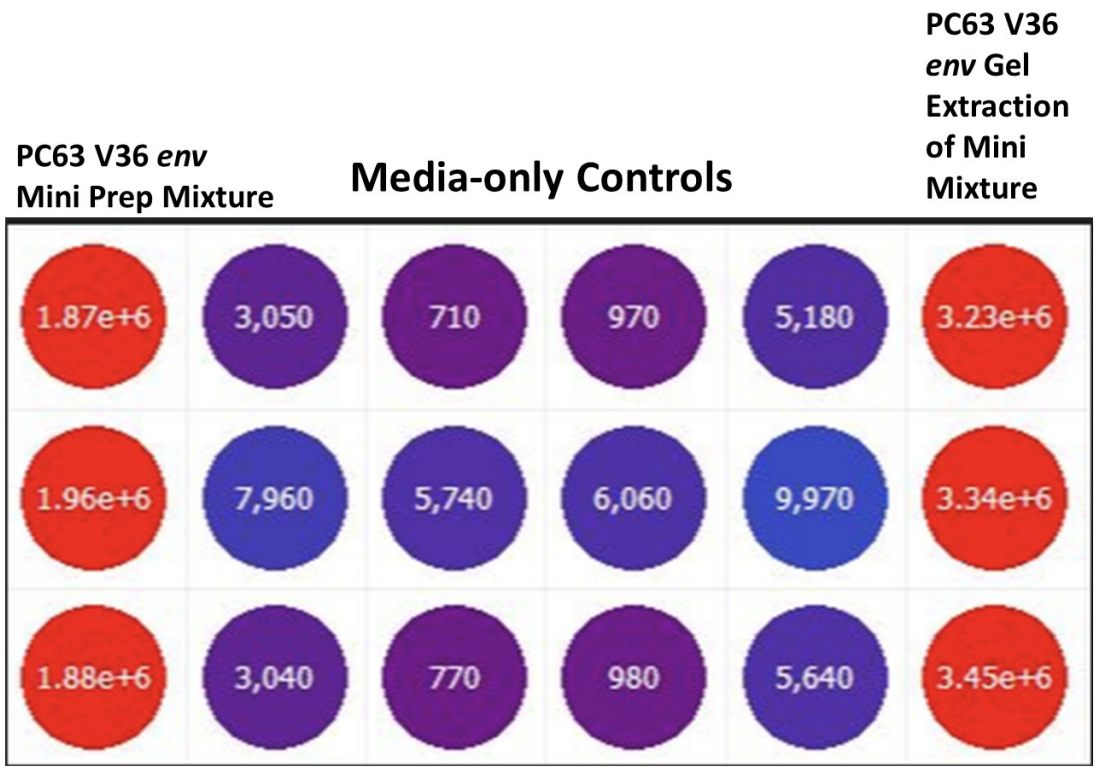

Figure 10. PC63 V36 in Q23Δenv Luciferase Assay Results. Level of infectivity is given here in Relative-Light Units (RLU). The DNA used to perform the original transfection of HEK 293T cells that produced virus was from the midi prep products of a retransformed pooled mixture of mini preps, and the retransformation of the gel extraction of the same pooled mixture of mini preps. Both types of DNA were able to create infectious virus in HEK293T cells.

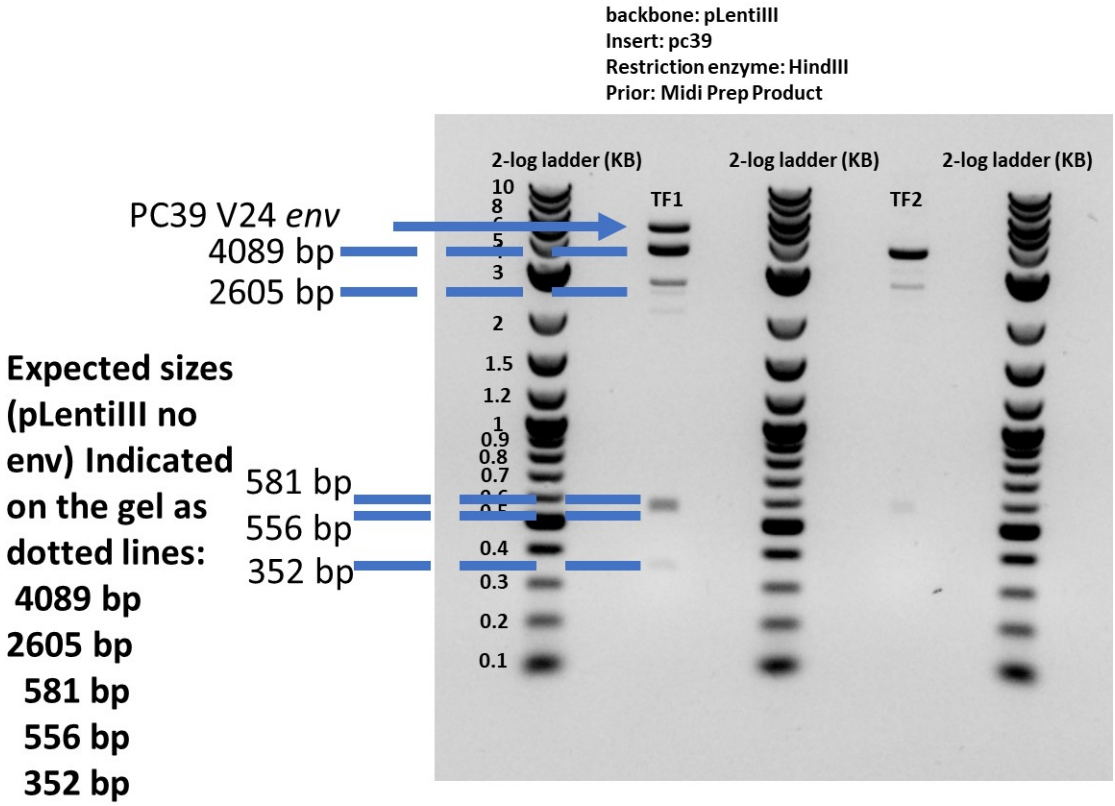

Figure 11. PC39 V24 *env* in pLentiIII-HA, digested with HindIII. The two products run on this gel were from the same date of transformation, and both were PC39 V24 *env*. However, only one of the transformants resulted in a cloned *env*. This transformation was successful but not particularly efficient.

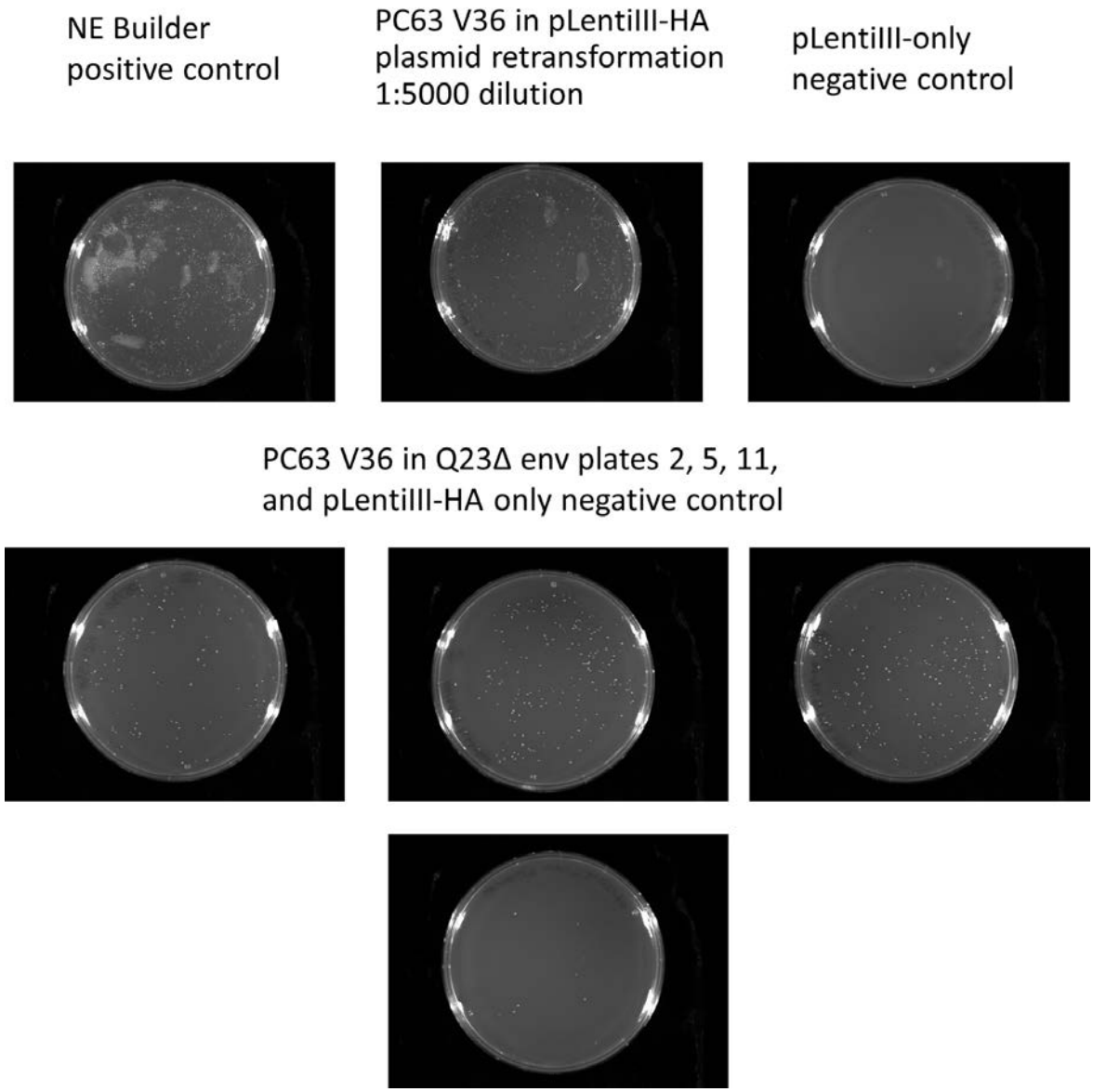

Figure 12. Colorimetric Images of LB-Antibiotic Plates from *env* molecular cloning workflow. Images taken using the ChemiDoc XRS+ (BioRad). All of the cloning reaction plates seen here satisfy our efficiency requirement of having more than double the colony number compared to the negative control plate. The positive control plate used the NEBuilder Positive Control (2 strands of dsDNA ligated together, facilitated by the Master Mix) indicated a functional enzyme mix from the cloning kit used. The negative control plates contain the transformation of an NEBuilder cloning reaction without insert, so the reaction only contains vector.

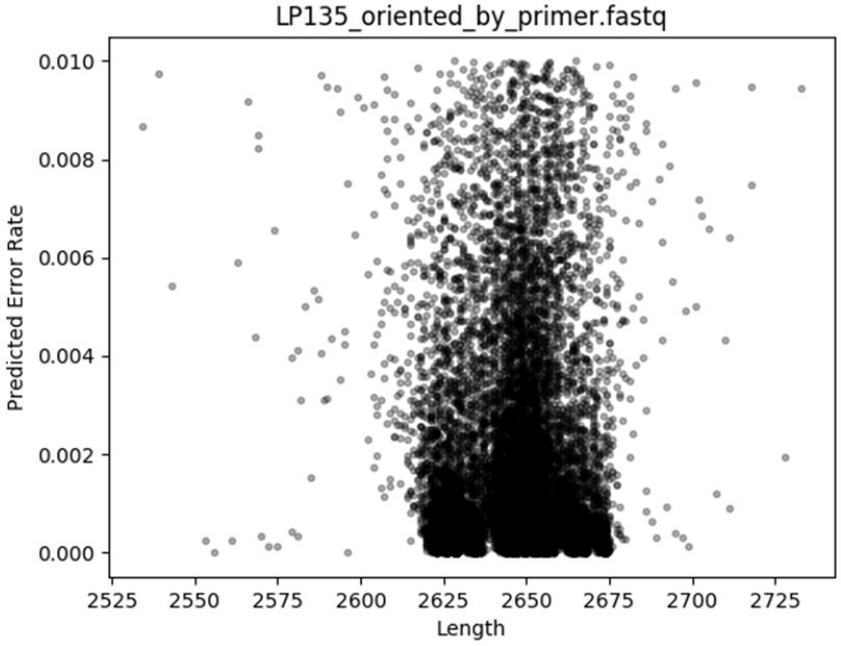

Figure 13. Dot Plot of the PC63 V36 in pLentiIII-HA *env* library. The plot indicates that the error rates for the *env* PacBio sequences range from 0.0 to 0.01. This plot represents the 19445 sequences run through the PacBio RS-II sequencer, which utilizes SMRT sequencing. Despite the noisiness of the reads, Ben Murrell's Robust Amplicon Denoising Program (RAD), which is a Julia package, determined that the library comes from 114 templates. 114 templates is a surprising number, because the DNA templates come from 96 wells. We hypothesize that 114 templates resulted from the unintentional inclusion of multiple colonies for some mini preps because the high colony density of the plate that the colonies came from caused multiple colonies being picked for one prep.

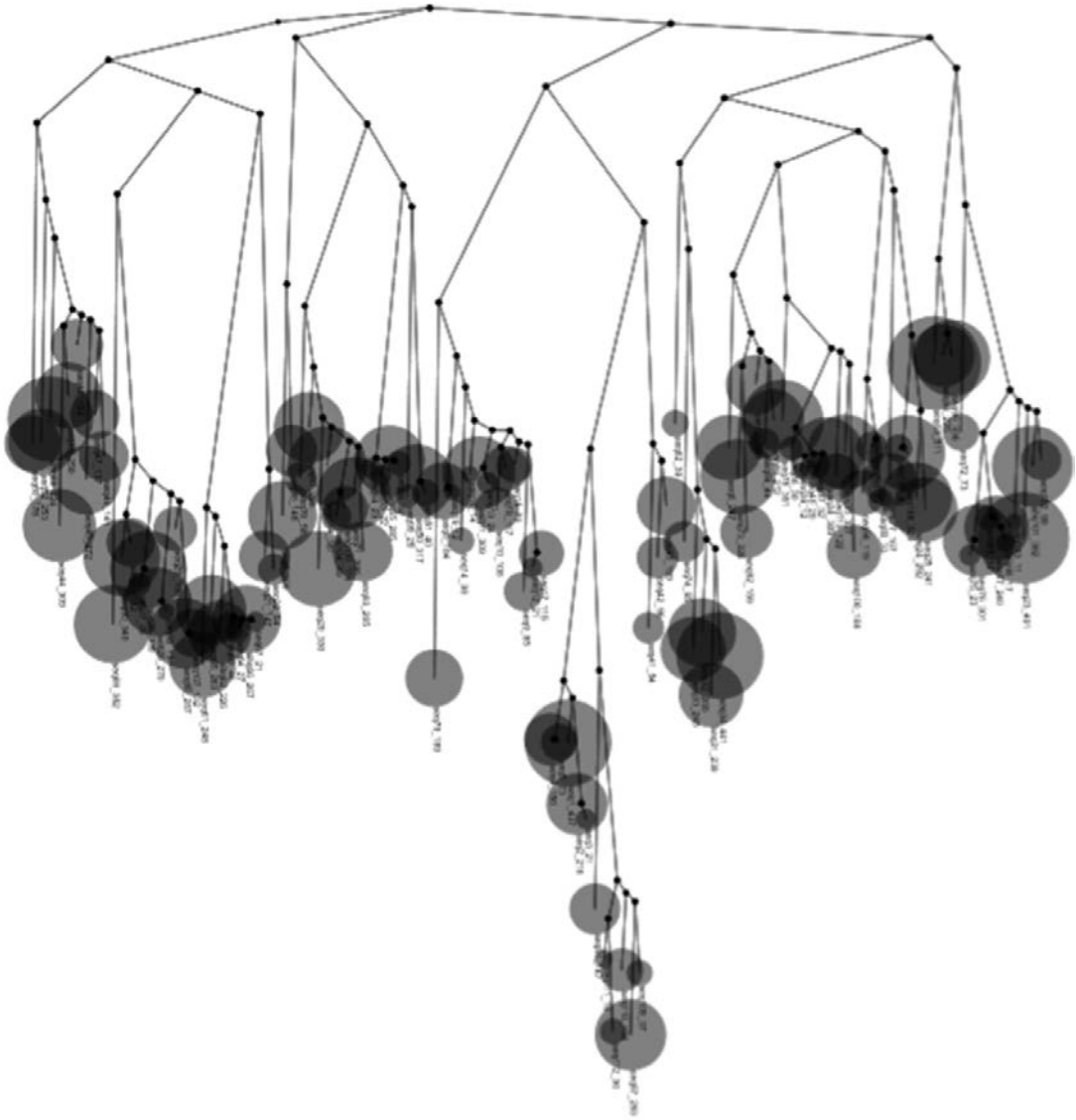

Figure 14. Cluster tree of the PC63 V36 in pLentiIII-HA *env* sequences. This tree is the result of an inference program in the Julia computing language that utilized the algorithm "neighbor joining" (Saitou and Mei, 1987) within the RAD webserver (https://tools.murrell.group/denoise). The sequences that comprise this tree come from the 114 templates that resulted from the denoising of the *env* library that was created. This tree indicates that the natural diversity was maintained in the *env* library that we created. Each bubble represents one template.

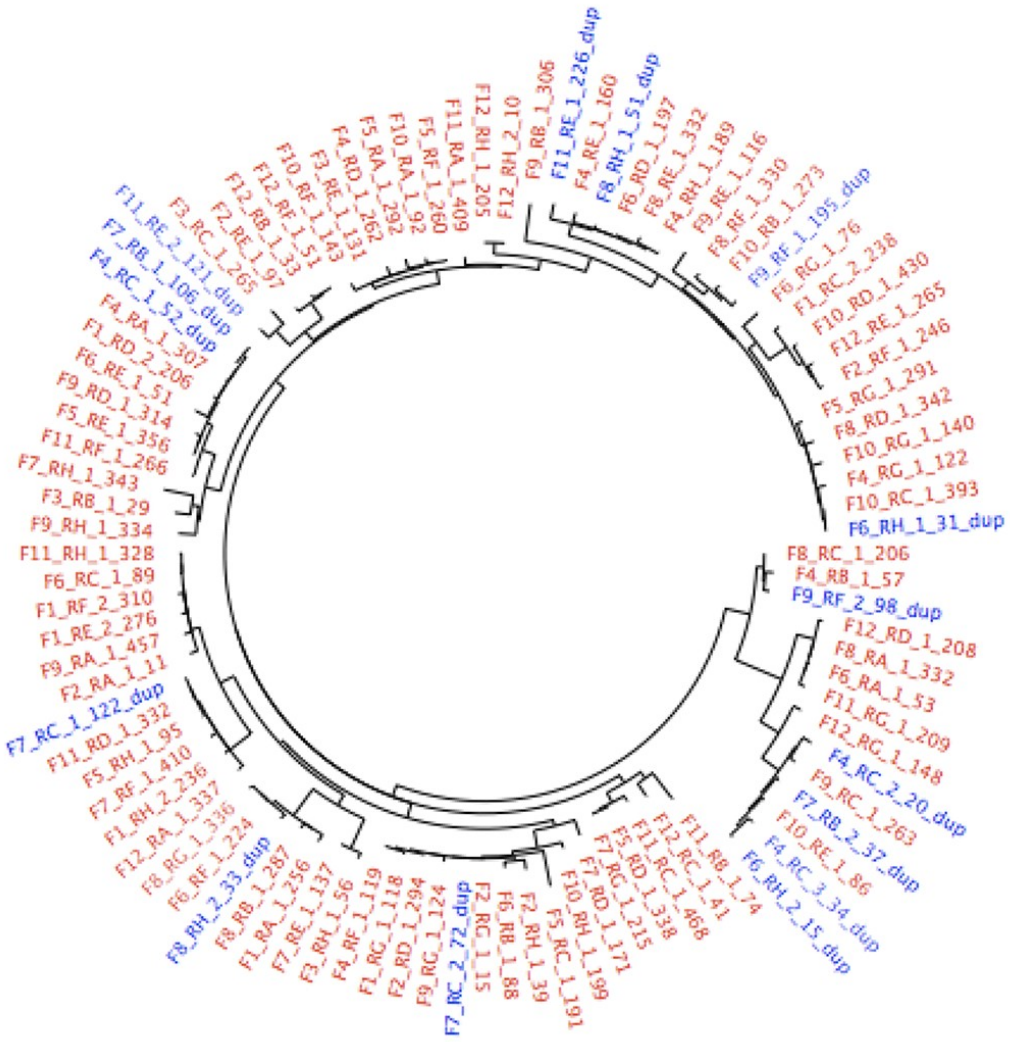

Figure 15. Phylogenetic tree of PC63 V36 *env* in pLentiIII-HA. Next Generation Sequencing results of our pLentiIII-HA plasmid indicate genetic diversity achieved within the cloning library created. Tree construction and sequencing analysis performed by Ben Murrell. This tree was constructed without the use of our Robust Amplicon Denoising technology. Instead, barcoded primers (attached to the library via PCR by Thomas Vollbrecht) were used to distinguish the samples from each other. Sequences in blue result from samples that have multiple templates, as a result of the well the DNA came from being composed of DNA from multiple colonies since the plate density was quite high.

CHAPTER 3: DISCUSSION AND CONCLUSIONS

3.1: DISCUSSION

Optimizations

T4 Ligase

Before we settled on NE Builder as our enzyme kit of choice for our high efficiency cloning, we employed a T4 Ligation strategy according to manufacturer's instructions (New England Biolabs). We moved away from this approach because the negative colony controls were too often equivalent in number to the cloning reaction colonies. In addition, because the T4 protocol requires digestion of the insert, we were cutting some of the *env* variants with the enzymes used to digest the *envs*, because the restriction enzyme was not listed as a cutter in the consensus sequence of our *envs.*

The biggest error that was performed in the workflow of generating env clones was using the incorrect type of insert for a given cloning reaction type. The T4 Ligation reaction requires that both the insert and the backbone be digested with the restriction enzymes of choice for the ligation to complete. The NEBuilder cloning kit requires a digested backbone with an undigested insert. If either insert types get mixed up with the other insert, the reactions will fail with a result of no colony formation upon transformation. This happened once, when we were running the T4 and NEBuilder protocols concurrently.

Primer design variation and Primer overhang structure

The primer design for both projects were quite similar, since the PCR step and the insert are the same for both projects with primers modified based on the backbone. The primers largely carried a structure where the first component of the primer matched the vector at a site that recognized a restriction enzyme cut site. For the forward primers for the Immunogen Fishing cloning, we then incorporated a Kozak sequence of GCCACC to aid with later transcription, followed by the start of the *env* type the primer was designed for (PC39 or PC63). Also for the Immunogen Fishing cloning, we attempted to restrict the primer lengths have overlap with 20 base pairs with the start of env, with the backbone portion of the primer acting as an overhang that will bind to the digested backbone with the aid of the NE Builder enzymes. In the case of our reverse primers, we varied the *env* portion for some of the pLentiIII-HA cloning attempts. Some reverse primers contained *env* complementarity right up to the cytoplasmic tail, and so the result was a shortened *env* of approximately 2.2 kb because the cytoplasmic tail was not amplified. We did this because *env* has been shown to have improved expression with a removed cytoplasmic tail (Stano, *et al.*, 2017). For some other primers, we kept the *env* complementarity of the end of the *env* gene, and thus attempted to insert 2.7 kb. For our cloning into pLentiIII-HA, we found that our more successful cloning attempts used primers that removed the cytoplasmic tail. We did not vary the primer design for attempting to clone into Q23Δenv, and all of the *env* that we were able to insert into Q23Δenv was full *env*, not shortened.

UV Exposure time

For our successful cloning attempts in pLentiIII-HA and Q23Δenv, we gradually switched to minimizing UV exposure to our cloning components. We did this by taking

gel images using a blue light setting on our Axygen Gel Documentation system, which was the primary imaging machine used to take our Agarose gel images. This light was listed as “Blue” but probably still used UV light in order to visualize the gel because the fluorescent molecule used was ethidium bromide.

Addition of rSAP

We originally did not add rSAP to our digested backbones, but we realized that one of the potential problems that our cloning was facing was potential religation of the backbone ends. To prevent religation of the ends, we dephosphorylated the ends with Shrimp Alkaline Phosphatase, using manufacturer-recommended guidelines (New England BioLabs).

Plating

All of the bacterial transformations were plated on the same day that the transformations into Stellar cells were performed. This is to minimize the number of doubling times that occurred during the bacterial growth periods while also preventing premature bacterial cell death from storage in a refrigerator or freezer. This increases the correlation of genetic diversity to colony number. The lower the doubling amounts, the greater the proxy for diversity.

Timepoints

All of the PC39 *envs* and PC63 *envs* come from a study where samples were collected from the donor in question at a specific clinical visit (Landais, *et al.,* 2016). As such, the clinical visit number for a given cDNA sample that we worked with was denoted with a V#. For example, PC63 V36 is *env* DNA that comes from the clinical visit that was intended to be 36 months after first identification of the infection (Landais, *et*

al., 2016). Three different timepoints were used for PC39 cloning workflows: PC39 V24, V36, and V60. The three different timepoints for PC63 were V09, V36, and V60.

Backbone Amount

We varied the amount of backbone in ng used for different cloning attempts, but we maintained the molar ratios for the corresponding cloning reaction types. For both the T4 and the NEBuilder reactions, we tried both 50 ng and 100 ng for the backbone amount. The T4 Ligation manual recommends a 3:1 insert:backbone ratio, while the NEBuilder manual recommends a 2:1 insert:backbone ratio. The calculations performed to determine the amount of ng to use with its corresponding backbone amount utilized the NEB Ligation Calculator. 100 ng was eventually decided as the proper backbone amount for our cloning reactions. The decision was made because the pmol number for the backbone size of pLentiIII-HA lies in the middle of the accepted pmol ranges for the reactions.

Different cell types

Stellar competent cells were not the only cell type used for our transformations. We also experimented with using 5α Competent Cells (New England Biolabs) and One Shot Stbl3 Chemically Competent *E. coli* (Thermo Fisher Scientific). When we needed to scale up our proviral vector Q23Δenv, we transformed 10ng of the vector per Stbl3 cell vial, but for all other transformations we found that Stellar cells were our preferred choice, with improved efficiency over both competitors for our cloning reaction purposes.

Cloning Efficiency and Diversity

A huge contribution to the diversity of *env* is the enzyme responsible for converting HIV single stranded RNA into the DNA that is incorporated into the genome, Reverse Transcriptase (RT). RT is highly error prone and introduces one error for every 10,000 bp (Korber, *et al*., 2001). This error rate contributes to the high number of variants of HIV-1 within one patient.

The purpose of this workflow is to maintain, to the maximum extent possible, the natural diversity of the donor virus throughout the cloning process. As many variants as possible need to be present in the cloned libraries, so that workflows, such as sorting transduced 293T cells via FACS, can potentially observe the potential Env variants that bind to the PC63 bNAb.

The premise of maintaining the natural diversity is simple: The PCR reaction will amplify the already-diverse patient sample. The main challenge with the molecular cloning of patient-derived sample is to have a high efficiency transformation step and a high efficiency cloning step, so that upon further downstream work such as transduction, each cell could possess a unique variant. Some strategies that we implemented to ensure higher efficiency cloning included dephosphorylation of the ends of the backbone vector for both vector sets using rSAP, shrimp alkaline phosphatase, to prevent religation of the DNA ends.

We considered a cloning reaction to be efficient when the experimental plate had more than twice the number of colonies grown on the plate than the negative control plate. It should be noted that this qualification was not to dictate that we had obtained diversity, but just to state that there was some kind of cloning that we felt comfortable proceeding with. We had difficulties achieving this level of efficiency when we were

attempting to complete the cloning reaction using the NEB T4 Ligase Protocol (New England BioLabs) because the number of colonies growing on the negative control plate would be close to equivalent with the colony number on the experimental plate. We hypothesized that this was possibly because the T4 Ligase was cutting the *env* out of the plasmid and ligating the two ends of the pLentiIII-HA plasmid together, resulting in colonies that expressed the Kanamycin Resistance Gene but did not contain *env.* This could be explained by the natural diversity of *env*, in that some *envs* could contain recognition sequences for restriction sites, but not enough *envs* would contain these sequences to appear in a consensus sequence, which was what we used in our decision-making process for selecting restriction enzymes.

Further attempts at transduction of these libraries are needed to assess the issues with cell surface expression of these *envs*. Once this challenge is resolved, future work will attempt to probe transduced HEK293T cells against bNAbs from the PC63 broad antibody lineage to assess which Envs bind (the positively sorted cells). The PC63 Envs that are positively selected will be expected to possess certain motifs currently unknown to us that will appear at a higher frequency compared to the unsorted HEK293T cell population.

The project involving PC63 envelopes cloned into the Q23Δenv plasmid looks more promising. The cloning efficiency was appropriately high, and we were able to produce live virus in 293T cells, which was also to infect Tzm-bl cells and run a successful Luciferase infectivity assay in this cell system (Fig. 10).

The Next Generation Sequencing determined that 80 of the 96 wells submitted for sequencing contained single env genes. The other 16 wells contained multiple *envs* due to

colony selection error during the mini prepping stage. Additionally, in terms of cloning performance, repeated side by side cloning reactions favored V36 over other timepoints. We were able to generate *env* amplicons for the other PC63 and PC39 timepoints that we have in stock, and our gel images showed this, but timepoints that were not V36, such as V24 of V60, were not able to incorporate into a full plasmid. This seems to indicate that the cloning is not robust.

Ultimately, what is the use of generating of generating an HIV-1 *env* library with natural diversity? The sequencing dataset is useful because normally, when donor DNA is sequenced using a Next Generation Sequencing technology like PacBio, sequences are noisy and full of errors (Laird-Smith, *et al.*, 2016). For example, consider a library of 20,000 sequences, and the challenge of determining the noise-free variants within those 20,000 sequences. My supervisor, Ben Murrell, and his team have created software that subtracts the noise and gives back clean sequences (Kumar, *et al.*, 2018). However, it is difficult to prove that this works. In order to take steps to prove that Next Generation Sequencing editing technology is effective, we took the pLentiIII_PC63 plasmid DNA product and performed a PCR with barcodes with ACGT sequences. We were able to evaluate how well our algorithm for editing Next Generation Sequencing is performing (Fig. 15). This is one example of the utility that a natural diverse library such as the ones generated in this project provide.

It is important to note that the dot plot and cluster tree listed in our results are based on 114 templates, but the number of colonies used for the Next Generation Sequencing was 96 (Fig. 13, 14). This was because of the accidental inclusion of the multiple colonies per some of the mini preps performed to obtain DNA for the

sequencing attempt. The plate from which the colonies came from had a high enough density to have colony overlap, which would explain a template number that is higher than the initial well number.

3.2: CONCLUSION

This project indicates that *env* cDNA from donor samples can be cloned into bacteria while maintaining their natural diversity. The downstream applications of the cloning libraries created for this thesis require more time and troubleshooting, but it remains that we have established a suitable protocol for the generation of naturally diverse *env* libraries that will enable more high-throughput experiments involving bNAb interactions with *env*.

WORKS CITED

ABM, inc. *PLenti-III-HA Vector Map | ABM Inc.* https://www.abmgood.com/pLenti-III-HA-Vector-Map.html. Accessed 20 Aug. 2018.

Beddows, S., S. Lister, R. Cheingsong, C. Bruck, J. Weber, James E. Robinson, Julie M. Decker, Yingying Li, Maria G. Salazar, Victoria R. Polonis, Koleka Mlisana, Salim Abdool Karim, Kunxue Hong, Kelli M. Greene, Miroslawa Bilska, Jintao Zhou, Susan Allen, Elwyn Chomba, Joseph Mulenga, Cheswa Vwalika, Feng Gao, Ming Zhang, Bette T. M. Korber, Eric Hunter, Beatrice H. Hahn, and David C. Montefiori. "Comparison of the Antibody Repertoire Generated in Healthy Volunteers Following Immunization with a Monomeric Recombinant Gp120 Construct Derived from a CCR5/CXCR4-Using Human Immunodeficiency Virus Type 1 Isolate with Sera from Naturally Infected Individuals." *Journal of Virology*, vol. 73, no. 2, American Society for Microbiology Journals, Feb. 1999, pp. 1740–45, doi:10.1128/jvi.01730-06.

Bonsignori, Mattia, Tongqing Zhou, Zizhang Sheng, Lei Chen, Feng Gao, M. Gordon Joyce, Gabriel Ozorowski, Gwo-Yu Chuang, Chaim A. Schramm, Kevin Wiehe, S. Munir Alam, Todd Bradley, Morgan A. Gladden, Kwan-Ki Hwang, Sheelah Iyengar Amit Kumar, Xiaozhi Lu, Kan Luo, Michael C. Mangiapani, Robert J. Parks, Hongshuo Song, Priyamvada Acharya, Robert T. Bailer, Allen Cao, Aliaksandr Druz, Ivelin S. Georgiev, Young D. Kwon, Mark K. Louder, Baoshan Zhang, Anqi Zheng, Brenna J. Hill, Rui Kong, Cinque Soto, NISC Comparative Sequencing Program, James C. Mullikin, Daniel C. Douek, David C. Montefiori, Michael A. Moody, George M. Shaw, Beatrice H. Hahn, Garnett Kelsoe, Peter T. Hraber, Bette T. Korber, Scott D. Boyd, Andrew Z. Fire, Thomas B. Kepler, Lawrence Shapiro, Andrew B. Ward, John R. Mascola, Hua-Xin Liao, Peter D. Kwong, and Barton F. Haynes. "Maturation Pathway from Germline to Broad HIV-1 Neutralizer of a CD4-Mimic Antibody." *Cell*, vol. 165, no. 2, NIH Public Access, Apr. 2016, pp. 449–63, doi:10.1016/j.cell.2016.02.022.

Burton, Dennis R., Rafi Ahmed, Dan H. Barouch, Salvatore T. Butera, Shane Crotty, Adam Godzik, Daniel E. Kaufmann, M. Juliana McElrath, Michel C. Nussenzweig, Bali Pulendran, Chris N. Scanlan, William R. Schief, Guido Silvestri, Hendrik Streeck, Bruce D. Walker, Laura M. Walker, Andrew B. Ward, Ian A. Wilson, and Richard Wyatt. "A Blueprint for HIV Vaccine Discovery." *Cell Host & Microbe*, vol. 12, no. 4, Elsevier, Oct. 2012, pp. 396–407, doi:10.1016/j.chom.2012.09.008.

Conley, A. J., J. A. Kessler, L. J. Boots, J. S. Tung, B. A. Arnold, P. M. Keller, A. R. Shaw, E. A. Emini, Khoa M. Le, Peter S. Lee, Yuanzi Hua, Michael S. Seaman, John P. Moore, Andrew B. Ward, Ian A. Wilson, Rogier W. Sanders, and Dennis R. Burton. "Neutralization of Divergent Human Immunodeficiency Virus Type 1 Variants and Primary Isolates by IAM-41-2F5, an Anti-Gp41 Human Monoclonal Antibody." *Proceedings of the National Academy of Sciences*, vol. 91, no. 8,

National Academy of Sciences, Apr. 1994, pp. 3348–52, doi:10.1073/pnas.91.8.3348.

Corning Life Sciences. *Axygen® Gel Documentation System | Gel Documentation Systems | Molecular Biology Equipment | Equipment | Life Sciences United States Consumer Site | Corning*. https://ecatalog.corning.com/life-sciences/b2c/US/en/Equipment/Molecular-Biology-Equipment/Gel-Documentation-Systems/Axygen®-Gel-Documentation-System/p/GDBL-1000?lang=en. Accessed 19 Aug. 2018.

Dingens, Adam S., Hugh K. Haddox, Julie Overbaugh, and Jesse D. Bloom. "Comprehensive Mapping of HIV-1 Escape from a Broadly Neutralizing Antibody." *Cell Host & Microbe*, vol. 21, no. 6, NIH Public Access, June 2017, p. 777–787.e4, doi:10.1016/j.chom.2017.05.003.

DNA MODIFYING ENZYMES Instruction Manual NEBuilder ® HiFi DNA Assembly Master Mix/NEBuilder HiFi DNA Assembly Cloning Kit. 2017, https://www.neb.com/-/media/catalog/datacards-or-manuals/manuale2621.pdf.

Doria-Rose, Nicole A., Rachel M. Klein, Marcus G. Daniels, Sijy O'Dell, Martha Nason, Alan Lapedes, Tanmoy Bhattacharya, Stephen A. Migueles, Richard T. Wyatt, Bette T. Korber, John R. Mascola, and Mark Connors. "Breadth of Human Immunodeficiency Virus-Specific Neutralizing Activity in Sera: Clustering Analysis and Association with Clinical Variables." *Journal of Virology*, vol. 84, no. 3, American Society for Microbiology Journals, Feb. 2010, pp. 1631–36, doi:10.1128/JVI.01482-09.

Falkowska, Emilia, Alejandra Ramos, Yu Feng, Tongqing Zhou, Stephanie Moquin, Laura M. Walker, Xueling Wu, Michael S. Seaman, Terri Wrin, Peter D. Kwong, Richard T. Wyatt, John R. Mascola, Pascal Poignard, and Dennis R. Burton. "PGV04, an HIV-1 Gp120 CD4 Binding Site Antibody, Is Broad and Potent in Neutralization but Does Not Induce Conformational Changes Characteristic of CD4." *Journal of Virology*, vol. 86, no. 8, American Society for Microbiology Journals, Apr. 2012, pp. 4394–403, doi:10.1128/JVI.06973-11.

Gao, F., S. G. Morrison, D. L. Robertson, C. L. Thornton, S. Craig, G. Karlsson, J. Sodroski, M. Morgado, B. Galvao-Castro, H. von Briesen, S. Beddows, J. Weber, P. M. Sharp, G. M. Shaw, and B. H. Hahn. "Molecular Cloning and Analysis of Functional Envelope Genes from Human Immunodeficiency Virus Type 1 Sequence Subtypes A through G. The WHO and NIAID Networks for HIV Isolation and Characterization." *Journal of Virology*, vol. 70, no. 3, American Society for Microbiology Journals, Mar. 1996, pp. 1651–67, http://www.ncbi.nlm.nih.gov/pubmed/8627686.

GENEWIZ. *Sample Preparation - Sample Submission Guidelines - Resources -*

GENEWIZ. https://www.genewiz.com/Public/Resources/Sample-Submission-Guidelines/Sanger-Sequencing-Sample-Submission-Guidelines/Sample-Preparation#sanger-sequence. Accessed 19 Aug. 2018.

Gift, Syna Kuriakose, Daniel P. Leaman, Lei Zhang, Arthur S. Kim, and Michael B. Zwick. "Functional Stability of HIV-1 Envelope Trimer Affects Accessibility to Broadly Neutralizing Antibodies at Its Apex." *Journal of Virology*, vol. 91, no. 24, American Society for Microbiology Journals, Dec. 2017, pp. e01216-17, doi:10.1128/JVI.01216-17.

Haddox, Hugh K., Adam S. Dingens, Sarah K. Hilton, Julie Overbaugh, and Jesse D. Bloom. "Mapping Mutational Effects along the Evolutionary Landscape of HIV Envelope." *ELife*, vol. 7, Mar. 2018, doi:10.7554/eLife.34420.

Haynes, Barton F. and David C. Montefiori. "Aiming to Induce Broadly Reactive Neutralizing Antibody Responses with HIV-1 Vaccine Candidates." *Expert Review of Vaccines*, vol. 5, no. 3, NIH Public Access, June 2006, pp. 347–63, doi:10.1586/14760584.5.3.347.

Korber, Bette, Brian Gaschen, Karina Yusim, Rama Thakallapally, Can Kesmir, and Vincent Detours. "Evolutionary and Immunological Implications of Contemporary HIV-1 Variation." *British Medical Bulletin*, vol. 58, no. 1, Oxford University Press, Sept. 2001, pp. 19–42, doi:10.1093/bmb/58.1.19.

Kumar, Venkatesh, Thomas Vollbrecht, Mark Chernyshev, Sanjay Mohan, Brian Hanst, Nicholas Bavafa, Antonia Lorenzo, Robert Ketteringham, Kemal Eren, Michael Golden, Michelli Faria Oliveira, and Ben Murrell. "Long-Read Amplicon Denoising." *BioRxiv*, Cold Spring Harbor Laboratory, Aug. 2018, p. 383794, doi:10.1101/383794.

Kwong, Peter D., Michael L. Doyle, David J. Casper, Claudia Cicala, Stephanie A. Leavitt, Shahzad Majeed, Tavis D. Steenbeke, Miro Venturi, Irwin Chaiken, Michael Fung, Hermann Katinger, Paul W. I. H. Parren, James Robinson, Donald Van Ryk, Liping Wang, Dennis R. Burton, Ernesto Freire, Richard Wyatt, Joseph Sodroski, Wayne A. Hendrickson, and James Arthos. "HIV-1 Evades Antibody-Mediated Neutralization through Conformational Masking of Receptor-Binding Sites." *Nature*, vol. 420, no. 6916, Nature Publishing Group, Dec. 2002, pp. 678–82, doi:10.1038/nature01188.

Kwong, Peter D. and John R. Mascola. "Human Antibodies That Neutralize HIV-1: Identification, Structures, and B Cell Ontogenies." *Immunity*, vol. 37, no. 3, Cell Press, Sept. 2012, pp. 412–25, doi:10.1016/J.IMMUNI.2012.08.012.

Laird Smith, Melissa, Ben Murrell, Kemal Eren, Caroline Ignacio, Elise Landais, Steven Weaver, Pham Phung, Colleen Ludka, Lance Hepler, Gemma Caballero, Tristan

Pollner, Yan Guo, Douglas Richman, Pascal Poignard, Ellen E. Paxinos, Sergei L. Kosakovsky Pond, and Davey M. Smith. “Rapid Sequencing of Complete *Env* Genes from Primary HIV-1 Samples.” *Virus Evolution*, vol. 2, no. 2, Oxford University Press, July 2016, p. vew018, doi:10.1093/ve/vew018.

Landais, Elise, Xiayu Huang, Colin Havenar-Daughton, Ben Murrell, Matt A. Price, Lalinda Wickramasinghe, Alejandra Ramos, Charoan B. Bian, Melissa Simek, Susan Allen, Etienne Karita, William Kilembe, Shabir Lakhi, Mubiana Inambao, Anatoli Kamali, Eduard J. Sanders, Omu Anzala, Vinodh Edward, Linda-Gail Bekker, Jianming Tang, Jill Gilmour, Sergei L. Kosakovsky-Pond, Pham Phung, Terri Wrin, Shane Crotty, Adam Godzik, and Pascal Poignard. “Broadly Neutralizing Antibody Responses in a Large Longitudinal Sub-Saharan HIV Primary Infection Cohort.” *PLOS Pathogens*, edited by Alexandra Trkola, vol. 12, no. 1, Public Library of Science, Jan. 2016, p. e1005369, doi:10.1371/journal.ppat.1005369.

Landais, Elise, Ben Murrell, Bryan Briney, Sasha Murrell, Kimmo Rantalainen, Zachary T. Berndsen, Alejandra Ramos, Lalinda Wickramasinghe, Melissa Laird Smith, Kemal Eren, Natalia de Val, Mengyu Wu, Audrey Cappelletti, Jeffrey Umotoy, Yolanda Lie, Terri Wrin, Paul Algate, Po-Ying Chan-Hui, Etienne Karita, Andrew B. Ward, Ian A. Wilson, Dennis R. Burton, Davey Smith, Sergei L. Kosakovsky Pond, and Pascal Poignard. “HIV Envelope Glycoform Heterogeneity and Localized Diversity Govern the Initiation and Maturation of a V2 Apex Broadly Neutralizing Antibody Lineage.” *Immunity*, vol. 47, no. 5, Cell Press, Nov. 2017, p. 990–1003.e9, doi:10.1016/J.IMMUNI.2017.11.002.

Li, Ming, Feng Gao, John R. Mascola, Leonidas Stamatatos, Victoria R. Polonis, Marguerite Koutsoukos, Gerald Voss, Paul Goepfert, Peter Gilbert, Kelli M. Greene, Miroslawa Bilska, Denise L. Kothe, Jesus F. Salazar-Gonzalez, Xiping Wei, Julie M. Decker, Beatrice H. Hahn, and David C. Montefiori. “Human Immunodeficiency Virus Type 1 Env Clones from Acute and Early Subtype B Infections for Standardized Assessments of Vaccine-Elicited Neutralizing Antibodies.” *Journal of Virology*, vol. 79, no. 16, American Society for Microbiology Journals, Aug. 2005, pp. 10108–25, doi:10.1128/JVI.79.16.10108-10125.2005.

Lipofectamine 3000 Reagent - US. https://www.thermofisher.com/us/en/home/brands/product-brand/lipofectamine/lipofectamine-3000.html?gclid=Cj0KCQjwquTbBRCSARIsADzW88xMtSW7cbkaogbA-UZfG8GqDT_D63IGNgRFnRHSYCWYDc75jDAyKpMaAgf7EALw_wcB&s_kwcid=AL!3652!3!265279845573!e!!g!!lipofectamine 3000 protocol&ef_id=WxmwLwAADUDm6-yQ:20180820051411:s. Accessed 19 Aug. 2018.

Martinez-Navio, José M. and Ronald C. Desrosiers. “Neutralizing Capacity of

Monoclonal Antibodies That Recognize Peptide Sequences Underlying the Carbohydrates on Gp41 of Simian Immunodeficiency Virus." *Journal of Virology*, vol. 86, no. 23, American Society for Microbiology Journals, Dec. 2012, pp. 12484–93, doi:10.1128/JVI.01959-12.

Miyauchi, Kosuke, Yuri Kim, Olga Latinovic, Vladimir Morozov, and Gregory B. Melikyan. "HIV Enters Cells via Endocytosis and Dynamin-Dependent Fusion with Endosomes." *Cell*, vol. 137, no. 3, Cell Press, May 2009, pp. 433–44, doi:10.1016/J.CELL.2009.02.046.

Moir, Susan, Tae-Wook Chun, and Anthony S. Fauci. "Pathogenic Mechanisms of HIV Disease." *Annual Review of Pathology: Mechanisms of Disease*, vol. 6, no. 1, Annual Reviews , Feb. 2011, pp. 223–48, doi:10.1146/annurev-pathol-011110-130254.

New England BioLabs. *Ligation Protocol with T4 DNA Ligase (M0202) | NEB*. https://www.neb.com/protocols/1/01/01/dna-ligation-with-t4-dna-ligase-m0202. Accessed 19 Aug. 2018.

---. *NEB 5-Alpha Competent E. Coli (High Efficiency)*. https://www.neb.com/-/media/catalog/datacards-or-manuals/c2987datasheet-lot2831402.pdf. Accessed 19 Aug. 2018.

Saitou, N. and M. Nei. "The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees." *Molecular Biology and Evolution*, vol. 4, no. 4, Oxford University Press, July 1987, pp. 406–25, doi:10.1093/oxfordjournals.molbev.a040454.

Seaman, Michael S., Holly Janes, Natalie Hawkins, Lauren E. Grandpre, Colleen Devoy, Ayush Giri, Rory T. Coffey, Linda Harris, Blake Wood, Marcus G. Daniels, Tanmoy Bhattacharya, Alan Lapedes, Victoria R. Polonis, Francine E. McCutchan, Peter B. Gilbert, Steve G. Self, Bette T. Korber, David C. Montefiori, and John R. Mascola. "Tiered Categorization of a Diverse Panel of HIV-1 Env Pseudoviruses for Assessment of Neutralizing Antibodies." *Journal of Virology*, vol. 84, no. 3, American Society for Microbiology Journals, Feb. 2010, pp. 1439–52, doi:10.1128/JVI.02108-09.

Siliciano, Robert F. and Warner C. Greene. "HIV Latency." *Cold Spring Harbor Perspectives in Medicine*, vol. 1, no. 1, Cold Spring Harbor Laboratory Press, Sept. 2011, p. a007096, doi:10.1101/cshperspect.a007096.

Simek, Melissa D., Wasima Rida, Frances H. Priddy, Pham Pung, Emily Carrow, Dagna S. Laufer, Jennifer K. Lehrman, Mark Boaz, Tony Tarragona-Fiol, George Miiro, Josephine Birungi, Anton Pozniak, Dale A. McPhee, Olivier Manigart, Etienne Karita, André Inwoley, Walter Jaoko, Jack Dehovitz, Linda-Gail Bekker, Punnee Pitisuttithum, Robert Paris, Laura M. Walker, Pascal Poignard, Terri Wrin, Patricia

E. Fast, Dennis R. Burton, and Wayne C. Koff. “Human Immunodeficiency Virus Type 1 Elite Neutralizers: Individuals with Broad and Potent Neutralizing Activity Identified by Using a High-Throughput Neutralization Assay Together with an Analytical Selection Algorithm.” *Journal of Virology*, vol. 83, no. 14, American Society for Microbiology Journals, July 2009, pp. 7337–48, doi:10.1128/JVI.00110-09.

Sok, Devin, Marit J. van Gils, Matthias Pauthner, Jean-Philippe Julien, Karen L. Saye-Francisco, Jessica Hsueh, Bryan Briney, Jeong Hyun Lee, Khoa M. Le, Peter S. Lee, Yuanzi Hua, Michael S. Seaman, John P. Moore, Andrew B. Ward, Ian A. Wilson, Rogier W. Sanders, and Dennis R. Burton. “Recombinant HIV Envelope Trimer Selects for Quaternary-Dependent Antibodies Targeting the Trimer Apex.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 49, National Academy of Sciences, Dec. 2014, pp. 17624–29, doi:10.1073/pnas.1415789111.

Stano, Armando, Daniel P. Leaman, Arthur S. Kim, Lei Zhang, Ludovic Autin, Jidnyasa Ingale, Syna K. Gift, Jared Truong, Richard T. Wyatt, Arthur J. Olson, and Michael B. Zwick. “Dense Array of Spikes on HIV-1 Virion Particles.” *Journal of Virology*, vol. 91, no. 14, American Society for Microbiology Journals, Apr. 2017, p. JVI.00415-17, doi:10.1128/JVI.00415-17.

Takara. *Advantage 2 PCR Kit*. https://www.takarabio.com/products/pcr/high-yield-pcr/advantage-2-products/advantage-2-pcr-kit. Accessed 19 Aug. 2018.

Thermo Fisher Scientific. *NanoDrop 1000 Spectrophotometer V3.8 User’s Manual*. 2010, www.nanodrop.com.

Thermo Fisher Scientific. *One Shot Stbl3 Chemically Competent E. Coli - Thermo Fisher Scientific*. https://www.thermofisher.com/order/catalog/product/C737303?gclid=Cj0KCQjwquTbBRCSARIsADzW88xry-juzmEI8N_P9llIpEjCzM7R3LN0ue6CvC-XxvED64rubPWUu-MaAmBUEALw_wcB&s_kwcid=AL!3652!3!283061444533!b!!g!!&ef_id=WxmwLwAADUDm6-yQ:20180820052024:s. Accessed 19 Aug. 2018.

Walker, Laura M., Michael Huber, Katie J. Doores, Emilia Falkowska, Robert Pejchal, Jean-Philippe Julien, Sheng-Kai Wang, Alejandra Ramos, Po-Ying Chan-Hui, Matthew Moyle, Jennifer L. Mitcham, Phillip W. Hammond, Ole A. Olsen, Pham Phung, Steven Fling, Chi-Huey Wong, Sanjay Phogat, Terri Wrin, Melissa D. Simek, Protocol G. Principal Protocol G Principal Investigators, Wayne C. Koff, Ian A. Wilson, Dennis R. Burton, and Pascal Poignard. “Broad Neutralization Coverage of HIV by Multiple Highly Potent Antibodies.” *Nature*, vol. 477, no. 7365, NIH Public Access, Sept. 2011, pp. 466–70, doi:10.1038/nature10373.

World Health Organization. “WHO | HIV/AIDS.” *WHO*, World Health Organization, 2018, http://www.who.int/gho/hiv/en/.

Wyatt, R. and J. Sodroski. “The HIV-1 Envelope Glycoproteins: Fusogens, Antigens, and Immunogens.” *Science (New York, N.Y.)*, vol. 280, no. 5371, American Association for the Advancement of Science, June 1998, pp. 1884–88, doi:10.1126/SCIENCE.280.5371.1884.

ZYMO RESEARCH CORP. Zymoclean™ Gel DNA Recovery Kit Catalog Nos. D4001T, D4001, D4002, D4007 & D4008. https://d2gsy6rsbfrvyb.cloudfront.net/media/amasty/amfile/attach/_D4001T_D4001_D4002_D4007_D4008_Zymoclean_Gel_DNA_Recovery_Kit_ver_1_2_1_LKN-SW_.pdf. Accessed 19 Aug. 2018.