DNA PLASMID PURIFICATION PROCESS OPTIMIZATION
AT LIFE TECHNOLOGIES CORPORATION

A Thesis
presented to
The Faculty of California Polytechnic State University,
San Luis Obispo

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in Biomedical Engineering

by
Trevor Shepherd
June 2013

© 2013

Trevor Shepherd

ALL RIGHTS RESERVED

COMMITTEE MEMBERSHIP

TITLE: DNA PLASMID PURIFICATION PROCESS
OPTIMIZATION AT LIFE TECHNOLOGIES
CORPORATION

AUTHOR: Trevor Shepherd

DATE SUBMITTED: April, 2013

COMMITTEE CHAIR: Dr. Lanny Griffin, Biomedical Engineering Department
Chair

COMMITTEE MEMBER: Dr. Trevor Cardinal, Associate Professor of
Biomedical Engineering

COMMITTEE MEMBER: Dr. Mark Cochran, Manufacturing Scientist at Life
Technologies

ABSTRACT

DNA PLASMID PURIFICATION PROCESS OPTIMIZATION

AT LIFE TECHNOLOGIES CORPORATION

Trevor Shepherd

This project focused on optimizing the plasmid DNA purification process at Life Technologies. These plasmids are designed to code for specialized proteins used by research universities, national laboratories, or research companies. Once cultivated and harvested, the plasmids must be analyzed for quality and quantity. The project is divided into improving three aspects of the process: 1) plasmid identification, 2) plasmid purity evaluation, and 3) process yield. Plasmid identification is now simpler, more robust and has zero ambiguity. Plasmid purity evaluation is now measured with computer software, which reduces user error and eliminates subjectivity. Using the nascent metrics provided by the improved identification and purity evaluation techniques, process yield was analyzed and improved. The hypotheses on yield improvement and the information gleaned from their resulting experiments provide a foundation for further process improvement.

ACKNOWLEDGEMENTS

I would like to thank Mark Cochran for his deep knowledge of the science and steady patience throughout my time at Life Technologies and while I was writing my thesis. Also, I would like to thank Trevor Cardinal and Lanny Griffin for being part of my thesis committee and providing guidance throughout this project. Finally, I would like to thank Life Technologies for providing the opportunity to work with them.

TABLE OF CONTENTS

	Page
LIST OF TABLES.....	viii
LIST OF FIGURES.....	ix
 CHAPTER	
I. Introduction.....	1
Life Technologies Synopsis.....	1
Project Overview.....	2
Technical Background.....	3
Shortcomings of the DNA Purification Process.....	19
II. Objectives and Approach.....	20
Improving the Identification Protocol.....	20
Improving the Purity Analysis Protocol.....	21
Improving the Yield of the Purification Process.....	22
III. Methods and Materials.....	23
Improving the Identification Protocol.....	23
Improving the Purity Analysis Protocol.....	31
Improving the Yield of the Purification Process.....	39
IV. Results and Discussion.....	42

Identification Assay.....	42
Purity Assay.....	45
Yield Assay.....	51
V. Conclusion and Future Works.....	58
Identification Assay.....	58
Purity Assay.....	58
Yield Assay.....	59
Bibliography.....	60

LIST OF TABLES

Table	Page
1. Raw Density Data From VisionWorksLS Software.....	38
2. MegaPrep/GigaPrep Specifications.....	40
3. Purity Calculations.....	46
4. Plasmid Characterization.....	48
5. GigaPrep Purity and Yield Assay Results.....	52
6. Maxicard Purity and Yield Assay Results.....	55

LIST OF FIGURES

Figure	Page
1. Storage Freezer.....	4
2. Individual “Glycerol Stock” On Ice.....	4
3. Plasmid Map.....	5
4. Diagram of Transformation.....	6
5. Plasmid Transformation Station.....	7
6. Illustration of Antibiotic Resistance Selection.....	8
7. Petri Dish with Transformed Bacteria Colonies.....	9
8. New Brunswick C-24 Incubator and Shaker.....	10
9. Large Scale Innoculation.....	11
10. Sorvall RC3B Centrifuge.....	12
11. Bacteria Pellet.....	12
12. Filtration and Purification Station.....	13
13. DNA Pellet.....	14
14. Nanodrop Spectrophotometer.....	15
15. Restriction Enzyme Reactions.....	16
16. Electrophoresis Gel Box.....	17
17. DNA Electrophoresis Gel.....	18
18. Previous Electrophoresis Gel.....	22

19. Raw Plasmid Sequence.....	24
20. Vector NTI Pictorial of Plasmid.....	25
21. Life Technologies Product Manual Sequence Description.....	26
22. Molecular Structure Illustration.....	27
23. Typical Sequencing Read-out.....	28
24. ID Sequence Lay-out.....	29
25. ContigExpress.....	30
26. E-Gel with “Smiley Faces”.....	31
27. Hand-Poured Gel.....	32
28. 2UV Transilluminator.....	33
29. Density Box.....	33
30. VisionWorksLS Imaging Software Interface.....	34
31. Electrophoresis Banding Pattern.....	35
32. Electron Micrograph Image of Bacterial DNA.....	36
33. Defined Density Boxes.....	37
34. Filtration Cartridge.....	39
35. Maxicard Purification Machine.....	41
36. Sequence Data.....	42
37. SeqID File.....	43
38. ContigExpress Identification Evaluation.....	44
39. Purity Evaluation Verification Gel.....	45
40. Pixel Density vs Volume.....	47
41. Genomic Contamination vs Plasmid Size (bp).....	49

42. Genomic Contamination vs Purification Scale.....	50
43. MegaPrep Purity vs Bacteria Quantity.....	53
44. MegaPrep Purity vs Efficiency.....	54
45. Maxicard Purity vs Efficiency.....	56
46. Maxicard Purity vs Plasmid Size.....	56

I. Introduction

Life Technologies Synopsis

Life Technologies is a life science tools company which employs approximately 11,000 people, holds approximately 4,000 patents, and markets over 50,000 products. The company's list of products can be divided into two basic categories: consumables and diagnostics.

The consumable products are specifically designed DNA, RNA and proteins. These are typically used by research universities, national laboratories, or research companies as the basic materials for investigating diseases, pharmacology, tissue engineering and genetic engineering. The diagnostic products are the tools used to manipulate and evaluate utilization of the consumable products. Some of Life Technologies' most notable diagnostic products are: Ion Torrent®, which utilizes a state-of-the-art technology to rapidly sequence genetic code; Dynabeads®, which uses magnetic and surface properties to separate cells; and Purelink® Separation Products, which automate the multi-step processes of isolating DNA or RNA.

The size and scope of Life Technologies evolved through numerous mergers and acquisitions over the last 30+ years, resulting in a strategic amalgamation of both complimentary and competing companies. While the succession of coalescing companies has been successful, each nascent product acquisition requires a re-worked production process to efficiently incorporate it into the aggregate.

Project Overview

This project focuses on optimizing Life Technologies' in-house DNA Plasmid Manufacturing Process. A plasmid is a small circle of DNA which can replicate independently of the host cell chromosome. Typically, this DNA has been ordered by a customer for the production of a desired recombinant protein. These recombinant proteins have been molecularly cloned from their natural source, typically bacteria, and engineered for specific expression systems.

When researching proteins, it is extremely useful to modify them so they are easily identified and separated. These strategies are effected by designing DNA into the plasmids which code for resistance to an antibiotic, and/or for a particular tag.

A simple yet effective way of separating a desired host-cell type from an undesired host-cell type is to destroy the undesired cells with an antibiotic. Implementation of this strategy involves proliferation in a medium containing an antibiotic. Any cell not containing the DNA that codes for protection against the antibiotic will be destroyed.

DNA designed to code for particular tags allow separation/identification at the protein level. Protein tags are genetically grafted onto the recombinant protein of interest. Two of the most common types of tags are epitope and fluorescence; these tags create high antibody affinity and enable fluorescent illumination, respectively.

Once a plasmid is designed, it must be proliferated in a cell culture and purified.

The basic outline of the DNA purification process follows these steps: 1) obtain the desired plasmid; 2) transform the plasmid into a specialized strain of *E. coli* bacteria; 3) proliferate the transformed bacteria; 4) harvest the desired DNA; and 5) analyze and verify the DNA identity, purity and concentration. Steps 4 and 5 needed the most improvement and are where the majority of the project was focused.

Technical Background

Step 1: Obtain Plasmid

Life Technologies categorically stores hundreds of plasmids, recombinant DNA designed to produce specific complimentary combinations of proteins. These stored plasmids are referred to as “Glycerol Stock”, and are stored in a mixture of glycerol and TE Buffer (TE) at -80°C. The mixture safely stores the plasmids by maintaining the appropriate pH of approximately 8.0, chelating the metal ions that activate degrading nucleases, and it does not crystallize at this temperature which greatly reduces shear stress on the DNA.

Figure 1. Storage Freezer. Plasmids contained in mini-tubes and stored in categorically organized boxes in a -80°C freezer.

Figure 2. Individual “Glycerol Stock” On Ice. These Mini-tubes contain the stored plasmid suspended in glycerol and TE buffer. The DNA is kept cold to reduce molecular interaction which prevents degradation.

Figure 3. Plasmid Map. Example of Life Technologies' visual depiction of a plasmid.

The map shows the regions where specified proteins are coded, the number of base pairs, and the restriction enzyme recognition sites.

Step 2: Plasmid Transformation

Once the plasmid has been selected, it must be transformed into a cell for proliferation and then harvested from the bacteria culture. The bacteria cell wall

is the biggest barrier to DNA uptake. To pass the cell wall, a modified non-infectious strain of *E. coli* bacteria is mixed with the plasmid in saline solution. The mixture is then heated to 42°C, causing the wall of the cell to become porous enough for the plasmid to enter. Once transformed, the plasmid can be multiplied within the cell independently of the bacteria genome because it contains its own origin of replication.

Figure 4. Diagram of Transformation. This illustration depicts the incorporation of a recombinant bacteria plasmid into a host cell.

Figure 5. Plasmid Transformation Station. This is the warm water bath (42°C) where transformation occurs. Hosts are briefly “shocked” by the warm water allowing the plasmid to enter the cell. Once incorporated, the cell has the molecular framework to facilitate proliferation of the plasmid in addition to its own genomic DNA.

Step 3: Proliferating the Transformed Bacteria

The bacteria transformed with the plasmid of interest are then smeared onto an agar plate in a petri dish and incubated for 12 hours. The agar provides a suitable growth environment and contains the appropriate antibiotic. After incubation, individual colonies appear on the agar where a successfully transfected bacteria has proliferated. The antibiotic destroys all cells which have not incorporated the plasmid of interest and its constituent antibiotic resistance gene.

Figure 6. Illustration of Antibiotic Resistance Selection. This is a very effective way of easily distinguishing between successfully and unsuccessfully transformed bacteria cells.

A sample from a successfully transformed bacteria colony is taken and used to progenerate a larger quantity of the plasmid of interest.

Figure 7. Petri Dish with Transformed Bacteria Colonies.

Next, the transformed bacteria are transferred into a test tube with growth media and antibiotic, and then swirled for 12 hours. The combination of growth media containing amino acids and sugars, and a continuous swirl provides the necessary building blocks, energy, and shear stimulation for proliferation.

Figure 8. New Brunswick C-24 Incubator and Shaker. This machine contributes to an ideal growth environment by maintaining the proper temperature and shaking the samples.

The bacteria culture is then scaled-up one more time by transferring the test tube contents into 1L Erlenmeyer flasks and then proliferated in a larger ideal growth environment.

Figure 9. Large Scale Inoculation. The Innova 4 provides a very similar environment as the New Brunswick C-24, but it is designed to shake 1L flasks instead of test tubes.

Step 4: Harvesting DNA

Once the transformed bacteria are proliferated to an adequate level, appropriate steps must be taken for safe and effective separation of the desired plasmid from the rest of the cell. First, the bacteria are separated from the growth media by 30 minutes of high speed centrifugation with a radial acceleration of 4,500 times that of gravity.

Figure 10. Sorvall RC3B Centrifuge. This centrifuge is used to spin down the bacteria and, later in the process, the purified DNA.

Figure 11. Bacteria Pellet. Pellets of bacteria stuck to the side of the container and growth media. The resulting pellet of bacteria is separated by carefully pouring off the growth media.

Next, the pellet is suspended in a solution of lysis buffer and ribonucleases which destroys the cell wall and degrades the newly exposed RNA, respectively. After a few minutes of gentle agitation, the lysate and ribonucleases are deactivated with a neutralization buffer. The buffer also precipitates the cell debris, proteins,

genomic DNA and lysate. This mixture is first pulled through a filtration cartridge which removes the precipitated material. The residual aqueous solution of dissolved RNA and DNA is then pulled through a binding cartridge.

Figure 12. Filtration and Purification Station. The solution is poured in the top container of the apparatus and a vacuum is attached to the bottom canister to facilitate flow through the filtration or binding cartridge. This setup is used for the cell debris filtering cartridge and the RNA and DNA binding cartridge.

The binding cartridge is composed of a positively charged resin which binds the negatively charged DNA and RNA molecules. Once the solution has passed through the cartridge and the DNA and RNA has been bound, a salty wash buffer is run through the cartridge twice, disrupting the electrical attachment of the nucleotides. The first wash of salt solution disrupts the loosely bound RNA which is captured in a beaker and discarded. The second wash removes the more

tightly bound DNA molecules from the resin and is captured in another beaker. At this point, the plasmid DNA is relatively free from RNA and protein contaminants; however, one more step is taken for further purification. The plasmid DNA is then precipitated with isopropanol, washed with 80% ethanol and centrifuged into a pellet. Finally, the DNA pellet is suspended in TE buffer which maintains an ideal pH and protects against degradation.

Figure 13. DNA Pellet. Centrifugation of the suspended DNA packs the DNA into a pellet enabling easy separation from the media.

Step 5: Analyzing and Verifying the DNA Identity, Purity and Concentration

Once the plasmid DNA has been purified, a small sample is taken to evaluate the amount of DNA yielded by the process. This is accomplished using a spectrophotometer. The spectrophotometer evaluates concentration of DNA based on the amount of light absorbed by the sample. Light is emitted at 260nm, which is the wavelength DNA absorbs, and the instrument calculates a

concentration based on the amount of light that passes through (not absorbed) the sample relative to a blank sample.

Figure 14. Nanodrop Spectrophotometer. 1-2 μL of purified plasmid is pipetted onto the Nanodrop to evaluate the DNA concentration.

Once the DNA concentration has been determined, it must be evaluated for purity and positively identified. To accomplish this, the purified DNA is digested by a set of restriction enzymes and then electrophoresed to separate the different pieces of DNA molecules based on size and shape.

A restriction enzyme, a particular type of endonuclease, cuts the backbone by catalyzing the hydrolysis of the phosphodiester bonds at specific sequences of

nucleotides. These specific sequences are known as recognition sites, most of which are palindromic and six nucleotides long.

Figure 15. Restriction Enzyme Reaction. The breaking of the phosphodiester bonds of the DNA backbone is a result of a nucleophilic attack by the enzyme complex on the electrophilic phosphorous atom.

Knowing the sequence of a purified plasmid and the recognition sites for many restriction enzymes allows the plasmid to be cut into predictable lengths. With this knowledge, Life Technologies designed their restriction digests by pairing plasmids with a specific group of restriction endonucleases. The expected lengths of DNA can be compared to the actual DNA strands resulting from the digest as a means of identifying the plasmid sequence.

Using this technique, a sample of the plasmid, digested by restriction enzymes, is pipetted into multiple wells in an electrophoresis gel. During electrophoresis, an electric field is set up across the gel box. Over time the negative DNA molecules

move through the gel towards the positive side. The smaller molecules are able to move more quickly through the gel creating a method of separation based on size.

Figure 16. Electrophoresis Gel Box. The electric field is set up so that the negatively charged DNA starts on the negative end of the box. Negatively charged molecules then travel toward the positive side at rates proportional to their size.

Once the electrophoresis is complete, the agar gel and DNA are stained with ethidium bromide (EtBr). EtBr binds with the DNA and moves in between its hydrophobic base pairs which stretches the DNA fragment and removes water molecules from the ethidium cation. The result of this dehydrogenation is an increase in fluorescence of the ethidium.

Figure 17. DNA Electrophoresis Gel. The column on the left is called the ladder. The columns to the right are DNA molecules. Each line of the ladder corresponds to a different size molecule with the lines closest to the bottom being the smallest molecules. The ladder provides a way of indirectly measuring the size of the pieces of cut DNA.

The purity of the DNA can be determined by comparing the amount of Genomic DNA, which is large and travels the shortest distance, to the Plasmid DNA. This ratio of Genomic DNA over total DNA (Genomic plus Plasmid DNA) is presented as a percent purity. The amount of plasmid DNA as a percent of total DNA can then be multiplied by the concentration to determine the amount of Plasmid DNA.

Shortcomings of DNA Purification and Analysis

The end product is sold at a specific concentration of plasmid DNA, so the product must be well characterized during manufacturing to accurately market the product. The lack of precision in the separation and purification of the plasmid DNA from the genomic DNA and cell debris resulted in decreased yield and a lack of purity. The final step of confirming the correct identity of the plasmid and the level of purity were also in need of improvement. The identity confirmation was inconclusive, and the purity check was subjective. In order to improve the end product, both the harvesting and yield analysis steps of the process were investigated for improvement.

II. Objectives and Approach

There were three fundamental objectives: 1) improve the conclusiveness of the plasmid identification protocol, 2) reduce the ambiguity of the plasmid purity evaluation protocol, and 3) improve the purity and yield of the plasmids with regard to genomic DNA.

Improving the Identification Protocol

The pre-existing identification protocol was inherently inconclusive and had an unacceptably high rate of returning a false negative. It utilized a set of restriction enzymes to cut the strand of plasmid DNA at multiple locations. The restriction digest product was then electrophoresed, and the lengths of the resulting strands of DNA were compared to the expected lengths. If they all matched, the plasmid was identified as correct.

There were a few problems with this technique: the restriction enzymes being used were highly susceptible to damage during storage which results in unpredictable behavior; some of the cut sites were highly susceptible to methylation which blocks the enzyme from cutting the DNA; and the inherent possibility that an incorrect plasmid could coincidentally yield the expected lengths of DNA strands. To improve the identification step, an entirely new protocol was developed.

The new identification protocol sequences the manufactured plasmid and then compares the sequence of nucleic acids unique to that plasmid, approximately 300 long, to what the sequence is supposed to be. The result is a direct comparison, one nucleic acid at a time, of the manufactured plasmid against the correct sequence, resulting in an unequivocally conclusive test. To implement this new protocol, a unique region of each plasmid had to be determined, a database of each plasmid and its identification sequence developed, and a method for sequencing and comparison installed.

Improving the Purity Analysis Protocol

The pre-existing protocol, which analyzed the percent genomic DNA in the purified sample, was highly subjective. The protocol called for the purified DNA to be electrophoresed in an agar gel, and then stained with EtBr. Because the strand of plasmid DNA is much smaller than genomic DNA, the two molecules travel at different rates during electrophoresis and are thus separated. A picture of the stained gel was taken under ultraviolet light, and an operator then made a subjective assessment of the purity by comparing the extent to which EtBr was bound to each type of DNA molecule. Because of the subjective nature of this protocol, the purity evaluation was neither precise nor reproducible.

Figure 18. Previous Electrophoresis Gel. An image of an electrophoresis gel used in the pre-existing purity protocol. DNA banding is imprecise.

One of the objectives of this project was to develop a new purity protocol utilizing software that essentially quantifies the luminescence of the stained DNA, which is directly proportional to the amount of DNA.

Improving the yield of the Purification Process

Until the reproducibility of the purity evaluation technique was improved, the purity and yield of the entire purification process could not be evaluated for improvement. Once the process could be accurately evaluated for purity, the correct yield could also be determined. With that metric now available, a strategic plan of investigation was developed to specialize the purification process for each plasmid.

III. Methods and Materials

Improving the Identification Protocol

Once the plasmid has been purified, it is important to verify its identity. The pre-existing method was inconclusive and its components were prone to malfunctioning. The new method provides a decisive evaluation of each nucleotide by sequencing the DNA in a region of the plasmid that is uniquely its own.

To achieve this, a sample of the purified plasmid of interest is sent to a local genetic sequencing company where the identifying DNA is sequenced. But first, a region to be sequenced must be identified. Determining the region appropriate for confirming the plasmid's identity was the first step to improving the identification process.

Life Technologies holds a database with the nucleotide sequence for each plasmid of interest.

GGGACAAGTCAACATGGAGGTCAAGGTAGAGAATGTTGGCAAGTCACAGGA
GCTTAAAGTCAAAGTCAAGAATTTTATAAAAAGGTCTGATTGCAAGAAAAAA
CTTTTGCCTTGATTTTAGGGCTAGTCAGCTTTGAACTCACTATGAATATAAT
GCTGTCTGTCATGTATGTGGAGTCAAATGAGGCCCTAAGTTTATGTAGGATCC
AAGGGACTCCTGCTCCAAGGGATAATAAGACAAACACAGAAAACGCAACAA
AGGAAACAACACTCCACACAACGACCACAACAAGGGATCCAGAGGTGAGGG
AAACAAAAACCACCAAGCCCCAGGCCAATGAAGGAGCAACAAACCCAAGCA
GGAACCTCACCACCAAGGGAGACAAACACCAACGACAAGAGCAACAACAG
AGGCAGAACTGGAAAAACAAAGCAAACAACCCACAGAGCCAGGCACCAGCA
CCCAAAAGCACACCCCCGCAAGACCAAGCAGCAAATCCCCACCACAACAC
AAGCAATAGCACAACCTGACAACACCAACAACCCCAAAAGCAAGCACAGCAC
CCAAGAACAGACAGGCAACAACCAAAAAAACCGAAACGGACACCACAACAG
CAAGCAGAGCAAGGAACACCAACAACCCACAGAGACAGCAACAACAACCTC
CCAAAGCAACAACAGAAACAGGCAAGAGCAAAGAGGGGCCAACACAGCAC
ACAACCAAAAGAACAGCCCGAGACAACAGCAGGAGTTAGTTTAAGAGCAAGT
TAGTTAATTAAAA

Figure. 19. Raw Plasmid Sequence. An example of a plasmid sequence before it is entered into Vector NTI. Analysis in this format without a bioinformatics program is tedious and inefficient.

These sequences were transferred into a bioinformatics program called Vector NTI. This software enables the search of specific sequences within the molecule so that genes, primer recognition sequences and enzyme recognition sites can be efficiently identified and labeled.

Figure 20. Vector NTI Pictorial of Plasmid. The markers identify restriction enzyme recognition sites and regions that code for particular proteins.

The majority of the plasmid's genetic code is not unique to itself. Most plasmids code for a multitude of proteins necessary to carry out basic cell life that are common amongst a group of similar cells. The unique region of the plasmid that is used for its identification is called the Multiple Cloning Site (MCS). The MCS is a synthetic region of the plasmid with a particularly high concentration of

restriction enzyme sites. It enables a wide range of insert locations for a scientist designing new genes into a plasmid and is the location of the genetically engineered genes of interest. Once the MCS is located, primer sequences on either side of it are identified for sequencing.

pBAD/gIII A

```

1  AAGAAACCAA TTGTCCATAT TGCATCAGAC ATTGCCGTCA CTGCGTCTTT TACTGGCTCT TCTCGCTAAC CAAACCGGTA

81  ACCCCGCTTA TTAAAAGCAT TCTGTAACAA AGCGGGACCA AAGCCATGAC AAAAACGCGT AACAAAAGTG TCTATAATCA

161 CGGCAGAAAA GTCCACATTG ATTATTTGCA CGGCGTCACA CTTTGCTATG CCATAGCATT TTTATCCATA AGATTAGCGG
      O1 Region          CAP binding site          pBAD Forward priming site
      I2 and I1 Region

241 ATCCTACCTG ACGCTTTTTA TCGCAACTCT CTACTGTTTC TCCATACCCG TTTTGTGGGC TAACAGGAGG AATTAACC
      -35          -10          RBS

319 ATG AAA AAA CTG CTG TTC GCG ATT CCG CTG GTG GTG CCG TTC TAT AGC CAT AGC ACCATGGAGC
    Met Lys Lys Leu Leu Phe Ala Ile Pro Leu Val Val Pro Phe Tyr Ser His Ser
      gene III signal sequence          gene III cleavage site          Nco I  Eco136 I Xho I

383 TCGAGATCTG CAGCTGGTAC CATATGGGAA TTCGAAGCTT TCTA GAA CAA AAA CTC ATC TCA GAA GAG GAT CTG
      Sac I Bgl II  Pst I Pvu II Asp718 I Kpn I  EcoR I  BstB I Hind III  Xba I          myc epitope
      Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu

457 AAT AGC GCC GTC GAC CAT CAT CAT CAT CAT CAT TGA GTTTAAAC GGTCTCCAGC TTGGCTGTTT TGGCGGATGA
      Asn Ser Ala Val Asp His His His His His His ***
      Sal I          Polyhistidine region          Pme I

531 GAGAAGATTT TCAGCCTGAT ACAGATTAAA TCAGAACGCA GAAGCGGTCT GATAAACAG AATTGCGCTG GCGGCAGTAG

```

Figure 21. Life Technologies' Product Manual Sequence Description. This was used in concert with the Vector NTI map for primer design.

With the identifying region located and the surrounding primers determined, the plasmid can be sent out for sequencing.

The DNA is sequenced by separating the two phosphate-deoxyribose backbones, creating two single-stranded DNA molecules. The primers surrounding the region of interest then coordinate with their corresponding sequence of nucleic acids and catalyze DNA Polymerase's elongation of the

complimentary strand. This is done in the presence of deoxyribonucleotide monomers (dNTP), the single units of a DNA strand, and a dideoxyribonucleotide triphosphate analog (ddNTP). The ddNTP's are dNTP's with the 3'-OH group replaced with a 3'-H. This substitution renders the ddNTP unable to form the phosphodiester bond between dNTPs, thus terminating elongation of the strand.

Figure 22. Molecular Structure Illustration. An Illustration of the structural difference between deoxyadenosine triphosphate (dATP) and dideoxyadenosine triphosphate (ddATP).

Four reactions are carried out in parallel, each with a different ddNTP:
 dideoxyadenosine triphosphate (ddATP), dideoxythymidine triphosphate
 (ddTTP), dideoxyguanosine triphosphate (ddGTP), and dideoxycytidine

triphosphate (ddCTP). Each of the four reactions incorporates a different fluorophore that modifies the corresponding ddNTP. The four reactions are then run through an electrophoresis gel to determine molecular length and constituent nucleotides are identified according to light emission properties of the different fluorophores. In this way a large population of randomly sized DNA molecules with a known number of base-pairs can be evaluated. This process is highly automated allowing thousands of nucleotides to be identified in minutes. The evaluation program then produces a read-out similar to the one below.

Figure23. Typical Sequencing Read-out. Each colored peak corresponds to a specific wavelength and nucleotide. The height of the peak is a representation of the strength of the nucleotide identity as determined by number of matches.

With the DNA sequenced, the identity of the purified plasmid can be compared, nucleotide by nucleotide, to the correct sequence for conclusive positive identification. Figures 24 and 25 illustrate how Vector NTI is used to compare

the actual plasmid DNA, as determined by the outsourced sequencing company, with the Identification Sequences in Life Technologies' database.

The ContigExpress application of Vector NTI compares the sequence identified by the outsourced sequencing company with the known Identification Sequence in Life Technologies' database.

Figure 24. ID Sequence Lay-out. Pictorial representation of how sequenced DNA is spatially coordinated with the Identification Sequence in Life Technologies' database.

Each MCS has three versions (A, B and C), one for each place in the reading frame. This allows for flexibility when designing the MCS into a particular plasmid. Accordingly, there are three Identification Sequences per MCS per plasmid. In Figure 24, the three bands in the middle correspond to the A, B and

C versions of the plasmid Identification Sequences. Above are the 3 forward priming sequences (there is a duplicate of each for insurance resulting in a total of 6). Below the Identification Sequences are the reverse priming sequences with the same redundancy as the forward priming sequences. As illustrated, neither the forward nor the reverse priming sequence individually spans the entire Identification Sequence; but together, with some overlap, the entire Identification Sequence can be compared.

Figure 25. ContigExpress. An image of how ContigExpress compares the sequences in Figure 24 Nucleotide by nucleotide. It is here that the determination of correct identification is made. The highlighted discrepancies in the middle are due to the single and double nucleotide shifts of the A, B and C reading frames.

Improving the DNA Purity Analysis Protocol

The first step was to change the type of electrophoresis gel used. The pre-existing protocol used Life Technologies' E-Gel®, a prepackaged electrophoresis gel with a built-in ladder and its own electric field source. These E-Gels, however, run too hot for our application, causing denaturation, which results in “smiling” of the DNA bands. Smiling DNA decreases the readability and precision of the banding. The E-Gels were replaced with hand-poured gels which run at an appropriate temperature and are more customizable. The result is a more precise and reliable DNA banding pattern.

Figure 26. E-Gel with “Smiley Faces”. These pre-packaged electrophoresis gels used in the pre-existing protocol ran too hot. This caused denaturation of the DNA resulting in “smiley face” bands. The bands of the ladder are also distorted.

Figure 27. Hand-Poured Gel. These gels run at a more appropriate temperature and provide a higher degree of customization. The result is a more precise DNA size gradation.

With the hand-poured electrophoresis gels, the banding is more consistent, but the ambiguity of the purity analysis still needed to be removed. To eliminate the subjective nature of the protocol, evaluation was changed from a visual assessment by a technician to pixel evaluation by computer software.

A 2UV Transilluminator was used to capture an ultraviolet image of the gels.

Then VisionWorksLS software was used to quantify the amount of each type of DNA based on the density of black and white pixels.

Figure 28. 2UV Transilluminator.

Figure 29. Density Box. An example of the defined region used to evaluate the quantity of DNA in the band.

After electrophoresis on the digested DNA is complete, the gels are stained with EtBr. EtBr binds to the DNA and fluoresces when exposed to ultraviolet light from the 2UV Transilluminator. The fluorescence of the gel is then captured by the 2UV Transilluminator and communicated to the VisionWorksLS imaging

software. The imaging program enables the user to view the DNA bands and employ objective numerical evaluation techniques.

Figure 30. VisionWorksLS Imaging Software Interface.

Once captured by the viewing software, the electrophoresed banding patterns become readily apparent. Each purified plasmid is electrophoresed into two

columns, one that has been digested by a restriction enzyme and one that has not been digested.

Figure 31. Electrophoresis Banding Pattern. An illustration of typical electrophoresis banding patterns for digested and undigested plasmids.

The first column, starting from the left, of each pair contains the purified sample that has not been digested. When the undigested purified plasmid is electrophoresed, 3 prominent bands show up. The largest molecular band is the genomic contamination, which is essentially what the purity analysis is evaluating for. The next largest band is nicked plasmid, or open circular DNA with one strand cut. And the third band, the smallest molecule, is the supercoils of DNA. The electron micrograph image below is a helpful illustration of these three common conformations.

Figure 32. Electron Micrograph Image of Bacterial DNA Plasmid.

The second column contains the same purified plasmid as the first, but has been cut once by a restriction enzyme. The predominate band in this column is the cut supercoil of DNA. The cut plasmid runs more slowly during electrophoresis because there is more friction associated with a long strand of DNA than with the balled up circular version of the same molecule. More surface area results in more molecular interaction at the surface. Above the cut plasmid is what is

called a “smear”. The smear is caused by the genomic contamination and is the result of the restriction enzyme cutting it into many different size molecules.

For our purposes, we used the program to evaluate the amount of DNA present in each band relative to other bands. Because the amount of DNA is directly proportional to the amount of fluorescence, the software can quantify the amount of DNA as a function of the pixel density.

Figure 33. Defined Density Boxes.

Region	Total Density	Total Background	Mean Density	Mean Background	Total Raw Density	Mean Raw Density
1	37050.3367	70099.5005	14.4728	27.3826	106816	41.725
2	240366.2647	70099.5005	93.8931	27.3826	310463	121.2746
3	228948.4995	70099.5005	89.433	27.3826	299048	116.8156
4	9536.3606	70099.5005	3.7251	27.3826	79621	31.102
5	53103.03	70099.5005	20.7434	27.3826	123201	48.1254
6	163776.4995	70099.5005	63.9752	27.3826	233876	91.3578
7	135507.2388	70099.5005	52.9325	27.3826	205590	80.3086

Table 1. Raw Density Data from VisionWorksLS Software. The “Total Raw Density” refers to the number of white pixels in a particular boxed region. This is the value used to calculate DNA quantities

To calculate the percent genomic contamination in the purified plasmid sample, the density of the genomic DNA was divided by the total DNA (supercoil plus genomic). The ratio of supercoil to linearized plasmid is an indication of the condition of the supercoil.

In order to validate the reproducibility of the software-analyzed protocol, three sets of purified plasmids were electrophoresed at three different volumes. Each set was evaluated for both linearity of quantity vs. volume, and the variance of percent genomic contamination at 95% confidence.

Improving the Yield of the Purification Process

With the new purity quantification technique, it became possible to accurately evaluate the purity and yield of the purification process. Retains of each manufactured batch of purified plasmids for the last several years had been preserved in a freezer at -4°C. They were easily categorized by purification kit type: 1) MegaPrep, 2) GigaPrep and 3) Maxicard. For all three categories, the functional chemistry is the same. MegaPrep and GigaPrep are essentially the same process except for a difference in the size of the filtration cartridges used to separate the DNA from proteins and larger cell debris.

Figure 34. Filtration Cartridge. Filtration cartridge used in MegaPrep and GigaPrep purification processes.

Specifications*	Megaprep	Gigaprep
Starting <i>E. coli</i> Culture Volume at OD 2.0	500 mL	2.5 L
Binding Capacity**	2.5 mg	10 mg
Lysate Filtration Cartridge Reservoir Capacity	230 mL	450 mL
DNA-Binding Cartridge Reservoir Capacity	200 mL	400 mL
Elution Volume	50 mL	100 mL
DNA Recovery	90–95 %	90–95 %
DNA Yield***	1.5–2.5 mg	7.5–10 mg

Table 2. Megaprep/Gigaprep Specifications. Specification table indicating the difference between Megaprep and Gigaprep purification capabilities.

Maxicard is an all-inclusive automated machine that carries out the entire purification process – from bacteria culture to purified DNA – but on a much smaller scale than either MegaPrep or GigaPrep.

Figure 35. Maxicard Purification Machine.

Retains purified by each category of filtration were digested with an appropriate restriction enzyme, as described above, and evaluated against a multitude of metrics in an effort to discern any indication as to which was most appropriate for each type of plasmid.

IV. Results and Discussion

Identification Assay

Nearly 25 plasmid characterization files were created in Vector NTI and verified with their sequenced genetic code. Below are the identification validation results of a typical plasmid family.

The out-sourced sequencing company returned the sequenced DNA information in the format below.

Use the text box to Copy contents to the clipboard.

[illegible]

Figure 36. Sequence Data. The sequenced plasmid file returned by the outsourced company. The “N’s” are a result of imprecision localized around the primers.

The sequence file was then imported into Vector NTI and compared to the expected DNA sequence file (SeqID) in Life Technologies' database.

Figure 37. SeqID File. Above is the SeqID file against which the sequenced DNA is compared to.

The sequenced DNA file and SeqID are compared in a VectorNTI application called ContigExpress, shown below.

Figure 38. ContigExpress Identification Evaluation.

The sequenced DNA matches each corresponding SeqID file. The highlighted region of the plasmid is due to the frame shifts of the A/B/C versions. From these matches the identity of each plasmid is confirmed.

Purity Analysis Assay Results

In order to determine the reproducibility of the VisionWorksLS density measurement software, three different volumes (250 μ L, 500 μ L and 1000 μ L) were electrophoresed for three different purified plasmids.

Figure 39. Purity Evaluation Verification Gel. Three different plasmids were evaluated, each at three different loaded quantities, in order to verify the reproducibility and linearity fit of the evaluation technique.

DNA	ng loaded	Total Density	Type	%SC	%Total
Plasmid 1	1000	252344	Supercoil		
Plasmid 1	500	169193	Supercoil		
Plasmid 1	250	101969	Supercoil		
Plasmid 2	1000	218039	Supercoil		
Plasmid 2	500	136203	Supercoil		
Plasmid 2	250	67540	Supercoil		
Plasmid 3	1000	222855	Supercoil		
Plasmid 3	500	104198	Supercoil		
Plasmid 3	250	82435	Supercoil		
Plasmid 2	1000	47703	Genomic DNA	19%	16%
Plasmid 2	500	22647	Genomic DNA	13%	12%
Plasmid 1	250	13165	Genomic DNA	13%	11%
Plasmid 2	1000	82142	Genomic DNA	38%	27%
Plasmid 2	500	38166	Genomic DNA	28%	22%
Plasmid 2	250	16971	Genomic DNA	25%	20%
Plasmid 3	1000	62736	Genomic DNA	28%	22%
Plasmid 3	500	23914	Genomic DNA	23%	19%
Plasmid 3	250	17839	Genomic DNA	22%	18%

Table 3. Purity Calculations. Software density calculations for gel shown in Figure 38.

Analysis on the resulting data showed satisfactory results with percent genomic contamination confidence intervals of 2.2%, 3.5% and 2.0% at 95% confidence, well within the desired 95% confidence range of 5%. An additional regression analysis, shown below, on the linear fit of the data points found excellent R^2 values for the slope of DNA quantity vs. volume: 0.9806, 0.9834, 0.9647, 0.9999, 0.9958 and 0.9572. These values indicate a high degree linear relationship

between the points among each DNA type of each plasmid and further suggest that this analysis technique is reproducible and robust.

Figure 40. Pixel Density vs Volume. This graph shows the linearity between pixel density and quantity of DNA.

With the new Purity Analysis Assay verified for reproducibility, all the available plasmids from the previous year were analyzed for percent genomic contamination. Using the technique described in *Methods and Materials*, a table was developed to: 1) evaluate the purity of the previous year's product and 2) build a spreadsheet of data from which to evaluate individual parameter contribution to yield and percent genomic contamination.

N a m e	HF Enzyme	Media type	Yield (ug)	% g DNA	SC/Linear	Size	Scale	ug DNA/g cells
A	BamHI	LB	750	9%	0.98	5500	MC	2532
B	BamHI	TB	7240	3%	1.02	3890	GIGA	724
C	BamHI	TB	6420	10%	1.91	3890	GIGA	321
D	BamHI	TB	7480	11%	1.13	3890	GIGA	748
E	BamHI	TB	9000	17%	1.47	3890	GIGA	450
F	BamHI	TB	6127	11%	0.34	3890	GIGA	306
G	BamHI	TB	8400	19%	1.43	3890	GIGA	420
H	BamHI	TB	13550	5%	0.89	3890	GIGA	678
I	BamHI	TB	13550	11%	0.83	3890	GIGA	678
J	BamHI	TB	13550	10%	0.76	3890	GIGA	678
K	BamHI	TB	13550	9%	0.75	3890	GIGA	678
L	BamHI	TB	13550	7%	0.85	3890	GIGA	678
M	BamHI	CW	14600	11%	0.81	3890	GIGA	1502
N	BamHI	TB	9150	14%	1.34	3890	GIGA	458
O	EcoRI	LB	1615	26%	0.96	4102	MEGA	508
P	EcoRI	LB	995	8%	0.77	4102	MEGA	332
Q	EcoRI	LB	435	14%	1.51	4102	MEGA	146
R	EcoRI	LB	1836	21%	1.05	4102	MEGA	291

Table 4. Plasmid Characterization. Abbreviated version of the table used to discern suggested parameter effects on yield and purity. From this table, patterns were suggested, hypotheses were developed, and experiments were designed to test those hypotheses.

The Plasmid Characterization Table suggested a few trends. The strongest suggestion was that the size of the plasmid had an effect on purity.

Figure 41. Genomic Contamination vs Plasmid Size (bp). This graph suggests the larger plasmids have lower contamination.

One explanation for this trend is that larger plasmids have more negative charge to attach to the positively charged purification column, resulting in a greater degree of purification. This trend may be useful in predicting purity of different size plasmids; however, the size of the plasmid is determined by the proteins coded in the DNA and cannot be altered to improve purity.

The data in Figure 41 also indicated a possible correlation between the amount of bacteria culture which was purified and the size of the purification columns used during purification.

Figure 42. Genomic Contamination vs Purification Scale. This table suggests the GigaPrep purification scale achieves the least contamination.

The MaxiCard (MC) purified plasmids had the lowest contamination; however, they always use the same volume and concentration of bacteria, which is

significantly lower than the GigaPrep and MegaPrep purified plasmids.

Unfortunately, MaxiCard purification cannot be scaled up to support the high volume purification requirement of most plasmids. Between the two scales of purification which can support high volume production, the data indicate Giga is better than Mega. As with most types of filters, GigaPrep and MegaPrep purification columns can be over-loaded. GigaPrep is the larger of the two and can most likely support higher volumes of media before reaching its over-load threshold. This reasoning led to the hypothesis that increasing amounts of bacteria to be purified results in an increasing level of contamination. To test this hypothesis, plasmids were purified at varying quantities using the same scale of purification, and evaluated for trends between quantity of bacteria purified and genomic contamination level.

Yield Assay

Based on the suggested correlations in Table 4, varying quantities of bacteria were purified using MegaPrep kits to provide insight into the correlation between bacteria quantity and purity.

Two different plasmids were purified at 4 different quantities, shown in Table 5, in an effort to evaluate the effect of the amount of bacteria and DNA purified on purity.

	PBAD HIS B				PBAD GIII B			
Culture (mL)	500	750	1000	1250	750	850	900	1000
Bacteria (g)	2.69	4.74	5.57	9.23	3.94	5.35	5.79	5.78
DNA Yield (ug)	408	1590	1000	680	3170	3480	7940	7940
Yield/Bacteria	152	335	180	74	805	650	1371	1374
% gDNA	11	18	13	15	11	21	17	17

Table 5. GigaPrep Purity and Yield Assay Results. “Culture” is the volume of bacteria culture. “Bacteria” is the mass of bacteria in the culture. “DNA Yield” is the mass of DNA resulting from the purification process. “% gDNA” is the percent of total DNA which is genomic.

To help provide perspective, the results from Table 5 were graphed as purity vs the raw amount of DNA yielded, and purity vs the ratio of DNA yielded over mass of bacteria purified. The first graph is the most direct measurement of how the amount of DNA purified affected the quality of purification. The second graph illustrates how purity and the density of DNA in the bacteria are related.

Figure 43. MegaPrep Purity vs Bacteria Quantity.

There appears to be a trend of increasing percent genomic contamination as amount of DNA purified increases. This corroborates the theory that the charged column used to separate the DNA from the protein and RNA loses its effectiveness as it approaches a saturation point.

Figure 44. MegaPrep Purity vs Efficiency.

The ratio of DNA yielded over bacteria purified was initially looked at as an indication of the efficiency of the purification process. Except for the outlier at approximately 800, the % genomic contamination seems to be increasing with efficiency. However, the amount and type of DNA produced by the bacteria can play as large a role here as the effectiveness of the purification process. For example, different plasmids replicate at different rates relative to the host cell's genomic DNA. This means that different plasmids result in varying amounts of plasmid DNA relative to genomic DNA before the purification process begins. Also, given an equal plasmid replication rate relative to the genomic replication rate, an 8K bp plasmid will have half the genomic contamination of a 4K bp plasmid. Even though the number of plasmids is the same, because the

measurement is based on the quantity of constituent amino acids, not the number of plasmids themselves, the level of genomic contamination will differ.

Four plasmids were purified using the MaxiCard purification machine. The results are shown in Table 6. Because this machine requires a specific volume and concentration, these values are the same for all four plasmids. This consistency provided an opportunity to investigate the effect of plasmid size while controlling volume and concentration.

	PDEST40#1	PDEST40#2	PEF5	PBADHIS
Size (bp)	7100	7100	7500	4100
Culture (mL)	125	125	125	125
Bacteria (g)	0.275	0.275	0.275	0.275
DNA Yield (ug)	609	766	890	23
Yield/Bacteria	2215	2785	3236	84
% gDNA	9	24	6	13

Table 6. MaxiCard Purity and Yield Assay Results.

The graphed results show a trend of decreasing genomic contamination as efficiency increase, except for the outlier at 24%. This data suggests a possible trend, but more data points are needed to a statistically significant trend.

Figure 45. MaxiCard Purity vs Efficiency.

Figure 46. MaxiCard Purity vs Plasmid Size.

As with the MegaPrep purified plasmids, genomic contamination, except for one data point, seems to decrease as plasmid size increases. This is possibly due to more attractive forces between the purification column and the plasmid due to the increased number of base pairs; or the result of a larger plasmid containing more DNA. The larger plasmid, given the same number of plasmid molecules and same amount of genomic DNA, will always have a lower level of genomic contamination.

Overall, the data suggests purity is most significantly affected by both purification scale and plasmid size. However, none of the experiments have enough data points to validate these theories with statistical significance. Further experiments are needed to confirm this hypothesis.

V. Conclusion and Future Works

Identification Assay

The new protocol for confirming the correct identity of the purified plasmid is an improvement in many ways. The new method is simpler, more robust and less expensive. There are no inconclusive results due to false negatives and methylation, and the unstable restriction enzymes have been obviated. Compared to the old paradigm, there is essentially zero ambiguity of a correct match.

While currently very effective, an internal control sequence within the plasmid may be worth investigating. In the event that there is a bad sequencing reaction or bad DNA, a control sequence would create a secondary control. Also, it may be worth looking into an alternative unique region for comparison.

Purity Assay

The new purity evaluation process is a vast improvement over the extreme ambiguity and variability of the previous method. However, the amount of time that the DNA is soaked in the stain, and the time it is de-stained is not regulated. Consistent staining would improve the probability that the DNA in each gel interacts equally with the EtBr. Different types of stains could possibly optimize the consistency and speed of the process as well.

Yield Assay

The yield assay suggested some interesting trends with regard to the most significant factors affecting genomic contamination. While plasmid size and purification scale seem to play a large role and should be further investigated, it would be useful to research the replication rate of different plasmids relative to the host cell chromosome. This relative replication rate would provide insight into the ratio of genomic DNA to plasmid DNA, inherent in the biology, which certainly affects the yield of the purification process.

BIBLIOGRAPHY

Alberts, Bruce, et al. *Molecular Biology of The Cell*. 4th ed. New York: Garland Science, 2002. Print.

Brown, T.A. *Gene Cloning & DNA Analysis*. 6th ed. Chichester: Wiley-Blackwell, 2010. Print

Dardel, Frederic and Kepes, Francois. *Bioinformatics: Genomics and Post-genomics*. Chichester: Wiley and Sons Ltd, 2006. Print.