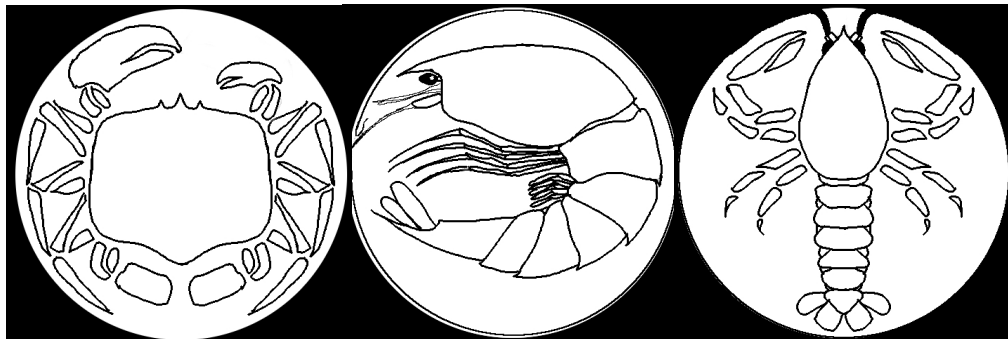


Edition

2018

BRACKEN-GRISSOM LAB

Crustacean Genomics and Systematics



Gene Sequencing: A Beginner's Guide

3000 NE 151st Street
Marine Science Building, Rm 330
North Miami, FL 33181

Email: brackengrissomlab@gmail.com
Phone: 305.919.5838
Website: www.brackengrissomlab.com



TABLE OF CONTENTS

<u>SECTION1</u> –INTRODUCTION.....	1
<u>SECTION2</u> –THE HBG DATABASE.....	1
<u>SECTION3</u> –WORKFLOW: EXTRACTING DNA TO SEQUENCING GENES	2
<u>PART 1:</u> EXTRACTING DNA	3
<i>Selecting an extraction protocol</i>	4
<i>DNA extraction form</i>	3
<i>Verifying the quality of your DNA extractions</i>	4
<i>Organizing your xDNA gels</i>	6
<i>Update the HBG database</i>	8
<u>PART 2:</u> AMPLIFYING DNA.....	9
<i>Polymerase chain reaction</i>	9
<i>Primers</i>	9
<i>Selecting primers</i>	11
<i>Ordering primers</i>	15
<i>Running a PCR</i>	15
<i>Verifying the success of your PCR</i>	15
<i>Organizing your PCR gels</i>	19
<u>PART 3:</u> SEQUENCING DNA	21
<i>Preparing your samples</i>	21
<i>Cleaning your sequences</i>	22
<u>SECTION4</u> ORGANIZATION!.....	23
<u>SECTION5</u> –CONTACTS	24

Section 1 – Introduction

For any researcher, learning a new field of study can be intimidating. The field of genetics is certainly no exception. This manual is written with the intention of streamlining the learning process; providing the reader with some basic protocols, and a simple start-to-finish workflow guideline for genetic sequencing. It assumes the work will be done using the Bracken-Grissom Lab (BG Lab) facility and equipment, however, the information contained herein should be easily transferrable to any work environment with adequate equipment.

Section 2 – The HBG Database

This section is relevant only to members of the BG Lab, but it does include information that others may find informative, specifically concerning how the BG lab organizes its specimen collection, and the tools it uses to do so. Skip to SECTION 3 if this does not pertain to you.

The BG Lab utilizes a databasing software called FileMaker Pro (FMP). FMP allows the lab to organize collected and donated specimens in a way that their information can be quickly and easily located. It records information like the voucher ID number, taxonomy, when and where the specimen was collected, tissue and DNA extractions, and much more for every single specimen in our collection, or that we've ever worked with. With nearly 10,000 entries (as of the creation of this manual), the HBG database is a critical component of the BG lab, without which would render the lab completely non-functional. All members are expected to learn and understand how to use it. A hands-on tutorial will be included with your training in the lab.

The most important thing to understand about how the BG Lab tracks its specimens is that each one is assigned a voucher number upon collection. This number is also known as a specimen’s “HBG number,” and stays with that specimen in perpetuity. By assigning an HBG number to every specimen we work with, we can readily track that specimen using the HBG database, and it allows for easy organization of all research data pertaining to that specimen. This will become clearer with experience working in this lab.

The BG Lab also has a sizeable collection of crustacean specimens. Officially, this collection is known as the Florida International Crustacean Collection (FICC). Every specimen within the FICC has been databased, and has been preserved to ensure the conservation of their morphological and molecular information. As a member of the BG Lab, you will become familiar with this collection.

Section 3 – Workflow: Extracting DNA to Sequencing Genes

This section of the manual follows the organization of the folders and files you were given with this manual. For example, “PART 1: EXTRACTING DNA” found below corresponds to the “PART 1 EXTRACTING DNA” folder found in the zip file you were given which included this manual. Any supplementary material that is referenced in PART 1 here will be found in the PART 1 folder. Every subsequent PART of this section will also have a corresponding folder with the same name where their referenced supplementary material can be located.

PART 1: EXTRACTING DNA

Assign voucher numbers to your samples

As mentioned in SECTION 1, the BG Lab gives every specimen a voucher ID number to help with tracking the sample throughout the process of gene sequencing. If your samples don't already have voucher numbers, a standard format is as follows:

YourInitials##

For example, if your name were John Doe Smith, you would name your first 5 samples:

JDS01

JDS02

JDS03

JDS04

JDS05

This format is useful because it's simple, yet informative. Of course, this is just a suggested format. Feel free to create your own system as you see fit.

DNA extraction form

Before beginning the extraction process, a "DNA extraction form" should be completed. The purpose of this form is to keep extraction data organized and clear, not only to the person performing the extraction, but also to anyone who may need to reference their progress on the project. If multiple people are working on the same project, each performing extractions, it can be very easy to lose track of who extracted what. This can lead to inefficiency as multiple unnecessary extractions are performed

on the same samples. An example of the form is included as supplementary materials (**xDNA_Form.docx**)

Selecting an extraction protocol

The first step in the process of gene sequencing is the isolation of your sample's DNA. To do that, the Bracken-Grissom Lab commonly uses one of two different protocols: DNeasy Qiagen Kit (Qiagen) Extraction, and Phenol Chloroform (PC) Extraction. Deciding which protocol to use depends on a few things. Qiagen is quicker and simpler than PC (~5 hours versus 3 days, respectively), but PC requires less tissue and generally yields higher quality extractions. For most applications, Qiagen is sufficient, and, due to its simplicity, is the BG lab's preferred extraction method. PC is used when tissue is limited, and/or when Qiagen yields low quality extractions. This might be due to the age of the sample, as DNA often degrades with age. PC might be a better choice for older samples (ex. 10-20 years old or more). The protocols for each method are included as supplementary materials (**Qiagen_Extraction.PDF**, and **PC_Extraction.PDF**).

Store your DNA extractions at -20°C

Verifying the quality of your DNA extractions

To verify the quality of your DNA extractions (xDNAs), two tests are performed. The first test determines the quantity of DNA in the xDNA, while the second measures the quality of the DNA.

Qubit test

To measure the quantity of DNA in your xDNAs, a machine called a fluorometer is used. This machine measures the light absorbance of your xDNA to determine its concentration. Light is passed through the sample, but some of it is absorbed by the DNA and fails to reach the sensor on the other side. The more DNA there is present in the sample, the less light passes through to be detected by

the sensor. In this way, it can be accurately determined how much DNA there is in the sample. Typically, the concentration is measured in units of (ng/uL). Any concentration above 5 ng/uL is ideal, but concentrations as low as 1 ng/uL have been used successfully.

The specific fluorometer model the BG Lab uses is the Invitrogen Qubit 2.0 by LifeTechnologies. It is referred to as “the Qubit.” The protocol for using the Qubit is included as supplementary materials (**Qubit_Protocol.PDF**).

xDNA gel electrophoresis

Depending on a few different factors, your xDNA may have some degradation. If the tissue was poorly preserved, if it’s considerably old, or if there was an error during the extraction process, the DNA may be fragmented to varying degrees of severity. If this fragmentation is too severe, the xDNA will be unusable for molecular work.

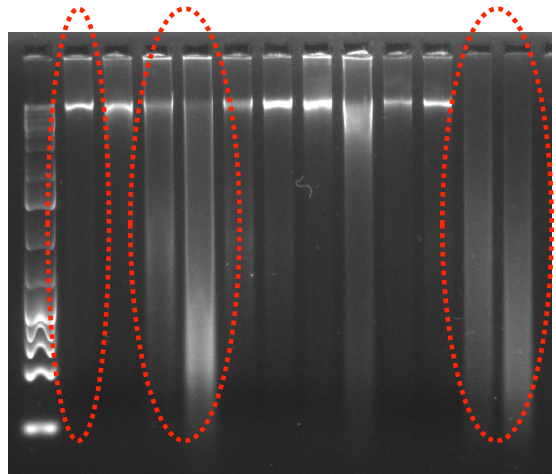
To measure the degradation of the xDNA, it is run through a process called “Gel Electrophoresis.” A sample of the xDNA is inserted into an agarose gel, which is simply a flat square of a gelatin-like substance. The gel is then subjected to an electric current, which causes the DNA to travel through it. The gel has a matrix-like structure that hinders the movement of the DNA, so it takes time for the DNA to travel from one end of the gel to the other. The longer a molecule of DNA is, the longer it will take to travel through the gel. After a period of time, the gel can be examined to see how far the DNA samples have traveled. In this way, the average length of the DNA molecules within a sample of xDNA can be measured. The protocol for running an xDNA gel is provided as supplementary material (**xDNA_gel_Protocol.PDF**).

Interpreting your xDNA gel

High-quality xDNA should contain whole chromosomal DNA, with limited fragmentation or degradation. In other words, it should have very long strands of

DNA. Therefore, an ideal xDNA sample will show a strong, single band that travelled very little distance across the gel.

If degradation of the DNA has occurred, it will appear as a long smear on the gel, with the fragments of various sizes travelling at varying speeds across the gel. This doesn't necessarily mean the xDNA is unusable. A molecule of DNA can be several hundreds of thousands of base pairs long. For gene sequencing, as long as some fragments are large enough, it's highly probable that the 200 to 1000 bp gene sequence you'll be targeting will still be intact, and available for gene amplification. Only the most severe degradation should be a cause for concern, and may require re-extraction of the DNA.



From left to right, each hashed circle shows high quality, medium quality (still useable), and low quality extractions (Possibly unusable).

Organizing your xDNA gels

The easiest way to add unnecessary difficulty to a research project is to disregard organization. This is especially true the larger your dataset becomes. As you perform more and more DNA extractions, it will inevitably become more and more difficult to keep track of which samples have been extracted, what the quality of the extractions were, their locations, etc. The following are helpful tips and suggestions that can make organization easy, and can help reduce feelings of frustration and self-hatred as you make progress with your research.

Keep your xDNAs together and organized

This may sound simple and obvious, but some projects might include 50, 100, possibly more xDNAs, and with 2 or 3 people working together, using the same samples, it can be surprisingly easy for some samples to get separated from the others. A good way to avoid this is to have a “Master Box” of xDNAs. From those xDNAs,

dilutions can be made for each member that needs them. This also increases the lifespan of each extraction. A simple 1:10 dilution is sufficient (assuming the concentration of the original xDNA is above 5 ng/uL), and should still provide enough xDNA to perform gene amplification protocols.

It is also HIGHLY SUGGESTED that the xDNAs be arranged in order by their voucher ID number. Again, this may sound simple, but it saves so much time, especially if the xDNA dataset is large.

Spreadsheet your xDNA results

As you perform extractions, you'll be extracting multiple samples at the same time. To help keep the results of these extractions organized, create a spreadsheet each time you run a gel. Start by creating a naming system for every xDNA gel that you run. In other words, every gel you run should have a unique name to help you track which samples were run on which gel. The naming system the BG Lab uses is:

xDNA_[YourInitials###]_[DateOfExtraction]

For example, if John Doe Smith runs 5 xDNA gels on random days in 2018, they would be named as follows:

xDNA_JDS01_01Jan2018

xDNA_JDS02_05Jan2018

xDNA_JDS03_10Jan2018

xDNA_JDS04_01Feb2018

xDNA_JDS05_23Feb2018

For each of these gels, a spreadsheet should be created that includes the name of the gel, an image of the gel, a table that describes which samples were run on that gel, the

Qubit concentration values for each sample on the gel, and any other relevant information specific to your project. Finally, these spreadsheets should be kept together in a separate folder for easy access and retrieval. An example spreadsheet is included as supplementary materials (**xDNA_HBG01_01Jan2018.xlsx**).

Update the HBG database

This section is only relevant to BG Lab members. If this does not pertain to you, skip to **PART 2**.

The final step upon completing a DNA extraction is updating the HBG Database. It needs to be updated with the concentration of the extraction, the date of the extraction, where the extraction is located (box # and location), and any other relevant information.

PART 2: AMPLIFYING DNA

Polymerase chain reaction

Now that the DNA has been isolated from the specimen, and your xDNA has been organized and databased, the next step is to begin targeting the genes within your samples for amplification. Gene amplification is essentially the process of extracting a specific region of DNA, and making millions of copies of it; “amplifying” its presence in the sample so that it can be more easily observed and sequenced. This is done via a process called Polymerase Chain Reaction (PCR).

A PCR works by exposing the xDNA sample to all of the necessary ingredients for DNA replication (dNTPs, polymerases, and primers), while also running it through a cycle of increasing and decreasing temperatures. Rather than trying to explain a moderately complex process with words, the following Youtube video gives an excellent illustration:

<https://www.youtube.com/watch?v=2KoLnIwoZKU>

If done successfully, a PCR results in millions of copies of identical double-stranded fragments of DNA, or “amplicons.”

Primers

So, how exactly does one specify which region of the DNA is to be amplified? That is determined by the sequence of the primers used during the PCR. At least two different primers are included in the PCR (the forward and the reverse primers), and they anneal to specific regions of the DNA based on their complementarity. During the heating phase of the PCR, as the DNA denatures into two single-strands, the single bases on each strand are exposed to the primers, allowing the primers to bind to their complementary regions on the DNA. One primer anneals to one strand, from which a polymerase can bind and begin building the complementary strand,

while the other primer does the same on the other strand, building its complementary strand in the opposite direction. The important part of this process is that the amplicon is “sandwiched” between the regions that the primers anneal to. After two cycles of heating and cooling, the target sequence is extracted as a copy from the original chromosomal DNA, and after the full 35 – 40 cycles are completed, the PCR sample is saturated with this amplicon.

Primer melting and annealing temperature

One of the most important characteristics to consider when selecting your primers is their melting temperature. This is the maximum temperature at which the primer can bind to the DNA during the PCR process. This temperature is determined by the number of hydrogen bonds between the primer and its complementary strand of DNA. Thinking back to undergrad genetics, you may remember that Adenine and Thymine pairs form with two hydrogen bonds, while Cytosine and Guanine pairs form with three hydrogen bonds. These hydrogen bonds work like Velcro, allowing the primer to “stick” to its complementary DNA substrate, and more bonds between the two makes the primer “stickier”; requiring more energy to prevent it from sticking. Therefore, the higher the number of C’s and G’s within a given region of DNA, the higher the melting temperature will be for its complementary primer. Their increased number of hydrogen bonds make them “stick” more easily, meaning more energy (heat) can be input to the system before the primer will be unable to anneal.

The actual primer temperature used during your PCR protocol is called the annealing temperature. It can be any temperature up to the lowest melting point temperature of the primers used in the protocol. This is important because the lower your annealing temperature is, the easier it is for your primers to anneal to the DNA substrate, but it’s also easier for the primer to anneal to mismatched regions of the DNA. This will result in your primers amplifying more regions than what is being

targeted. The higher your annealing temperature is, the more precisely your primer has to match its complementary DNA region. A primer with a higher melting temperature can be run at a higher annealing temperature, thus reducing the likelihood of non-specific annealing. In general, an annealing temperature above 50°C is ideal. Below that, non-specific annealing becomes more likely, but annealing temperatures as low as 38°C have yielded success.

To determine the melting temperature of your primer, navigate to:

<http://biotools.nubic.northwestern.edu/OligoCalc.html>

Copy the ~25bp primer sequence and paste it into the OligoCalc tool. Click “Calculate” and observe the “Salt-Adjusted Temperature.” This is the melting temperature of your primer.

Selecting primers

Universal primers

Throughout the history of molecular research, certain genes have been selected as the standard for phylogenetic informativity. Amongst the top of this list are the 16S and COI genes. These genes are somewhat unique in that they are relatively highly conserved across multiple taxa. They are widely used in research, and have had developed a set of primers that tend to yield high amplification success across wide ranges of taxa. The name and sequence of each of these primers is listed on the following page:

COI Universal Primers	
LCOI-1490 (Forward)	5' - GGT CAA CAA ATC ATA AAG ATA TTG - 3'
HCOI-2198 (Reverse)	5' - TAA ACT TCA GGG TGA CCA AAA AAT CA - 3'

16S Universal Primers	
16S L2 (Forward)	5' - TGC CTG TTT ATC AAA AAC AT - 3'
16S L9 (Forward)	5' - CGC CTG TTT ATC AAA AAC AT - 3'
16S 1472R (Reverse)	5' - AGA TAG AAA CCA ACC TGG - 3'

For someone beginning their research using these genes, these primers are a good place to start.

Review the literature

It's entirely possible that the universal primers will be insufficient for amplifying your target sequence, or you may be targeting genes other than COI and 16S. Depending on the taxa and genes you're targeting, there may already be published primers to choose from. The next step to selecting your primers is to read literature on your taxa / target genes to see if any molecular work has already been done. If so, the primers used in those studies should be published, and should be viable for your taxa as well.

Design your own primers

In the event that both the universal primers, and primers that you find in the literature fail to produce results, it's possible to design your own primers. This is a fairly simple process, but it requires some introductory bioinformatics knowledge and software. Since teaching this knowledge is outside the scope of this manual, this section might seem confusing. If you have any questions, please contact a member of

the BG Lab directly. Moving forward, it will be assumed you have this basic knowledge.

To begin designing your new primers, you must have some reference sequences of your target gene from your target taxa to compare. Ideally these sequences are from your species, but it may be the case that the only sequence data you can find are from close relatives - members of the same genus, or even family. The most common places to find these reference sequences are through online gene sequence databases such as GenBank.

The more reference sequences you can find, the more certain you can be that your primer designs are robust. Once you've acquired the reference sequences, align them so you can compare regions of the gene across all reference sequences. There are a few things you will want to look for:

1. You want to find 5' to 3' regions of 20 – 30 bases that are conserved across all reference sequences. These may be ideal regions to design a primer for, since all taxa exhibit the same sequence at these regions, increasing the likelihood of success of your primer design.

It is possible that a perfectly conserved region won't exist. In this case, it's all right to select a region with 1 or 2 ambiguities. The primer sequence will have to specify these ambiguities with the IUPAC Nucleotide Nomenclature System.

2. These regions must also be fairly rich in cytosine and guanine. This is important for setting the annealing temperature of the primer during the PCR protocol (See "**Primer melting and annealing temperature**" on page 10 for details).
3. Make sure you design a primer for both strands, and try to design them with similar annealing temperatures. To design the reverse primer you

may need to reverse compliment your reference sequences, and then flip the order of the primer design sequence to ensure it's arranged in a 5' to 3' direction.

4. Make sure the targeted region of the gene falls between the two primers, accounting for the fact that the region will be amplified in a 5' to 3' direction on both strands. Starting with the primer region, the amplified regions on each strand should overlap.
5. Give your newly designed primers cool names that will help you remember which is which, and what they are used for.

M13 tails

In the case of the BG Lab, we sequence our PCRs through a company called Genewiz. If you do the same, you will also have to submit samples of your primers. This adds a considerable amount of time and effort to the submission process, and also increases the price of the order. However, this step is only necessary if you submit different amplicon samples with different primers. If all of your samples are submitted with the same primers, the primer plate can be omitted from the order. However, because we sequence 96 samples at a time, it's often not very feasible to ensure all samples use the same primers, especially since multiple projects might be included on a single sequencing plate. To avoid the extra time and work of preparing and submitting a separate primer plate, we design all of our primers with an M13 tail attached to the 5' end of the primer sequence. The M13 tails are universal sequences that all sequencing services should recognize if they require primer samples. Their sequences are given on the following page:

M13 Forward	5' - GTA AAA CGA CGG CCA GT - 3'
M13 Reverse	5' - CAG GAA ACA GCT ATG ACC - 3'

By incorporating M13 tails in your primer designs, you can avoid having to send in a second sequencing plate of primers, which will save time, effort, and money.

Ordering primers

The BG Lab orders primers through Integrated DNA Technologies (IDT). Their website is fairly straightforward and intuitive. You'll need your primer sequence and name. Primers are also referred to as Oligos, which is how IDT refers to them. Simply navigate to their ordering page, and input your Oligo names and sequences. Alternatively, you can order your primers through our lab after approval from Dr. Bracken-Grissom.

Running a PCR

GoTaq Green

There are many different protocols available for running a PCR. The BG lab uses Promega's GoTaq Green for basic PCRs. It's simple and effective for most applications. A copy of the protocol is included as supplementary material (**GoTaqGreen_PCR.PDF**). Also available is a template document for organizing your samples and measurements for the mastermix solution. It's included as supplementary material (**GoTaqGreen_AmplificationForm_Template.docx**)

All PCR amplicons should be stored at -20°C

Verifying the success of your PCR

Gel electrophoresis

Just as with xDNA, verifying the amplification of your PCR product is done via gel electrophoresis. In fact, the protocol is very similar. The PCR Gel Electrophoresis protocol is included as supplementary material (**PCR_Gel_Protocol.PDF**).

Interpreting your PCR gel

The concept of a PCR gel is similar to that of an xDNA gel (See **xDNA Gel Electrophoresis** and **Interpreting your xDNA gel** on page 4-5 for details). The key difference is that, with an xDNA gel, you're measuring whole genome-sized DNA molecules, while with a PCR gel, you're only measuring small fragments of DNA, usually between 200 and 1,000bp long. With a PCR, you're also expecting to have a discrete amplicon of a single length, while with a DNA extraction you may observe degradation that is measured as a gradient of fragment lengths, or a smear along the length of the sample lane.

In short, when you visualize your single-gene PCR gel, you should expect to see a single, discrete band that travels farther along the gel the shorter the amplicon is. Depending on the protocol you use, you might also see one or two secondary bands of <100bp lengths. The lowest band represents the loading dye, which flows the fastest through the gel. The next highest band, usually around ~50-100bp, is your primer dimer band. Unless every primer is incorporated into the amplicon copies, they primers themselves might anneal to each other, forming double-stranded DNA fragments. These two bands are harmless and will be ignored moving forward. Just remember they may be present, but this manual will assume they aren't.

However, this isn't what you'll see. Common issues with PCRs, and how to troubleshoot them are listed on the following pages

Troubleshooting your PCR Results

1. **Completely empty lane** – In the event that you observe a completely empty lane, this may be indicative of a few different things:
 - a. The first step to resolving this issue is ensuring there was no error in conducting the protocol; specifically making sure the xDNA was actually added to the PCR. It may sound silly, but it's

certainly possible that you might have forgotten to add the DNA, in which case it would be impossible for gene amplification to occur. The solution in this case is to rerun the PCR, making sure to add the xDNA when prompted in the protocol.

To further aid in the troubleshooting process at this step, a positive control can be added to help rule out protocol errors. This is simply the inclusion of a sample that has been successfully amplified during a previous PCR. If there is no protocol error, the positive control should amplify.

b. Next, verify the quality of the xDNA used. If the xDNA is too degraded, or too low concentration, it may be possible that the PCR failed because the primers failed to anneal to the limited and/or severely fragmented DNA. The only solution in this case is to re-extract DNA to get a better sample.

c. Another possible issue could be the annealing temperature used during the protocol. If the temperature were set too high for the primers used, the primers would have failed to anneal to the DNA, preventing gene amplification. The solution in this case is rerun the PCR using a lower annealing temperature.

2. **Smearing** – Smearing along a lane is almost always indicative of an xDNA concentration that is too high, especially if the smearing is present without an amplicon band. The cause of the smearing is due to “overcrowding” of the DNA within the sample, which prevents the primers from accessing the DNA strands, preventing gene amplification.

The smear is the high concentration of genomic DNA that has degraded to varying degrees prior to, or during the PCR process. If smearing is observed, try diluting the xDNA (1:10) and rerunning the PCR. It's possible more than one dilution may be required.

It's also possible that you may observe smearing while also observing a visible PCR band. Depending on the contrast of the band to the smear, the PCR product might be sufficient for sequencing, but smearing increases the chance that the sequence results will be “noisy” and too difficult to clean, resulting in a failed sequence. Only send out smeary sequences if the smear is very faint, or in instances of desperation.

3. **Multiple amplicon bands** – This can be caused by one of two things:
 - a. The annealing temperature is too low. The energy in the system drops so low that the primer binds with mismatched regions of the DNA, resulting in more than one region being amplified. The solution in this case is an easy adjustment to the PCR protocol, increasing the annealing temperature by a few degrees to remove the non-targeted amplicons.
 - b. Inadequate primer design, resulting in primers that match more than one region of the DNA. This is the case when increasing annealing temperatures fail to remove the secondary amplicon bands, or remove the primary band at the same time as the secondary bands. In this case, the only solution is to find / design new sets of primers (See **Selecting Primers** on pages 11-14).

4. **Amplicon band present in negative control** – This is indicative of contamination. At some point during the protocol, DNA entered the negative control tube. If it occurred in the mastermix, then it's certainly the case that every sample is contaminated, in which case your PCR results cannot be trusted. To resolve this issue, try rerunning the PCR, making no changes to the protocol. If contamination occurs again, all reagents (excluding expensive taq) and liquids must be replaced, and pipettes cleaned. If contamination persists, the taq might be contaminated. Replace the taq, thoroughly clean EVERYTHING, and try again.

Organizing your PCR gels

The organization of your PCR gels should be a similar process to that of your xDNA gels. Just as you name each xDNA gel, you should also name each PCR gel. The naming system is also very similar. The BG Lab uses the following:

PCR_[YourInitials###]_[DateOfPCR]

For example, if John Doe Smith runs 5 PCRs on random dates throughout 2018, they would be named as follows:

PCR_JDS01_01Jan2018

PCR_JDS02_25Jan2018

PCR_JDS03_13Feb2018

PCR_JDS04_05Mar2018

PCR_JDS05_17Aug2018

Each of these PCRs should be recorded in separate spreadsheets, including the name of the PCR, an image of the PCR gel, the voucher numbers and species names for

each specimen included in the PCR, the primers used, and any other information that might be relevant to your specific project. A spreadsheet example template is included as supplementary materials (**PCR_HBG01_01Jan2018.xlsx**).

PART 3: SEQUENCING DNA

As previously mentioned, the BG Lab sequences our PCR amplicons through a company called Genewiz. If you are sequencing your samples through another service, skip to **Cleaning your sequences** (pg 22). For basic, single-gene PCRs, Genewiz utilizes a process called Sanger Sequencing. A detailed explanation of this process is beyond the scope of this manual, but can easily be researched if you're curious. The only necessary information you need regarding the Sanger Sequencing process is that samples are sequenced on a plate, 96 samples at a time. That means, we only sequence our PCR amplicons when we have 96 of them to submit for sequencing.

Multiple people can fill the plate across multiple projects, however, for non-members of the BG Lab, it will be necessary that you be capable of submitting your own separate plate to avoid possible contamination of BG Lab samples, unless prior approval is granted by Dr. Bracken-Grissom. It is possible to submit partial plates, but the cost remains the same as if it were a full plate. At our current quoted rate, the BG Lab spends \$270 per plate. For non-BG Lab members, it is possible to sequence your samples through our lab at that price. If this is the best option for you, please contact Dr. Bracken-Grissom, or her lab manager to negotiate the details of the arrangement.

Preparing your samples

To keep track of which samples go on which sequencing plate, the BG Lab creates sequencing plate spreadsheets, or "reference plate maps." These are strictly for maintaining organization, especially if there are multiple projects being incorporated onto a single plate. It's imperative that, upon successfully amplifying a gene for sequencing, it be added to the reference plate map. This helps to ensure your samples make it onto the plate, but also helps to keep everyone updated on how much space is available, and roughly how long it will be before the next plate can be

submitted for sequencing. A template of a reference plate map is included as supplementary materials (**Ref_Plate_HBG01_01Jan2018.xlsx**)

As soon as the 96-sample mark is reached, the next step is to place the order for the sequencing service on the Genewiz website. Unfortunately, the ordering process can be somewhat confusing. If you do run into issues, you can contact the BG Lab manager, or Genewiz customer support. You will also need two additional spreadsheet files, which you must fill out yourself. They are included as supplementary materials (**SampleMap_HBG01_01Jan2018.xlsx**, and **Container_HBG01_01Jan2018.xlsx**). Also included as supplementary materials is the ordering protocol (**GenewizOrder_Protocol.PDF**)

After the order is placed, the amplicons need to be prepared for shipment to Genewiz. A protocol is included as supplementary materials (**GenewizShipping_Protocol.PDF**).

Cleaning your sequences

Once the order is submitted to Genewiz, it generally takes around 1 week before you will receive your results. These are given as raw sequence reads, which require some cleaning before they can be used. To clean them, the BG Lab uses a powerful software called Geneious. While it is very powerful, it also has a moderate learning curve. A protocol is included to walk you through the necessary steps to clean your sequences (**Cleaning_Sequences.PDF**), but it's also advised that you refer to the Geneious user manual, which can be found online, to help learn what the software has to offer.

Section 4: ORGANIZATION!

This part of the manual is very important. As mentioned earlier, the easiest way to make mistakes, and create unnecessary frustration is through poor organization. Hopefully, by this point in the process, you're realizing how many things there are to keep track of. Between knowing what DNA has been extracted, the quality of your extractions, which samples you've PCR'd, which PCRs worked and which failed, how many you've sent out for sequencing, and how many were successful.... There are so many steps throughout where a mistake can be made. Therefore, you need a way to effectively organize your data. Included as supplementary materials is a template spreadsheet meant to help with this (**MASTER_SHEET.xlsx**).

In its default configuration, it is partially filled with mock data to help you understand how each column should be filled in. In general, the organization is as follows:

The far left columns are for sample identification, including voucher ID numbers, and taxon information.

The next group of columns includes the DNA extraction information, such as which extraction protocol was used, the xDNA concentration and remaining volume, and which gel the extraction was run on.

The next group of columns contains PCR and sequencing information. Each gene consists of two columns; 1 for the PCR results, and one for the Sequencing results. Each cell is filled with either the PCR name or the sequence plate name, depending on the column it's in, and is color-coded based on whether the PCR or sequence succeeded or failed. Green cells represent successful PCRs, while red represent failed

PCRs. Blue cells represent successful sequence results, while red (with white text) represents failed sequences.

The next group of columns contains tissue and xDNA locations, and brief metadata information about where the specimen was collected from. You can include more information here, like the date it was collected, or anything else that is relevant to your project.

Finally, the last column is for any other relevant notes specific to your project.

As mentioned, the master sheet is meant to keep every aspect of your project organized in a way that you can readily determine which samples have been worked with. It is the template we follow in the BG Lab, but certainly feel free to adapt and modify it for your personal needs.

Section 5 – Contacts

Dr. Heather Bracken-Grissom – Principle Investigator

Office Phone: 64190

Email: heather.bracken@gmail.com

Dr. Danielle DeLeo – Postdoctoral Associate

Office Phone: 64109

Email: ddeleo@fiu.edu

Charles Golightly – Lab Manager

Cell Phone: 786.659.3795

Email: cgoli005@fiu.edu