CHAPTER FOUR

DNA Extraction

Regardless whether a researcher wants to sequence a single locus or an entire genome the usual first step toward the acquisition of phylogenomic data is to perform what are called *DNA extractions*. DNA must be isolated and purified from the cells of animals, plants, fungi, or microorganisms before it can be used in downstream applications such as PCR, NGS library preparation, and DNA sequencing. That means that all RNA (usually), proteins, salts, and other cellular components must be separated from the DNA (genomic and organellar). When the process is complete, the purified DNA is usually in an aqueous buffer solution that is compatible with downstream enzymatic-driven applications. Because the process of DNA extraction involves the use of chemicals and specialized laboratory equipment, the work must be done in a molecular laboratory and appropriate safety procedures followed. As we will see in this chapter, the time required for doing DNA extractions ranges from less than an hour to several days depending on extraction method used and the nature and number of the genetic samples. The cost per sample will also depend on the type of extraction method to be used. In general DNA extraction "kits" represent the simplest and fastest methods for extracting DNA but they also cost significantly more than methods based on "old fashioned" molecular biology.

The goals of this chapter are to provide an overview of the common methods for extracting DNA, discuss the advantages and disadvantages of each method, and consider some troubleshooting strategies. Although I discuss the basic steps involved in these methods for DNA extraction, I do not provide the actual protocols that should be followed in the laboratory. Thus, the reader should consult actual step-by-step molecular biology recipes and protocols found in the literature or in the instructions provided with commercially available DNA extraction kits. Kit manuals often contain helpful technical advice on the DNA extraction process, which can further aid the researcher in better understanding the methodology involved and hence improve the quality of extraction results, but often do not sufficiently describe the chemical processes. A discussion and detailed step-by-step protocols for some older (but still very useful) DNA extraction methods can be found in Hillis et al. (1996).

4.1 DNA EXTRACTION METHODOLOGY

Until the late 1990s, molecular biologists relied upon the *phenol–chloroform–isoamyl alcohol* or "PCI" method as it is sometimes called. Though effective, this method consisted of a laborious process involving noxious organic solvents and requiring the use of a chemical fume hood. However, the introduction of molecular biology kits, just prior to the dawn of the genomics era, represented a major innovation that greatly simplified the DNA extraction process. Moreover, these extraction kits offered an additional advantage—freedom from the chemical fume hood. Hazardous volatile chemicals (e.g., phenol) are no longer necessary for high-quality DNA extractions. Most work involving an extraction kit is done using only pipettes, a vortex mixer, and microcentrifuge.

4.1.1 Summary of the DNA Extraction Process

Although the various extraction methods differ from each other in terms of specific procedures

and reagents used, they involve the same basic sequence of steps: (1) *lysis of tissues/cells and degradation of proteins*; (2) *degradation of RNA* (can be optional); (3) *isolation of DNA from proteins*; (4) *desalting the DNA*; and (5) *resuspension of DNA in storage buffer solution*.

Step 1: Lysis of Tissues/Cells and Degradation of Proteins This initial step involves the breaking down of tissues into cells, the lysis of those cells to release the DNA into a common "soup," and the enzymatic destruction of proteins. A detergent-containing buffer lyses the cells while the broad spectrum protease *Proteinase K* destroys most proteins. *Proteinase K* is used for two reasons: (1) some structural proteins must be destroyed in order to facilitate the breaking down of the tissues and (2) DNA-degrading enzymes such as nucleases, endonucleases, and exonucleases must be inactivated to prevent DNA degradation. Performing this starting procedure is simple and can be done in less than an hour. For example, vertebrate tissues are processed in the following manner: the lysis buffer + *Proteinase K* are first pipetted into a 1.5 mL microfuge tube, which contains the tissue sample; the tube is then heated to around 55°C until the tissue has dissolved in the solution. The heat-incubation time required ranges from several hours to several days (or more) depending on the size and nature of the tissue sample. The step is deemed complete when a solid tissue sample is no longer visible. If the starting tissue was pigmented, then the solution will change color as pigments are extracted from the cells. These do not affect subsequent DNA purification. This step is carried out in almost identical fashion among most DNA extraction methods. The main difference is due to whether the researcher uses buffers that were "homemade" in the lab or are components of a kit.

Step 2: Degradation of RNA The enzyme *Ribonuclease A*, more commonly called *RNase A*, is another enzyme that is often added to the initial lysate with the purpose of destroying all RNA molecules. Prior to implementing this step, careful thought should be given as to whether it is absolutely necessary to destroy the RNA in a DNA extract. There are two important considerations: (1) *RNase* is a very expensive reagent and thus it may not be needed if the primary downstream goal is to use PCR and Sanger sequencing and (2) the RNA component might be needed at a later time for another application such as constructing a cDNA library. However, in many types of NGS-based studies, the RNA component is often eliminated from a DNA extract before starting the construction of a sequencing library otherwise it can confound attempts to estimate the concentration of double-stranded DNA (dsDNA) in a sample. The decision to skip the *RNase* step during the DNA extraction process is not critical as RNA can later be removed via *RNase* treatment followed by DNA precipitation.

Step 3: Isolation of DNA from Proteins The next major step during the extraction process consists of separating the DNA from the other cellular components, especially proteins. Different extraction methods achieve this task by different means. A very popular DNA extraction method known as "salting out" (or "saline method") by Miller et al. (1988), uses a saturated sodium chloride solution (5–6M) to precipitate the proteins from the solution, while the DNA remains in an aqueous supernatant following centrifugation. Ice-cold ethanol (≥95%) or isopropanol (pure) is then added to the aqueous solution and mixed. The presence of salt and alcohol causes the DNA to lose its ability to remain dissolved in water thereby causing it to precipitate. After mixing the solution, the tubes are placed into a −20°C freezer for a minimum of 2 hours (preferably overnight). Note that the longer the time spent in the freezer the greater the DNA yield, up to a point—most of the DNA will have precipitated after an overnight precipitation period and thus longer waiting times are unnecessary. However, as this is often a convenient stopping point, the tubes can be left in the freezer indefinitely until the researcher is ready to resume the extraction process with the next step. After the tubes are removed from the freezer, they should immediately be placed into a microcentrifuge where they are spun at very high speed (~14,000 rpm) for ~5–15 minutes. This causes the DNA to form a solid pellet at the bottom of the tube, which, in a good light, can sometimes be visible to the naked eye. Some commercially available kits (e.g., Wizard® Genomic DNA Purification kit, Promega) are largely based on the salting out principle.

A second method for separating the protein and DNA components, which is used by many kits, involves mixing proprietary buffers from the kits with the lysate before adding the solution to a plastic spin column containing membranes with silica, glass, or diatomaceous earth (e.g., QIAquick® PCR Purification kit, Qiagen®). DNA

binds to the membranes at a low pH, and tubes are centrifuged at high speed in order to draw the liquid through the column and into a waste liquid collection tube. The bound DNA is retained while the proteins and other unwanted materials to pass through into the waste tube.

The traditional PCI method, which uses phenol to separate the proteins from the DNA, is accomplished in the following manner. First, after adding phenol to the aqueous (lysate) solution and mixing, the solution is centrifuged at high speed (~14,000 rpm), which stratifies the solution into two roughly equal layers in the microtube: a clear aqueous layer above that contains mostly DNA and salts and an "organic" layer below that contains mostly phenol, proteins, and other cellular debris. The aqueous supernatant is saved while the organic layer is discarded into appropriate waste containers inside a chemical fume hood. Remember that phenol is a very hazardous chemical and should be handled carefully following all safety precautions. This phenol addition step is repeated at least one more time before a similar step is performed using chloroform. With the retained aqueous solution, a small volume of a salt solution (e.g., 2M NaCl or 3M NaOAc pH 5) is pipetted into the solution and mixed followed by the addition of ice-cold 95% ethanol. This is the same DNA precipitation step described earlier. The tubes are placed in the −20°C freezer for at least one night before pelleting the DNA with centrifugation.

Step 4: Desalting the DNA Regardless of the extraction method used, the DNA pellet recovered via precipitation or spin columns will contain excess salts. If not removed, these salts can interfere with various downstream enzymatic procedures such as PCR. The salts therefore need to be "washed" from the DNA pellet. In this procedure, a fairly substantial volume of ethanol (e.g., ~500 μL of 70%–80% ethanol) is *gently* pipetted into each tube, which effectively dilutes away the salts. If using spin columns, the ethanol and salts are removed via a centrifugation step after the ethanol is added. Other methods simply require the user to pour the ethanol out of the microcentrifuge tubes or use the pipet to *carefully* remove the supernatant. When pouring or pipetting the ethanol from the tube, extreme caution should be exercised to avoid losing the DNA pellet, which can at any time become dislodged from the bottom of the tube—and hence lost. Touching the pellet with the pipet tip can also result in a lost pellet. DNA extraction kits usually require the user to purchase their own stock of ethanol for purposes of washing the recovered DNA. Although a single wash may be sufficient to desalt the DNA, two wash steps are usually recommended.

Step 5: Resuspension of DNA in Storage Buffer Solution The final procedures in DNA extraction are to (1) dry the DNA pellet to remove residual alcohol, which can interfere with downstream enzymatic applications and (2) resuspend or elute the DNA into an appropriate storage buffer solution for DNA. When using spin column kits, the drying step is simply carried out via a brief period of centrifugation. Afterward, an "elution" buffer is pipetted onto the membrane in the spin column, which releases the bound-DNA from the membrane. A final centrifugation step causes the DNA and buffer solution to pass into a clean collection microtube, which completes the extraction process. In the salting out or PCI protocols, drying of the pellet can be accomplished via air-drying the open and inverted tube overnight or by using a vacuum centrifuge *with no heat* for ~10 minutes depending on how much residual alcohol remains. After the pellet is dry, a DNA storage buffer is pipetted into the tube and allowed to dissolve the DNA pellet.

4.1.2 A Note about DNA Storage Buffers

As we just saw, the final step in the extraction process consists of resuspending or eluting the DNA into a "storage buffer" solution. Historically, this buffer consisted of 1× strength Tris–EDTA or "1× TE" buffer (i.e., 10 mM Tris–HCl, 1 mM EDTA, pH 8.0). The Tris component acts as a pH buffer keeping the DNA in a slightly alkaline environment, which helps to inhibit DNA-destroying DNases. The EDTA also acts as a preservative but does so by chelating metal cations such as Mg^{2+}, which are cofactors for most nucleases (Carter 2000). Although 1× TE is a good preservative for DNA, a drawback is that EDTA can interfere with DNA polymerases during PCR if too many of the Mg^{2+} cations are chelated (DNA polymerases also require these cations to function). This problem, however, can be overcome if the researcher takes the time to optimize the concentration of MgCl buffer in PCR, which would then allow the DNA polymerases to function well despite the presence of EDTA. However, this procedure adds additional optimization work to PCR. Another strategy to

improve the performance of enzymatic reactions involving your extracted DNA is to use a buffer containing a lower concentration of EDTA such as a 0.5× TE. Some DNA extraction kits use 0.5× TE as a DNA elution buffer (e.g., QIAquick, Qiagen). Some researchers even use a 0.1× TE, which is also called "ultra-low TE." However, keep in mind that as the strength of TE is lowered the DNA extract becomes more suitable to PCR and more vulnerable to destruction by nucleases. This is an important tradeoff. The researcher has two options that can accommodate the short-term requirements of PCR and the long-term storage of stock DNA. First, you can use 1× TE and optimize the Mg^{2+} in PCR. Instructions that accompany commercially available PCR polymerases often provide guidance on how to optimize the amount of Mg^{2+} (via a MgCl buffer solution) in a PCR reaction. Alternatively, you can mix up some DNA in low strength TE aliquots (0.1–0.5× TE) and maintain the primary DNA extract in 1× TE strength for longer term storage. For this latter strategy, it may be preferable to initially elute (or resuspend) the DNA in a lower volume of 1× TE in order to increase the DNA concentration for the long-term storage stock. This will facilitate the making aliquots of desired concentrations for PCR (Chapter 5) or NGS library construction (Chapter 7).

Many researchers have simply used molecular biology grade water for making aliquots for PCR or storing the extracted DNA. While such samples are ideal for PCR, they are not safe for long-term storage—and not just because of the threat of nuclease destruction. Water is not a good storage medium because it will facilitate degradation of the DNA sample. Thus, the DNA should be stored in a buffer solution. A number of additional methods might prove effective for long-term storage of DNA including (1) using 1× TE and freezing the sample at −80C; (2) using 95%–100% ethanol and freezing at −20 or −80C; and (3) drying the DNA before freezing at −80C. Unfortunately, no clear consensus seems to exist yet as to which storage method works best therefore the best course of action may be to divide a sample into different aliquots and try various methods (Yates et al. 1989). In other words, this is the "don't keep all your eggs in a single basket" strategy. More research is clearly needed for this very important topic.

Chemicals such as nucleases and water are not the only agents that can degrade DNA. Ultraviolet (UV) light as well as repeated episodes of freezing and thawing of extract samples will also degrade the DNA into smaller fragments. Thus, following a DNA extraction procedure, a "working stock" aliquot of each DNA sample should be made so that the number of freeze–thaw cycles for a given DNA sample can be minimized. New aliquots can be taken from the original extracts whenever DNA is needed for a procedure. This approach also makes it easy for the researcher to dilute a 1× TE-preserved DNA sample to a 0.1–0.5× concentration thus making the sample ideal for PCR or other enzyme-driven procedures.

4.1.3 Extracting DNA from Plants, Fungi, and Invertebrates

The previously discussed DNA extraction procedures are effective for many types of tissues especially those from vertebrates (my main taxonomic group of interest). However, it is well known that tissues from plants, fungi, invertebrates, and other organisms may require a modified protocol in order to obtain purified DNA samples. Readers who are interested in learning more about these special DNA extraction procedures for these organisms should consult the materials and methods sections of recent research articles on a particular group of interest in order to determine which procedures or kits are being used. Also see Palumbi (1996) for advice and DNA extraction protocols that deal with difficult-to-extract tissues.

4.1.4 Extracting DNA from Formalin-Fixed Museum Specimens

Tissue samples obtained from formalin-fixed specimens also require specialized DNA extraction methods. This problem represents one of the greatest challenges in molecular phylogenetics and phylogenomics because millions of archived specimens in natural history museums and other research collections may contain DNA suitable for sequencing and hence are of potentially great importance to tree of life studies. For example, most herpetological and ichthyological specimens are initially fluid-preserved in a 10% buffered formalin solution for a period of time lasting days or weeks. However, because formalin is extremely toxic, these specimens are subsequently transferred to a 70% ethanol solution for long-term storage. Although this preservation technique is important for preserving specimens,

once specimens are fixed in formalin it becomes exceedingly difficult or impossible to extract useable DNA from their tissues. Even if DNA is isolated from formalin-fixed tissues, depending on how long the specimen was fixed and other factors, there is no guarantee that the extracted genetic material will be amenable to DNA sequencing.

Formalin adversely affects DNA in a number of ways including (1) forming cross-linkages between DNA bases and between DNA and proteins; (2) altering the nucleotides; and (3) fragmenting the DNA (Tang 2006). If specimens are preserved in unbuffered formalin (which is acidic), then the formalin will become oxidized into formic acid. This, in turn, results in substantial DNA degradation including depurination and fragmentation (Tang 2006). Fortunately, the use of unbuffered formalin to preserve natural history specimens seems to be a long discontinued practice.

Formalin-induced cross-links between proteins and DNA cannot easily be broken and thus it is difficult or impossible to isolate DNA using conventional DNA extraction methods (Shedlock et al. 1997). In order to extract DNA from formalin-preserved specimens, tissue samples must first be subjected to laborious and time-consuming "washings" in special buffers and heat treatments in order to break down cross-linked DNA–protein complexes in an effort to free DNA (Shedlock et al. 1997; Tang 2006). However, even if some DNA is obtained this way, the yield of high molecular weight DNA is usually low and is fragmented into pieces no larger than several hundreds of base pairs (Shedlock et al. 1997). Still, some studies have successfully obtained DNA sequence data from formalin-fixed museum specimens.

Shedlock et al. (1997) described a modified DNA extraction method for obtaining DNA from formalin-fixed fishes. Included in their study were 12 deep-sea fish specimens of varying condition and with in-jar storage times ranging from several years to more than 85 years (a specimen collected in 1909). Moreover, the study only focused on obtaining template DNA suitable for amplifying a 470 bp portion of the cytochrome oxidase b (*Cyt b*) gene and a 570 bp fragment of the 16 s rRNA gene from the mitochondrial genome. Of the 34 total PCR reactions attempted, 28 (82%) were successful and one of these products (*Cyt b*) was obtained from an 85-year-old specimen. However, these authors noted that most of their successful amplifications were with specimens preserved for 25 years or less. Although their modified extraction procedure is laborious and requires several days to complete, this study nonetheless produced encouraging results showing that it is possible to at least obtain PCR-amplifiable mtDNA even from specimens nearly a century old.

An unfortunate aspect of preserving specimens in formalin is that extracted DNA consists of fragments that are less than a few hundred bases long. However, a recent study by Hykin et al. (2015) suggested and showed that the short-read NGS technology used by the Illumina platforms (Chapter 7) is perfectly suited to such degraded DNA samples. These authors presented a modified DNA extraction protocol for formalin-fixed reptile tissues and tested it using the Illumina platform. Their study included *Anolis* lizard specimens with preservation ages of ~30 years and ~100 years. In addition, they subsampled different anatomical parts of each specimen (i.e., liver, leg muscle, and tail tip) and used different extraction methods (i.e., PCI and a Qiagen kit) in order to determine which tissue types and extraction methods are best suited for their protocol. Their results showed that potentially useful amounts of DNA were only obtained from the 30-year-old *Anolis* specimen and only when PCI was used. An analysis of the fragment-size distributions revealed that the main peak fragment size for the 30-year-old specimen was centered over the 200–300 bp range. Interestingly, although the 100-year-old specimen yielded a lower concentration of DNA, results indicated that the peak fragment size for this extract was in the 200–400 bp range suggesting it might be possible to obtain sequence data from very old museum specimens. However, Illumina sequencing of these two samples only produced sequence data for the 30-year-old specimen extract—including enough data to fully reconstruct its entire mitochondrial genome. Hykin et al. (2015) pointed out that further refinements of their protocol including the possible use of alternative library-making methods could lead to better results.

4.2 EVALUATING THE RESULTS OF DNA EXTRACTIONS

After the extraction process is completed the next task is to check the *quality*, *concentration*, and *purity* of the extracted DNA samples. This step will

provide valuable information about the quality of your tissue samples, the quality of your DNA extraction procedures, and allow you to appropriately dilute these often highly concentrated DNA samples in ways that allow for consistent PCR or NGS sequencing library results while conserving stock DNA. Three approaches for evaluating DNA samples include: *agarose gel electrophoresis*, *UV spectrophotometry*, and *fluorometry*. As we will see, each of these methods has a particular strength and thus together they can provide critical information about DNA extracts.

4.2.1 Agarose Gel Electrophoresis

Agarose gel electrophoresis has changed little over several decades owing to its simple, inexpensive, and reliable nature. For work involving the sequencing of DNA, researchers use this method for several important purposes: (1) to assess the quality (i.e., degree of degradation) and concentration of DNA obtained from an extraction; (2) to assess the quality of PCR results (Chapter 5); and (3) to assess the quality of an NGS sequencing library and for size-selecting DNA fragments for NGS (Chapter 7).

The equipment for agarose gel electrophoresis consists of two main parts: a rectangular box containing an agarose gel slab submerged in a buffer solution and a power supply unit (Figure 4.1). When the power is switched on an electrical current passes through the gel. Because DNA is a negatively charged molecule, the current will mobilize the DNA fragments causing them to travel through the gel from the negative side of the box to the positive side. A key principle of electrophoresis is that the linear DNA will travel through the gel matrix at a rate proportional to its size with shorter DNA molecules traveling at a faster rate than longer ones. Agarose gel electrophoresis is an invaluable tool for fractionating DNA of different sizes, which has important applications for checking the quality of DNA extractions, PCR, and DNA sequencing.

The agarose gel electrophoresis procedure is comprised of a series of steps, which includes: (1) *casting the gel*, (2) *loading samples into the gel*, (3) *running the gel* (i.e., *the actual electrophoresis step*), and (4) *photographing the gel*. The entire procedure requires 2–3 hours of time depending upon actual electrophoresis settings (e.g., voltage). We will now review these steps.

Step 1: Casting the Gel The first step of agarose gel electrophoresis begins when the researcher heats a 0.7%–1% agarose solution (~50–75 mL) in a microwave oven for a minute, then lets the boiling-hot liquid cool down before pouring into a casting dish and adding the combs. The comb creates the individual sample wells in the gel. Note, that if the agarose solution is too hot when poured into the casting dish then the dish itself can be damaged through warping. On the other hand, letting the liquid cool too much prior to pouring is not good either because the agarose will then start to solidify before it is poured into the dish thus making a lumpy gel. Thus, an ideal gel is made with a "warm" agarose solution—one that is still in liquid form yet has a viscous consistency. For purposes of checking the results of a DNA extraction, a 0.7%–1% gel is ideal because less concentrated gels are too delicate to handle (i.e., they break into pieces too easily when handled) and more concentrated gels are not necessary. Using a more concentrated gel means you are only wasting agarose, which is an expensive

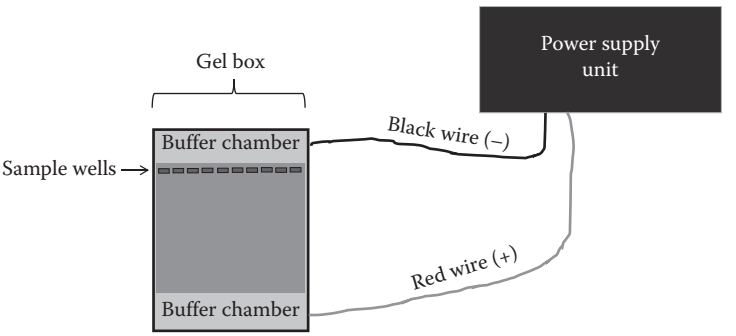

Figure 4.1. Agarose gel electrophoresis apparatus consisting of a gel box and power supply unit (top view). The gel (dark gray) sits between the two buffer chambers.

reagent. In Chapter 5, we will discuss using more concentrated gels to evaluate the results of PCR. More concentrated gels provide better resolving power for discriminating fragments of DNA by their sizes. Thus, a 2% gel is preferred for determining the sizes of PCR products, whereas a 1% gel is sufficient for evaluating the results of DNA extraction. After approximately 30 minutes the gel cools and hardens (much like fruit-flavored gelatin-type desserts) at which time it is submerged in a gel box containing a buffer, which is usually 1× TAE (Tris–Acetic Acid–EDTA) or 1× 0.5× TBE (Tris–Boric Acid–EDTA; Sambrook et al. 1989).

Step 2: Loading Samples into the Gel Next, a sample of each DNA extraction (usually ~5 µL) is mixed with a loading buffer that includes sucrose, glycerol, and/or ficoll along with one or multiple negatively charged dyes (usually ~1 µL of a 6× solution) before being pipetted into each open well in the gel. The purpose of the loading buffer (sometimes called "loading dye") is to help the DNA sink into the gel well (otherwise the DNA will float and disperse itself) and provide a visual marker in the gel to confirm that the electrophoresis process is running smoothly (i.e., colored bands can be seen migrating across the gel under white light after, say, 5–10 minutes of running). Note that these dyes are visualized in normal light and thus they do not allow for the visualization of the migrating DNA. A separate nucleic acid stain, which fluoresces under UV or colored light, must be used to actually see the DNA in a gel. In addition to the sample lanes in the gel, one well is reserved for a sample of "ladder" or (sometimes called "molecular ruler"). The ladder is a solution comprised of a set of *linear* dsDNA fragments of known size (e.g., 100 bp increments), which migrate across the gel at a rate proportional to their length. Although the ladder acts as a reference allowing you to judge the approximate sizes of DNA bands in the gel, another major function of the ladder is to provide *a positive control* for DNA in the gel. In other words, this also is a control for the electrophoresis process.

Step 3: Running the Gel Once the gel is loaded with the samples, electrical current flows through the buffer solution containing the gel, which "pulls" the negatively charged DNA through the porous interstitial spaces of the solidified agarose. Gels are typically run for 20–60 minutes depending on the voltage setting (e.g., 90 volts). When electrophoresis is complete, for safety reasons, don't forget to switch off the voltage box *and* disconnect the gel box from the power source by unplugging the red and black wires.

Step 4: Photographing the Gel In order to visualize the DNA in a gel, the investigator must first incorporate into the DNA some type of nucleic acid "stain" that fluoresces under UV light. The nucleic acid stain can be applied to the DNA in several different ways. In practice, stains have been applied to gels either before or after electrophoresis. If applied before electrophoresis, then the stain is mixed into the molten agarose prior to casting. If applied after electrophoresis, then the gel is submerged in a box containing a solution of TAE (or TBE) buffer + nucleic acid stain solution followed by a 10–30-minute waiting period, the length of which depends on the type and strength of the staining chemical. When the stain comes into contact with the DNA either during electrophoresis or while the gel is submerged in the buffer and stain box, the stain molecules bind to or become intercalated into the DNA. A major advantage to staining gels after electrophoresis is that investigators can avoid contaminating the entire electrophoresis apparatus and workspace with a potentially toxic stain; instead, a plastic dish, which can be placed inside a chemical fume hood, is used for staining gels. Some commercially available stains allow you to premix a nucleic acid stain with normal 6× loading dye. Thus, the stain comes into contact with the DNA samples at the time immediately before pipetting samples into the wells of a gel. Always consult the stain manufacturer's instructions for safe proper use.

What exactly are these nucleic acid stains? For many years, ethidium bromide (EtBr) was the primary staining reagent for DNA. However, because EtBr has long been presumed to be extremely toxic (Sambrook et al. 1989; also see Lowe 2016 for a different viewpoint), putatively safer alternative nucleic acid stains have become commercially available. Regardless whether or not these nucleic acid stains pose health hazards, common sense dictates that researchers always handle these chemicals in as safe a manner as possible. At the very least, researchers working with these chemicals should wear protective gloves and safety goggles, minimize contamination of other equipment and surfaces, perform this work in a chemical fume hood, and dispose of chemical waste in an institution-approved manner.

After staining and electrophoresis (or vice versa), the investigator places the gel on a blue light box or UV transilluminator box so that the results can be visualized and photographed for later reference. UV light will cause injuries to eyes and skin. Therefore, always wear appropriate eye and skin protection and use a shield when examining a gel on an UV transilluminator box. DNA stains such as SYBR Green or Gel Green can be visualized on a blue light box through an orange filter minimizing concerns of UV exposure. Gel photos represent invaluable records of each "experiment" and therefore should be printed, labeled, and included in laboratory notebooks. Figure 4.2 is a gel photo showing a successful set of DNA extractions. When examining a gel photo the main things to look for include the presence of any DNA and its level of degradation. If a fresh tissue sample was used in an extraction, then the resulting gel image should show two major bands of DNA in the gel: *genomic or chromosomal DNA* indicated by a band that is inside or just outside the well; *mitochondrial DNA* shown as the next discrete band (~16 kb); and some degraded DNA (mixture of nuclear and mitochondrial DNA) shown as a smear of smaller sized fragments (Figure 4.2). If the DNA was not treated with *RNase*, then you may also see a high concentration of RNA fragments, which appear as a bright band on the gel. See Sambrook et al. (1989) for discussion of agarose gel electrophoresis and protocols.

4.2.1.1 Troubleshooting

The results of electrophoresis not only provide important information about the quality and concentration of each DNA sample, but each gel image can also contain invaluable clues about the causes of any failures relating to the DNA extraction and electrophoresis processes. For example, if no DNA—not even the ladder—is visible on the gel (i.e., a "blank" gel), then the investigator can only conclude that one or more problems arose during the electrophoresis process. When this happens, nothing can be concluded about the quality of the DNA extraction process or the samples until the electrophoresis problems are solved. Various problems can arise during electrophoresis making the visualization of any DNA difficult or impossible. For example, a common culprit is a nucleic acid stain solution that has become too weak (due to overuse and/or having been degraded by light). Running a new gel with a fresh nucleic acid stain may solve the problem. Other possible causes of a failed gel include using a buffer solution contaminated by nucleases; switching the red and black wires causing the DNA to travel in a reverse direction and into the buffer chamber (using the mnemonic "run to red" will help avoid this issue); and problems with the cast gel such as having holes in the bottom of the wells causing the DNA to travel *under* the gel instead of through the gel matrix.

In other gel photos one might see normal-looking ladder bands but no other DNA elsewhere on the gel. In these cases, the investigator can conclude that a problem arose during the extraction process because existence of the ladder bands proves that the basic electrophoresis equipment, reagents, and running conditions were normal.

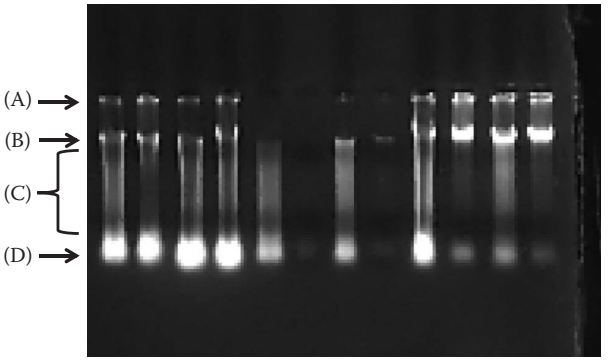

Figure 4.2. Photo of 1% agarose gel stained in ethidium bromide. DNA samples shown are from DNA extractions using the phenol–chloroform–isoamyl method. (A) Undegraded genomic DNA; (B) undegraded mitochondrial DNA; (C) degraded genomic DNA and RNA; (D) RNA. Note, loading wells are located along the top edge of the A bands. Ladder not shown.

DNA extractions can fail for a number of reasons including using defective reagents or kits, experimenter error (pipetting mistakes), and/or bad quality tissue samples.

4.2.2 UV Spectrophotometric Evaluation of DNA Samples

Another standard piece of equipment in a DNA laboratory is a tabletop device called a *UV spectrophotometer*, which is used to estimate the concentration and purity of a DNA sample. These machines operate by shining UV light through a DNA solution and then measure the light absorbance by the DNA. Since nucleic acids maximally absorb light in the UV range, the amount of light absorbed in a solution is directly proportional to the amount DNA/RNA in that sample. This property of nucleic acids enables molecular biologists to use a simple method for quantifying DNA samples and for determining their purity. There are two main limitations of UV spectrophotometry. First, UV spectrophotometry cannot discriminate RNA from DNA and thus it will only produce accurate estimates of genomic DNA concentration if the RNA component has been removed/destroyed. Thus, if a genomic DNA extract is not treated beforehand with *RNase*, then the actual concentration for the dsDNA component will be lower than the total concentration because of RNA contamination. This can be particularly important to consider when quantifying genomic DNA during NGS library construction. A second limitation is that UV spectrophotometry cannot determine whether or not a DNA sample is degraded. Thus, a sample containing only nucleotides can have the same concentration as a sample containing undegraded genomic DNA.

Spectrophotometers are simple to use. The first step is to run a "blank" before you measure your DNA sample so that a "background" absorbance value is established for your DNA solution minus the DNA. To perform the blank step, you simply run a sample of the *same* liquid your DNA is sitting in through the spectrophotometer, which will provide a "zero" concentration value. For example, if the final step of your DNA extraction kit specifies that you elute or dilute your DNA sample using a buffer provided in the kit (e.g., TE), then you should blank using the *same* buffer. If instead you performed the elution or dilution step using pure water, then you should blank using pure water. Blanking with the same solution will help ensure accuracy. The next step is to measure a sample of your DNA. If using an older model spectrophotometer, then you will likely be pipetting your sample into a quartz glass or clear plastic cuvette—the size of your small finger, whereas newer "microvolume" machines such as the NanoDrop 2000 (Thermo Scientific™) do not use a cuvette. One nice advantage of the NanoDrop is that it consumes far less sample than the older machines (1–2 μL vs. 5–10 μL of DNA). Moreover, checking the accuracy of your machine is easily done by pipetting 1 μL of a known concentration template (e.g., some 100 bp or 1 kb ladders are sold at a concentration of ~100 ng/μL).

4.2.2.1 UV Spectrophotometry to Determine Concentrations of Nucleic Acid Samples

DNA (and RNA) maximally absorbs light that is of 260 nm in wavelength. This constant is useful because the absorption of 260 nm light by dsDNA, single-stranded DNA (ssDNA), and RNA can be standardized in terms of optical density (OD) units and hence yield the following *conversion factors* (Sambrook et al. 1989):

1 OD = 50 μg/mL

1 OD = 40 μg/mL RNA and ssDNA

UV spectrophotometers typically do not perform well with highly concentrated DNA samples (e.g., >10 μg/μL). Therefore a DNA sample should be diluted 10-, 50-, or 100-fold before being assayed plus an appropriate *dilution factor* will also be needed to calculate the concentration. Using the above information we can calculate DNA concentration with the following equation:

$$\text{DNA concentration} = \text{OD} \times \text{conversion factor} \times \text{dilution factor} \quad (4.1)$$

Here is an example concentration calculation using Equation 4.1: begin with 50 μL of extracted genomic DNA. Next, dilute 20 μL of that DNA into 980 μL of water ("1/50 dilution" or a "50× dilution factor"). Next, place the diluted sample in a cuvette and determine the OD of the sample at 260 nm. Let's say the spectrophotometric

reading = 0.2 OD, then the dsDNA concentration (μg/mL) will be

$$\text{DNA concentration} = 0.2 \times 50\ \mu g/mL \times 50$$

$$\text{DNA concentration} = 500\ \mu g/mL$$

Many routine DNA methods typically involve minute volumetric measurements in microliters (μL). Accordingly, it is generally more useful to express DNA concentrations in terms of ng/μL, which is the equivalent to μg/mL. Older spectrophotometers usually measure concentration as μg/mL, whereas newer models (e.g., NanoDrop) output concentrations in ng/μL thus one does not need to manually convert OD units into concentration units. Lastly, spectrophotometers tend to not only have trouble in estimating the OD of concentrated samples, but they also tend to lose reliability with samples that are too dilute (<5 ng/μL). It is important to keep in mind that samples with very low readings—even some with a 0.0 ng/μL reading—may still contain enough DNA template to result in a successful PCR reaction—PCR is extremely sensitive as we will see in Chapter 5. Thus, these evidently DNA-free samples should not be dismissed as failed extractions and discarded. On many occasions I have observed such "zero concentration" samples yield useful PCR results and thus it is always worthwhile to try PCR anyway with such samples.

4.2.2.2 UV Spectrophotometry to Determine the Purity of DNA Samples

Ideally, a DNA sample will be free of all types of contaminants, which can interfere with downstream laboratory procedures such as PCR. Proteins, salt, and ethanol are the most common contaminants mixed with DNA following an extraction procedure. Fortunately, in addition to measuring concentration, a spectrophotometer can also measure the purity of a DNA sample. An index of sample purity called the "260/280 ratio" can be used to gauge the purity of a DNA extract (Sambrook et al. 1989). This ratio is simply produced after a spectrophotometer generates OD values at 260 and 280 nm wavelengths. Pure DNA has a ratio of 1.8 but values ranging from 1.3 to 2.1 indicate a relatively pure DNA sample that can generally perform well in PCR. The presence of proteins, which have a maximum absorbance at 280 nm will increase absorbance at 280 causing the 260/280 to lower in value, whereas the presence of ethanol, salts, or RNA will raise the ratio above 1.8. Applying careful techniques throughout the DNA extraction process can help minimize contamination problems. For example, using two separate 70% ethanol wash steps can help to fully desalt DNA samples as well as using extra care to ensure that ethanol has been evaporated from DNA pellets can help improve the purity of your DNA samples.

4.2.3 Fluorometric Quantitation of DNA Samples

Another common lab device used for DNA quantitation is the *fluorometer* such as the Qubit® (Thermo Scientific). The fluorometric approach has a couple advantages over microvolume UV spectrophotometric devices. First, a fluorometer can distinguish ssDNA from dsDNA (Life Technologies 2014). Thus, a fluorometer has the ability to accurately measure the concentration of dsDNA while ignoring the presence of other molecules such as RNA and lone nucleotides. Moreover, fluorometric-based concentration estimates of DNA are not affected by the presence of contaminants such as proteins, salts, and some organic compounds (Life Technologies 2014). Microvolume UV spectrophotometers measure DNA concentrations over an effective range from 2 ng/μL to 15 μg/μL, whereas the Qubit operates within the range 10 pg/μL to 1 μg/μL (Life Technologies 2014). Thus, the Qubit can effectively estimate the concentrations of very dilute DNA samples. Both the microvolume spectrophotometer and Qubit require small volumes (e.g., 1 μL) of samples to conduct a quantitation assay. Accordingly, the strength of the fluorometric approach is that it can provide the best estimates of dsDNA concentration—even with samples containing RNA, which is especially critical when quantifying DNA during NGS library construction. However, as fluorometric devices are unable to provide information about the quality (i.e., degree of DNA degradation) and purity of samples, agarose gels and microvolume UV spectrophotometers are able to provide this complementary information.

4.3 THE HIGH-THROUGHPUT WORKFLOW

Now that we have examined the basic methodology of DNA extraction, we will conclude this

chapter by introducing the idea of using high-throughput procedures to further economize lab work resulting in lower labor and consumables costs per sample while dramatically increasing the numbers of sequences obtained. For research projects in which researchers need to obtain many DNA sequences from many different individuals, loci, or both, the high-throughput laboratory format can be employed to great advantage. In practice, "high-throughput" simply refers to the process of using multichannel pipettes or robotic equipment (e.g., liquid handlers) to process many samples simultaneously, which can be in groups of 8, 12, 16, 96, and even 384 samples. High-throughput methods can be easily incorporated into each stage of the basic DNA extraction-template acquisition-sequencing workflow provided the appropriate equipment upgrades are made in the laboratory. These upgrades include purchasing: 8 or 12 channel pipettes, a tabletop centrifuge that can spin 96 deep well (2 mL) or 96 well (0.2 mL) PCR/sequencing microplates, a 96 sample microplate rotor for the vacuum centrifuge, and an agarose gel electrophoresis apparatus that can accommodate 96 samples of extracted DNA or 96 PCR products. Some large molecular ecology labs and especially commercial or institutional genome institutes also have robotic equipment such as liquid handling machines, each of which can cost hundreds of thousands of dollars. However, much of the time, researchers may not have enough samples to justify using such expensive equipment; a handheld 8-channel pipette will suffice for most applications. Although there is a significant up-front cost associated with these additional equipment purchases, the savings in labor and consumables over the long-term will likely far exceed the initial investment. Another exciting aspect of this high-throughput methodology is that thanks to the outsourcing of Sanger sequencing (Chapter 6) and NGS (Chapter 7), even small laboratories that adopt some or all of these procedures can scale up their production such that their *per capita* data acquisition can be comparable to the larger molecular biology labs. Moreover, as we will see in later chapters, high-throughput practices can be integrated into PCR, Sanger sequencing, and NGS template acquisition. We will now discuss the first opportunity for incorporating high-throughput into the workflow: high-throughput DNA extractions.

4.3.1 High-Throughput DNA Extractions

There are at least four opportunities for including high-throughput procedures during the process of DNA extraction: (1) using high-throughput DNA extraction kits (or other nonkit method) to simultaneously extract DNA from 96 tissue samples; (2) using high-throughput agarose gel electrophoresis to simultaneously analyze the results of 48–96 DNA extractions; (3) using high-throughput UV spectrophotometry to analyze the results of 96 DNA extractions; and (4) preparing diluted DNA template in a high-throughput format to facilitate large-scale PCR and Sanger sequencing. We will now review each of these procedures in turn. Keep in mind that although it may not be practical or affordable to follow all of these procedures, even adopting one or two of them can greatly improve the economy and output of your lab work.

4.3.1.1 Extracting DNA from 96 Tissue Samples

If you want to extract DNA from at least 96 tissue samples at one time, then you should consider purchasing a DNA extraction kit in the 96 well format. If you have access to one of the "automatic" DNA extraction machines (e.g., autogen), then the process of extracting DNA from 96 tissue samples will be even easier and almost hands free. These large-scale DNA extraction methods typically rely on 96 deep well (2 mL) blocks throughout the procedure including the final step in which the purified DNA is eluted into a clean deep well block. One note of caution when using these methods—given the large investment in terms of genetic starting material (i.e., tissue) that is at risk in case something goes wrong—you should carefully consider beforehand the desired concentration range and volume. Of course this will depend on the amount of tissue you input into the extraction process relative to the volume of buffer used to elute the DNA in the final step. Ideal concentrations for your newly extracted DNA will be 50–1,000 ng/μL. This "concentrated stock" can be stored in the freezer until new aliquots of appropriately diluted (working stock) template DNA are needed for downstream procedures (e.g., PCR and NGS libraries). What are the ideal working concentrations for these applications? The ideal working concentration will depend on the particular procedure under

consideration and thus we will wait to visit this issue when we discuss PCR (Chapter 5) and NGS (Chapter 7).

4.3.1.2 High-Throughput Agarose Gel Electrophoresis

Once a set of 96 DNA extractions has been processed, the researcher should proceed on to evaluating the quality and concentrations of all DNA samples on an agarose gel. Using a single channel pipet to load this many samples onto a single gel is not practical. Thus, using a multichannel pipet can greatly improve the speed of loading samples into a gel while reducing the incidence of errors. Anyone who has painstakingly pipetted a large number of samples—one sample at a time—into an agarose gel will appreciate the simplicity and effectiveness of loading eight samples at a time onto a gel.

Unfortunately, the agarose electrophoresis boxes and combs most often found in molecular ecology laboratories are inadequate for accommodating this high-throughput approach. Thus, a special agarose electrophoresis apparatus for high-throughput must be purchased. Be careful to obtain combs that match the spacing of the tips on the multichannel pipette.

After some practice, using a multichannel pipette can be easy. However, there are two important things to consider. First, you need to learn how much even pressure to put on the pipette while affixing pipette tips. Too little pressure and one or more tips may not form a good seal and either fall off or be incapable of holding the specified volume of liquid. That is, after drawing liquid into the tips you will see that the volumes are uneven among tips, which is not good. The second important thing to learn is how to pipette small volumes (0.5–20 μL range) without adding air bubbles to the solution. Multichannel pipettes can be difficult to use in the beginning because it can be difficult to simultaneously and correctly affix all 8 (or 12) clean tips to the pipette. If all tips are not properly affixed, then one or more tips will likely withdraw air instead of liquid reagent. Obviously adding a 0.5 μL of air instead of enzyme will not lead to a successful reaction in that tube. If you have trouble with this, then either you need to improve your technique, change the brand of tips you are using, or check to make sure the pipette is not defective or broken. Mastering a multichannel pipette is essential for high-throughput molecular biology.

4.3.1.3 High-Throughput UV Spectrophotometry

Measuring the purity and concentration of 96 extractions one tube at a time using a standard UV spectrophotometer would be arduous. Thus, for measuring a large number of extractions, it is much easier to use a 96-sample cuvette-plate UV spectrophotometer. If your DNA extractions are already arrayed in a 96-well (8 × 12) storage plate, then it will be easy to use a multichannel pipette to transfer a small volume of each sample to the wells in the cuvette plate. The plate-reading spectrophotometer can then quickly read the 96 samples and the results are then sent to a printer or saved in a spreadsheet file.

4.3.1.4 Preparation of Diluted DNA Templates for High-Throughput PCR

In preparation for high-throughput PCR, the spectrophotometric results for the original DNA extractions should be used to guide the making of appropriately diluted template in a clean 96 well (0.2 mL) microplate. Keeping in mind that the desired PCR template concentration should ideally be in the 10–100 ng/μL range, spectrophotometric data will allow you to make informed decisions on how to dilute the samples. Your original concentrated DNA extractions can be stored in the freezer for long-term storage. By having a separate plate of diluted templates, you will not only be ready to conduct PCR experiments in high-throughput fashion, but you will subject the concentrated DNA stock to fewer freeze–thaw cycles thereby better preserving the quality of those samples. The PCR template plate should be stored in the freezer when not in use. One last tip: always remember that whenever the original DNA extraction plate or the template plate are retrieved from the freezer, it needs to be spun in a centrifuge for a minute to force all the liquid to the bottom of each well, which will help minimize the risk of cross-contamination across wells and make it easier to pipet template for a new PCR. A nice alternative to using a 96 plate to hold your PCR templates, is to use 8-strip (0.2 mL) PCR tubes (i.e., 8 tubes connected together forming a strip). This approach gives you additional flexibility for managing

your template DNAs for PCR. For example, if you don't have 96 extractions to load into a single microplate, then the 8-strip tubes allow you to still use the multichannel pipette for setting up PCR reactions in a quasi high-throughput manner. Of course, 12 strip-8 tubes is the same as a 96 well microplate.

Another aspect of the high-throughput approach to consider is the quality of the consumables. One of the keys to having consistent success in the extraction and sequencing lab is to use high quality consumables. Unfortunately, variation exists among products such as pipette tips, microcentrifuge tubes, and PCR tubes. For example, all the time and money invested to prepare a microplate full of PCR products for sequencing would be for naught if the plate is leaking! Be sure you are satisfied with all the products you are using. Also keep in mind that while kits often work well enough to save time and trouble, don't exclude the possibility of making some reagents yourself or using some "old fashioned" protocols. It is remarkable that gel electrophoresis, which was developed in the 1960s, continues to be an indispensable lab method in the age of genomics. Doing some things yourself can save you much money and in some cases can provide results that are comparable, and sometimes preferable, to kits.

REFERENCES

Carter, R. E. 2000. General molecular biology. *Molecular Methods in Ecology* 6:1.

Hillis, D. M., B. K. Mable, A. Larson, S. K. Davis, and E. A. Zimmer. 1996. Chapter 9. Nucleic acids IV: Sequencing and cloning. In *Molecular Systematics*, 2nd edition, eds. D. M. Hillis, C. Moritz, and B. K. Mable, 321–381. Sunderland: Sinauer.

Hykin, S. M., K. Bi, and J. A. McGuire. 2015. Fixing formalin: A method to recover genomic scale DNA sequence data from formalin-fixed museum specimens using high-throughput sequencing. *PloS One* 10:p.e0141579.

Life Technologies. 2014. Invitrogen Technical Note, Comparison of fluorescence-based quantitation with UV absorbance measurements. Qubit® fluorometric quantitation vs. spectrophotometer measurements.

Lowe, D. 2016. The myth of ethidium bromide. In the Pipeline, blog of *Science Translational Medicine*. http://blogs.sciencemag.org/pipeline/archives/2016/04/18/the-myth-of-ethidium-bromide (accessed April 18, 2016).

Miller, S. A., D. D. Dykes, and H. F. R. N. Polesky. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16:1215.

Palumbi, S. R. 1996. Chapter 7. Nucleic acids II: The polymerase chain reaction. In *Molecular Systematics*, 2nd edition, eds. D. M. Hillis, C. Moritz, and B. K. Mable, 205–247. Sunderland: Sinauer.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual*, 2nd edition. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Shedlock, A. M., M. G. Haygood, T. W. Pietsch, and P. Bentzen. 1997. Enhanced DNA extraction and PCR amplification of mitochondrial genes from formalin-fixed museum specimens. *Biotechniques* 22:394–400.

Tang, E. P. ed. 2006. *Path to Effective Recovering of DNA from Formalin-Fixed Biological Samples in Natural History Collections: Workshop Summary*. Washington, DC: National Academies Press.

Yates, J. R., S. Malcolm, and A. P. Read. 1989. Guidelines for DNA banking: Report of the Clinical Genetics Society working party on DNA banking. *J Med Genet* 26:245.