# CHAPTER 2

## Laboratory Methods

### 2.1 Introduction

In recent years molecular techniques have been extensively used to determine phylogenetic relationships among taxa down to the level of populations. DNA is usually extracted from freshly collected specimens that have been preserved (e.g. in salt-saturated DMSO or 70-80% ethanol). Over the past years the number of higher taxonomic level phylogenetic studies has increased. It can be difficult however, to generate species-level phylogenies, as some groups are highly speciose and/or some species are difficult to obtain. The ability to use museum specimens for phylogenetic studies is therefore advantageous, especially for species-level studies. It enables sampling of species that are extinct, endangered and difficult to obtain fresh from the wild. Museum specimens may be quite old (from collections at the beginning of the 20$^{th}$ century) and have generally been fixed in formalin (10%) prior to being preserved in alcohol. Formalin-fixed tissue produces very low DNA yields and DNA is substantially degraded (Dillon et al. 1996; Wirgin et al. 1997), which has resulted in avoiding the use of museum specimens for molecular studies.

However, a few studies have described and compared DNA extraction methods from tissues preserved in various ways, including formalin-fixed specimens of tapeworms (Li et al. 2000), trout (Shiozawa et al. 1992), Atlantic coast bass (Wirgin et al. 1997),

molluscs (Chase et al. 1998), a range of other taxa (amphibian, reptile, fish, invertebrate) (Shedlock et al. 1997), and insects (Dillon et al. 1996). All these studies, with the exception of the last one, successfully extracted mitochondrial DNA from formalin-fixed samples.

This chapter describes the extraction and PCR (polymerase chain reaction) method for amplification of DNA from fresh samples. More importantly, I also describe the extraction protocol developed to obtain DNA from museum collections of reef fish for amplification of mitochondrial and nuclear markers, specifically from the genera *Naso* (Family: Acanthuridae), *Scarus* and *Chlorurus* (Family: Scaridae). The technique described here differs from previous studies, particularly with respect to extraction times and reagents used. Furthermore, I compare different PCR optimisation and amplification procedures for DNA from ethanol preserved (fresh) and formalin-fixed tissues. This study is the first to document successful sequencing of a nuclear marker from formalin-fixed samples.

**2.2 General extraction method from freshly preserved tissue**

DNA was generally extracted from tissue which had been fresh frozen (-20°C) or preserved either in 80% ethanol or salt-saturated 20% DMSO (dimethyl sulfoxide). Frozen specimens (which have been stored at -20°C for up to 10 years) were sub-sampled whilst frozen and the sub-sample placed immediately into 80% ethanol.

Standard extraction procedures were used (Sambrook et al. 1989). Briefly, a small piece of tissue was chopped into pieces, washed 2 – 3 times in TE buffer (10mM Tris pH 8.0, 1mM EDTA) for 30 – 45 min each wash. Tissue pieces were transferred to 500µl extraction buffer containing (100mM Tris-Cl pH 8.0, 1.4 M NaCl, 20mM EDTA; 2%

CTAB hexadecyltrimethyl ammonium bromide; 2% PVP (polyvinyl pyrolidone) and 20µl of a 20mg/ml proteinase K stock solution). Tissue was incubated at 55°C overnight. Samples were incubated for 30min at 90°C to inactivate proteinase K. To digest RNA, 10µl of a 10mg/ml RNAseA solution was added to each tube and samples were shaken at room temperature for 1 hr. A standard salt-chloroform extraction procedure (Sambrook et al. 1989) was used to isolate purified DNA from tissue extractions. DNA was then precipitated in 2.5 times volume of 100% ethanol and washed in 70% ethanol. DNA pellets were finally dried and resuspended in TE. Before DNA was used, it was checked on an 0.8% agarose gel to determine quality and concentration.

### *2.2.1 PCR amplification from fresh preserved tissue*

Universal primers (developed to be used across a broad range of organisms) were used to amplify selected regions of the DNA (Kocher et al. 1989; Simon et al. 1994; Lyons et al. 1997). Three gene regions, ETS2 (nuclear marker), 16SrRNA (mitochondrial large ribosomal unit) and cytochrome *b* (mtDNA protein coding region) were amplified for phylogenetic analysis (Table 2.1). The d-loop region of the control region of the mitochondrial genome was amplified for phylogeographic studies (Table 2.1).
ETS2 is an intron from a nuclear oncogene. It was originally used as a conserved mammalian single-locus DNA marker (Lyons et al. 1997), but the primers, based on flanking exonic regions, also amplify fish DNA.

Each 20µl PCR reaction volume contained 2.5mM Tris-Cl (pH 8.7), 5mM $KCl(NH_4)_2SO_4$, 200□M each dNTP, $MgCl_2$ concentrations varied from 1.5mM - 6.0mM

(Table 2.1), 10□M each primer, 1 unit of Taq Polymerase (Qiagen) and 10ng template DNA. ETS2 and cyt *b* amplifications followed the same basic cycling: initial denaturing for 2 min at 94°C was followed by 30 cycles, the first 5 cycles at 94°C for 30s, 30s at the highest annealing temperature $T_a$ (Table 2.1), followed by 1 min 30s at 72°C. An additional 25 cycles were performed, at the same denaturing and extension times and temperatures as before, but at a lower $T_a$°C (Table 2.1). 16S rRNA was amplified using touchdown PCR in 3 phases of 5, 5 and 20 cycles per annealing temperature (Table 2.1) with the same denaturing and extension temperatures and times as above. D-loop amplification was done at a single annealing temperature (Table 2.1) for 35 cycles. PCR products of 16S, cyt *b* and d-loop were checked visually on 2% agarose gels and purified by isopropanol precipitation, dried and resuspended in 20-50µl nuclease-free water. PCR products of ETS2 had to be gel-purified on 2% agarose gels as two bands (500bp and ~ 300bp sizes) appeared routinely. The 500bp fragment was excised and column-purified following manufacturer's protocol (Qiagen). Two examples of the 300bp fragment were sequenced to check if this fragment was from the same gene family. The 300bp sequences aligned with the 500bp fragment but had several large deletions.

### *2.2.2 Sequencing of fresh preserved tissue*

PCR templates were sequenced directly in both directions, using full reactions for ETS2 and 3/4 reactions for 16S, cyt *b* and d-loop. Each 20µl or 15µl reaction volume contained 8 or 6µl ABI dye terminator, 3.4pmol primer and 8-10µl PCR template. Cycle conditions for sequence reactions: initial incubation for 5 sec at 94°C, followed by 25 cycles of denaturing for 30s at 96°C, annealing for 15s at 50°C and 4 min extension at 60°C (with a

ramp rate of 1°C/sec between temperatures). Sequence products were isopropanol-precipitated as previously described for PCR product purification. Sequences were analysed on an automated ABI Prism 377 sequencer at the Griffith University Sequencing Facility.

**2.3 Extracting DNA from museum preserved tissue**

To obtain enough DNA, 1-3g of formalin-fixed muscle tissue was required. I found that 3g of tissue was needed for samples that had been preserved for more than 10-15 years, with less being necessary for more recently preserved tissue. The tissue was finely cut and equally divided into 5-6 x 2ml aliquots. Each aliquot was washed 2-3 times (each 1-2hrs, using a rotary shaker on low rpm at room temperature) in a solution containing ıml TE (10mM Tris pH 8.0, 1mM EDTA) and 0.5mM $MgCl_2$. Tissue was subsequently washed at 4°C overnight in 1ml TE without $MgCl_2$. Samples were briefly washed 2-3 times for several minutes in nuclease-free water prior to adding 1ml CTAB buffer per aliquot containing 50µl proteinase K (20mg/ml stock solution) to break down tissue and release DNA. Digestion was done overnight at 55°C in a water bath. If samples were not completely digested, another 20-30µl proteinase K was added and samples incubated for an additional 2-3 hours at 55°C. Proteinase K was denatured by heating to 90°C for 30 mins. RNAse digestion was omitted, as no RNA persists, due to the nature (age and preservation) of the tissue.

Standard DNA purification was performed, using 0.5ml phenol:chloroform:isoamyl alcohol (25:24:1) (Sigma) buffered to pH 8.0, followed by two 0.5ml per aliquot chloroform:isoamyl alcohol (24:1) extractions (Sambrook et al. 1989). DNA was

Table 2.1: Comparison of magnesium (Mg) concentrations, annealing temperatures and no. of cycles required during PCR amplification of 4 markers from museum and fresh specimens for several species.

| Primer used | Species (Museum I.D.) | Fresh sp. Mg mM conc.* | Museum sp. Mg mM conc.* | Fresh sp. annealing $T_a$ °C | Museum sp. annealing $T_a$ °C |
|---|---|---|---|---|---|
| **ETS2F**<br>5' -AGC TGT GGC AGT TTC TTC TG- 3'<br>**ETS2R**<br>5' -CGG CTC AGC TTC TCG TAG- 3'<br><br><br>(Lyons et al. 1997) | *Luvarus imperialis* (AMS I.38647002)<br>*Naso caeruleacauda* (ROM 66936, 67197)<br>*N. fageni* (ROM 68109)<br>*N. lopezi* (AMS I.37801002)<br>*N. mcdadei* (LFNMMST-62, CSIRO H3745)<br>*N. maculatus* (BPBM 24905)<br>*N. minor* (ROM 67152)<br>*N. reticulatus* (BPBM 23425)<br>*N. thynnoides* (ROM 67179,USNM 122069) | N/A<br>N/A<br>4.0<br>4.0<br>4.0<br>4.0<br>4.0<br>N/A<br>4.0 | 7.0<br>9.0<br>9.0<br>9.0<br>9.0<br>9.0<br>9.0<br>9.0<br>9.0 | 53°C - 51°C (5 + 25 cycles) | 51°C - 49°C (5 + 30 cycles) |
| **16SrRNA**: **LR-J-12887**<br>5' -CCG GTC TGA ACT CAG ATC ACG T- 3'<br>**LR-N-13398**<br>5' -CGC CTG TTT ACC AAA AAC AT- 3'<br><br>(Simon et al. 1994) | *L. imperialis* (AMS I.38647002)<br>*N. brachycentron* (USNM 205542)<br>*N. caeruleacauda* (ROM 66936, 67197)<br>*N. mcdadei* (LFNMMST-62, CSIRO H3745)<br>*N. reticulatus* (BPBM 23425)<br>*Chlorurus rhakoura* (AMS I.37146001)<br>*Scarus oedema* (WAM P.22865-01) | N/A<br>1.5<br>N/A<br>1.5<br>N/A<br>3.5<br>3.5 | 5.0<br>9.5<br>9.5<br>9.5<br>9.5<br>9.5<br>10.0 | 51°C - 49°C 47°C (5 + 5 + 25 cycles) | 49°C - 47°C (5 + 30 cycles) |
| **Cyt *b***: **L14841** 5'-AAA AAG CTT CCA TCC AAC ATC TCA GCA TGA TGA AA- 3'<br>**H15149** 5'-AAA CTG CAG CCC CTC AGA ATG ATA TTT GTC CTC A- 3'<br>(Kocher et al. 1989) | *L. imperialis* (AMS I.38647002)<br>*N. caeruleacauda* (ROM 66936, 67197)<br>*N. mcdadei* (CSIRO H3745)<br>*N. reticulatus* (BPBM 23425)<br>*N. thynnoides* (ROM 67179) | N/A<br>N/A<br>6.0<br>N/A<br>6.0 | 6.0<br>9.5<br>9.5<br>9.5<br>9.5 | 50°C - 48°C (5 + 25 cycles) | 50°C - 48°C (5 + 30 cycles) |
| **D-loop**: **NA1-F-dloop** 5' -AGC ATT CTG AAC TAA ACT AC- 3' **NA1-R-dloop** 5' -TGT CCC TTG ACT CTC AAT A- 3' (This study) | *N. tuberosus* (JLB007828, 019543, 013017, 001226) | 1.5 | 9.0 | 50°C (35 cycles) | 50°C - 48°C (5 + 30 cycles) |

Note: *****: amounts are for total Mg concentrations and include 1.5mM Mg already present in the PCR buffer. **N/A**: lack of fresh specimens. **AMS**: Australian Museum Sydney, **BPBM**: Bishop P. Bernice Museum, **CSIRO**: Commonwealth Scientific & Industrial Research Organisation, **ROM**: Royal Ontario Museum, **LFNMMST**: National Museum of Science & Technology, Taiwan, **USNM**: Smithsonian National Museum of Natural History, **WAM**: Western Australian Museum. **J.L.B**. Smith Institute of Ichthyology, Durban, South Africa.

precipitated in 2.5 times vol. of 100% EtOH and 0.05 times vol. of 3M NaAcetate at –20°C overnight. Following precipitation DNA was pelleted at 11,000 rpm for 30 mins, washed in 1ml 70% EtOH, and dried in a DNA SpeedVac (Savant) prior to being resuspended in 10µl TE (10mM Tris pH 8.0, 1mM EDTA) at 4°C overnight. DNA aliquots (in TE) from the same specimen were combined and stored at –20°C to prevent DNA from degrading. PCR amplification was performed without dilution and without checking on an agarose gel.
A 1–3g piece of muscle tissue produces sufficient DNA for PCR optimisation with 1 primer pair and subsequent sequencing reaction in one direction only. If more than one gene region is to be examined, extractions have to be repeated from additional muscle tissue. I therefore recommend that a minimum of 5–8g tissue is required in total.

### *2.3.1 PCR amplification of museum preserved tissue*

Due to the presence of EDTA in the resuspended DNA aliquots, the concentration of Mg in the PCR reaction had to be increased nearly two- or more -fold compared to DNA aliquots that have been diluted in water (see Table 2.1). A high fidelity Taq Polymerase (e.g. *ProofStart* Qiagen) had to be used for PCR amplification of DNA from formalin-fixed tissue. *ProofStart* Taq requires $MgSO_4$ rather than the generally used $MgCl_2$. PCR reactions were performed in 20µl volumes containing 2.5mM Tris-Cl (pH 8.7), 5 mM $KCl(NH_4)_2SO_4$, 300□M each dNTP, 10□M each primer, 2.5 unit of *ProofStart Taq* Polymerase (Qiagen) and 6-8µl template DNA. Additionally, $MgSO_4$ concentrations had to be re-optimised for each species and marker (Table 2.1). Annealing temperatures had to be reduced by 2°C during touchdown PCR amplification of museum specimens, compared to fresh specimens for 2 of the 3 markers used (Table 2.1). All 3 markers were amplified

under the same cycling conditions: Initial *ProofStart Taq* activation step of 5 min at 95$^{o}$C, prior to 5 cycles of denaturing for 30s at 94$^{o}$C, 30s at the highest annealing temperature (Table 2.1) followed by 1 min 30s extension at 72$^{o}$C. An additional 30 cycles are performed at the same denaturing and extension times and temperatures as before, but at a lower annealing temperature than was used initially (total of 35 cycles) (Table 2.1). The total volume of PCR product was gel-purified from 2% agarose gel, followed by column purification as per manufacturer's instruction (QIAquick Gel Extraction Kit, Qiagen). Gel-purified PCR products were then precipitated and dried as previously described. Dried pellets were resuspended in 10□l of nuclease-free water and immediately used for direct sequencing in one direction using dye terminator chemistry (ABI), following manufacturer's instructions.

Sequence reaction were done using the same protocol as previously described, but only full reactions were performed.

**2.4 Comparative quality of museum sequences and recommendations**

Using the methods described (section 2.3), DNA was successfully extracted, amplified and sequences obtained for all species and markers, including a nuclear marker (Table 2.1). Obtaining sequences from the full fragment of the mitochondrial DNA had a higher success rate than obtaining complete sequences from the nuclear marker. All sequences obtained for mitochondrial and nuclear markers from formalin-fixed tissue using this protocol were of the same quality (and length) as sequences obtained from fresh samples. In contrast, prior to this study a higher success rate in amplifying DNA from formalin-fixed samples was obtained only for short mitochondrial DNA fragments (Wirgin et al. 1997; Chase et al.

1998; Li et al. 2000). When sequences obtained from fresh specimens were compared to those obtained from museum specimens very few base changes or ambiguities were observed (Figure 2.1). To obtain full-length sequences from museum specimens, both forward and reverse sequence reactions were required. One example (Figure 2.1), sample NMAC 24905, was only reverse-sequenced and sample NFAG 68109 was only forward-sequenced, resulting in about 40 bp missing from the opposite end of each sequence. As the yield from PCR products is sometimes small, depending on the age and nature of preservation, this product can be used only for a single sequencing reaction. Therefore, it is recommended that 2-3 replicate PCR reactions be performed, in order to ensure that sufficient product is obtained for both forward and reverse sequencing reactions. We also observed that with different preservation methods (and ages of specimen), each individual specimen had to be re-optimised accordingly (for Mg conc. and annealing temperatures during PCR) as presented in Table 2.1. These specific conditions should be treated as a guide to changes required for amplifying museum specimens successfully.

**Analytical methods**

Analytical methods are presented separately in each data chapter (3, 4 and 6, 7) as they differed.

In conclusion, the method described here differs from previously published studies in that it clearly describes the methodology for successfully obtaining DNA, which can be amplified from formalin-fixed tissues. This is also the first report of successful amplification and sequencing of a nuclear gene using DNA from formalin-fixed tissue. The most important features of this protocol for producing amplifiable DNA are; 1) extending the washing

times to re-hydrate tissue, 2) increasing Mg concentrations in the PCR and 3) decreasing annealing temperatures during PCR. I feel confident that this method will also work for other taxa (vertebrates and some invertebrates).

```
                 *         20          *         40          *         60
           agtttcttctggagctactgactgacaagtcttgccagtCCTTCATCAGCTGGACAGGCG
Nmac1077  : -............................................................ :  59
Nmac1078  : -............................................................ :  59
Nmac1079  : -............................................................ :  59
NMAC24905 : -............................................................ :  59
Nfag1113  : ...........-.................................................. :  59
Nfag1114  : ...........-.................................................. :  59
Nfag1115  : ...........-.................................................. :  59
NFAG68109 : ----------------------------------------..................... :  21

                   *         80          *        100          *        120
           ACGGCTGGGAGTTCAAGCTGTCTGACCCAGATGAGGTGAGAGGTtGTGACCCCACTGGTC
Nmac1077  : ............................................................ : 119
Nmac1078  : ............................................................ : 119
Nmac1079  : ............................................................ : 119
NMAC24905 : ............................................................ : 119
Nfag1113  : .............................................C.............. : 119
Nfag1114  : .............................................C.............. : 119
Nfag1115  : .............................................C.............. : 119
NFAG68109 : .............................................C.............. :  81

                   *        140          *        160          *        180
           AGCGCTACACGTGCTGGGAGGTCCTAAGTAGTTAGCCAGATGAACAAGATGTCCTCTGAC
Nmac1077  : ............................................................ : 179
Nmac1078  : ............................................................ : 179
Nmac1079  : ............................................................ : 179
NMAC24905 : ............................................................ : 179
Nfag1113  : ............................................................ : 179
Nfag1114  : ............................................................ : 179
Nfag1115  : ............................................................ : 179
NFAG68109 : ............................................................ : 141

                  *        200          *        220          *        240
           TTTGTTAGAAAAATACATTTGCATGCCTTGTTCAGCtCCATCAATCTTTTGGGATCTGTT
Nmac1077  : ............................................................ : 239
Nmac1078  : ............................................................ : 239
Nmac1079  : ............................................................ : 239
NMAC24905 : ....................................Y....................... : 239
Nfag1113  : ............................................................ : 239
Nfag1114  : ............................................................ : 239
Nfag1115  : ............................................................ : 239
NFAG68109 : ............................................................ : 201
```

Note: Figure 2.1 continued overleaf

```
                  *       260          *       280          *       300
           AACTTTGTTGGTTTTACaTaAGtAGTTTTTCATTCAAAATATGGATGgCTGCATTCCAAG
Nmac1077  : ..................................................... : 299
Nmac1078  : ..................R.................................. : 299
Nmac1079  : ..................................................... : 299
NMAC24905 : ..................T..C.....................A.......... : 299
Nfag1113  : ..................C.....................A.......... : 299
Nfag1114  : ..................C.....................A.......... : 299
Nfag1115  : ..................C.....................A.......... : 299
NFAG68109 : ..................C....C.....................A.......... : 261

                  *       320          *       340          *       360
           ATGATTGTCAGATTTAAAGAATAaTGTGGTAGTTTGTTGCAAGTCTGTTGTACTATTTTT
Nmac1077  : ..................................................... : 359
Nmac1078  : ..................................................... : 359
Nmac1079  : ..................................................... : 359
NMAC24905 : ......................G.......................................... : 359
Nfag1113  : ..................................................... : 359
Nfag1114  : ..................................................... : 359
Nfag1115  : ..................................................... : 359
NFAG68109 : ..................................................... : 321

                  *       380          *       400          *       420
           AAGTGTTTGCAAGTCATCTTGACAagCTTCTTCTTTGTGTTGTTTCAGGTTGCTAGGAGG
Nmac1077  : ..................................................... : 419
Nmac1078  : ..................................................... : 419
Nmac1079  : ..................................................... : 419
NMAC24905 : ..................................................... : 419
Nfag1113  : ......................GT.............................. : 419
Nfag1114  : ......................GT.............................. : 419
Nfag1115  : ......................GT.............................. : 419
NFAG68109 : ......................G.............................. : 381

                  *       440          *       460          *       480
           tggggcaagaggaaaaacaagcccaagatgaactacca
Nmac1077  : ..................................C.---------------------- : 457
Nmac1078  : ..................................C.---------------------- : 457
Nmac1079  : ..................................----------------------- : 455
NMAC24905 : ------------------------------------------------------------ :   -
Nfag1113  : .................................G.GAACG----------------- : 462
Nfag1114  : .................................G.GAAC------------------ : 461
Nfag1115  : .................................G.GAA------------------- : 460
NFAG68109 : ..------------------------------------------------------- : 384
```

Figure 2.1: Sequence alignment of nuclear marker ETS2 comparing sequences obtained from fresh and museum specimens. Nmac: *Naso maculatus*, Nfag: *N. fageni*. Sample ID's in capitals and bold are formalin-fixed museum specimens. NMAC 24905 Bernice P. Bishop Museum, NFAG 68109: Royal Ontario Museum. Otherwise all samples were sequenced from fresh specimens. **(-)** indicates gaps/missing data, **(.)** indicates same base as shown on top line, **(*)** indicates every $10^{th}$ base. Based on the IUPAC - IUB Ambiguity Set, ambiguities in bases (between two or more possible nucleotides, *ie*. A, C, G, T/U) are represented as separate letters: R (puRine) can be either an A or G; Y (pYrimidine) stands for either C or T/U.