Snake species identification using mitochondrial gene loci
Snake species identification using mitochondrial gene loci – Ine van Velzen

Version: 2

Authors: Ine van Velzen
i.vanvelzen@student.avans.nl

Place: Flinders University, Adelaide, South Australia

Period: 20/02/2012 – 06/07/2012

Supervisors: Dr. Adrian Linacre
adrian.linacre@flinders.edu.au

Kirsten Baken
ka.baken@avans.nl

[2]
FOREWORD

This document is a final paper which is developed after an internship of 4.5 months at the Flinders University in Adelaide, Australia. This paper mainly contains results and its conclusions about the research I have been doing. I would like to thank Adrian Linacre for the opportunity doing a research project at the Flinders University. But also for all the support, ideas and guidance that have been given during the entire project. I would like to extend those thanks to Sherryn Ciavaglia for the help and ideas during the lab work.

Ine
ABSTRACT

Many species have become extinct or threatened with extinction due to a number of reasons. This is predominantly due to human activities or industry. Snake species are also being driven out of their habitat and by the introduction of exotic animal species. For example, in the United States, more than half of all snake species currently listed are endangered [1].

Illegal trade of wildlife is a serious and growing crime worth more than US $20 billion dollars per year. In Australia, reptiles are the biggest target in wildlife trade because of the high financial value. From all trade crimes, less than 1% result in prosecution of the person involved [2]. When is it suspected that a wildlife crime has been committed, it is important to determine what species it is and whether the snake is endangered or not to enforce legislation. A databank with particular sequences of all species could help.

Therefore, investigation and design of DNA markers for forensic application is needed. During this study, genes (ND6 and cytochrome b) within the mitochondrial DNA from the carpet python, *Morelia spilota*, is investigated to see if these could function as a species identification test. Furthermore, phylogenetic relationships based on these genes between and within this snake species is investigated and compared to the taxonomy which is based on morphology.

ND6 and cytochrome b both showed a lot of intraspecies variation within the subspecies of the carpet python. The variation between two samples in ND6 is between 1.07% and 6.09%. In cytochrome b this was between 0.72% and 6.82%. In this last gene there was less variation between the samples from the east coast, but more variation between those samples and samples from the rest of Australia.

Both genes produced similar trees and could be used for species identification of the carpet python. Although the samples from the east coast are very similar to each other, all other species group separately.

These results compared to the taxonomy and unpublished research which assumed there are only two of three subspecies instead of six, it seems like the *Morelia spilota* can be divided into four subspecies; east coast of Australia, north Australia (*Morelia spilota variegata*), central Australia (*Morelia spilota bredli*) and west-south Australia (*Morelia spilota imbricata*).
CONTENTS

FOREWORD

ABSTRACT

1. INTRODUCTION

1.1 Wildlife crime

1.1.1 Forensic science related to wildlife crime

1.1.2 Legislation

1.2 Species

1.2.1 Definition of a species

1.2.2 Speciation

1.2.3 Taxonomy

1.2.4 Habitation of *Morelia spilota* subspecies

1.3 The origin of genetic variation

1.4 Species identification using mitochondrial gene loci

1.5 Mitochondrial DNA

1.5.1 Snake mtDNA

1.6 Phylogenetics

1.6.1 Phylogenetic tree

1.6.2 Software

1.7 Primer information

1.8 Aims of this research

2. METHODS AND MATERIALS

2.1 Extraction

2.2 PCR

2.3 Gel electrophoresis

2.4 Purification

2.5 Sequencing

2.6 Tree construction

3. RESULTS AND DISCUSSION

[5]
3.1 PCR and primer optimizing 22
3.2 Extraction of different snake samples 23
3.3 Sequence quality 26
3.4 Species identification using ND6 28
  3.4.1 Intraspecies variation 28
  3.4.2 Smallest fragment length needed 31
3.5 Species identification using Cytochrome b 32
  3.5.1 Intraspecies variation 32
  3.5.2 Smallest fragment length needed 33
3.6 Taxonomy of the carpet python 36
  3.6.1 Comparing with results of previous research 38
4. CONCLUSION 41
5. REFERENCES 42
1. INTRODUCTION

1.1 Wildlife crime

Wildlife crime exists in many forms, from trafficking in live specimens, hunting out of season, cruelty to animals, habitat destruction, poaching for meat, poaching to use animal parts in medicines, horns and tusks used for jewellery and ornaments, etc. Many animal species have become extinct due to these reasons with the biggest cause of number loss due to human activities [3].

Wildlife crimes happen at international and at national level. At the international level is the trade in endangered species where protected species are traded across international borders. Second level is any crime against wildlife or protected species afforded protection at a national level. This includes: cases of alleged cruelty; illegal poaching, or fishing, or poisoning; disturbing protected habitats; or the collecting or selling of protected whole animals and plants or products derived from them. The increase in the decline of the numbers of animal species and extinctions emphasize the need for legal protection and enforcement of relevant legislation [4, 5].

The biological material that is traded most of the time is not the whole animal but body parts such as skin, bone or powdered horn. For identification of a species, morphology and microscopy are the starting points [8]. Morphological analysis of parts of an animal, or even a live specimen, will often have to be undertaken by a specialist, often from a zoo or a museum. Microscopy of hairs is another skill what requires much experience in order to be able to identify with confidence that the material is, for instance, that of a protected species which is protected and CITES listed, compared to those species that are not protected. However, even with experience, a microscopic comparison of hairs may not yield a definitive identification [6]. Much material that is traded is not in a condition where species identification can be made by microscopy or morphology. For instance, the material may present as powders, potions and oils. Therefore, using DNA markers for species identification has become a very important and helpful tool with such investigations [6]. Genetic species identification relies on the isolation and analysis of DNA markers that show variation between species, but are conserved within species, see paragraph 1.4.

1.1.1. Legislation

There are two prime issues in wildlife crime which relate to the different types of legislation. The first being the ability to identify a particular species and the second is the ability to determine if the biological material can be assigned with confidence to a particular individual member of that species. The reason that a particular species has to be determined is that many species are listed as being
protected both at a national and international level. There are currently 175 countries that are signatories to the Convention for the International Trade in Endangered Species of flora and fauna (CITES). CITES is an organization that monitors the trade in protected and endangered species across international borders [7].

In Australia, CITES is regulated under Part 13A of the Commonwealth Environmental Protection and Biodiversity Conservation Act 1999 (EPBC Act). EPBC regulates the export of Australian native species, exports and imports of all CITES listed species, and the import of plants and animals that could affect native species or their habitats. Penalties for infringing this Act can go up to $110,000 for an individual or $550,000 for a corporation, and up to 10 years of imprisonment.

It is difficult to estimate the exact amount of illegal trade as there are not the same international surveillance teams that are used for drug enforcement for the prosecution of offences involving wildlife [8].

1.2 Species

1.2.1 Definition of a species

There are a number of definitions of a species. The two main definitions are biological species and phylogenetic species. The biological species definition is based on gross morphological features. A biological species is a group of living organisms consisting of similar individuals capable of exchanging genes or interbreeding. The phylogenetic definition of species relies on genetic markers (RNA, DNA or proteins) thought to be characteristic of a species. This is being used in taxonomy, biodiversity, and evolutionary studies. Even with the advent of DNA and sequencing technologies, there is still no clear guideline to how many genetic variations forms a separate species [9, 10].

1.2.2 Speciation

Speciation can occur by allopatric, sympatric or parapatric methods. Allopatric speciation, or geographic speciation, occurs when a population becomes isolated due to geographical changes, for example mountain range. The isolated population then undergoes divergence as they will undergo different selected pressures, they independently undergo genetic drift and different mutations arise in the populations’ gene pool. Sympatric speciation is the process when new species evolve from a single ancestral species while living the same geographic region. The organisms are similar or even identical, and they occur together in at least some places. Selection for specialization has to be extremely strong to cause the population to diverge. This is because the gene flow in a random population would tend to break down differences between the two species. Parapatric speciation is
the evolution of geographically adjacent populations into distinct species. Divergence occurs by limited interbreeding where the two groups come into contact. The environment is so large that gene flow among the members is limited, resulting in genetic drift among different sub-populations [11]. These last two methods are both very rare and therefore not very common.

1.2.3 Taxonomy

Taxonomy is the science of identifying and naming organisms, and arranging them into a classification system. To distinguish different levels of similarity, each classifying group, called taxon, is subdivided into other groups. The taxa in hierarchical order: Domain, Kingdom, Phylum, Class, Order, Family, Genus and finally Species [12]. Figure 1 shows an overview of the taxonomy of the *Morelia spilota* and its subspecies.

![Taxonomy Diagram](image)

*Figure 1: The taxonomy of the Morelia spilota and its subspecies*
1.2.4 Habitation of *Morelia spilota* subspecies

It is assumed that the different subspecies of the *Morelia spilota* are arised by the different areas where they live because of, for example, mountain chains of rivers. Figure 2 shows a map of Australia with the locations of the seven subspecies of *Morelia spilota*.

![Figure 2: Map of Australia with the location of the different subspecies. *Morelia spilota cheynei* is coloured red, *Morelia spilota imbricata* is coloured orange, *Morelia spilota mcdowell* is colored yellow, *Morelia spilota metcalfei* is coloured green, *Morelia spilota spilota* is coloured blue and *Morelia spilota variegata* is coloured purple.](image)

1.3 The origin of genetic variation

The process of evolution is a change in the inherited characteristics of a population or species. It is an alteration of the genetic composition of a population. Genetic changes in populations and species begin with changes in the genetic material carried by individual organisms: mutations [13].

Mutations in a coding region have different effects depending on which base is altered. Mutations in coding regions are synonymous mutations, when they have no effect on the amino acid sequence of the polypeptide of protein. Non-synonymous mutations, in contrast, result in amino acid substitutions. They may have little or no effect on the functional properties of the polypeptide or protein, and may have substantial consequences. A mutation at the third base of the codon will not affect the resulting amino acid. The second base of a codon cannot mutate without alteration to the encoded amino acid [14].

Within gene sequences an insertion or a deletion of a base(s) (together called an indel) are another common kind of mutation. Indels may be a single base pair or many. If a single base pair becomes
inserted into or deleted from a coding sequence, the codon is shifted by one nucleotide, so that downstream triplets are read as different codons and translated into different amino acids. Thus, insertions of deletions often result in frameshift mutations [13].

The twenty amino acids fall into four categories; acid, basic, polar a non-polar. Depending on the function of the protein it may be that an alteration from one amino acid to another in the same group has little effect on the protein’s activity. Alteration of the amino acid between the different groups of proteins has much more effect, unless the amino acid plays little part in the activity of the protein [14].

Mutations can be a result of one or more incorrect nucleotide which is added randomly during DNA replication or induced by external factors. On a larger scale, the gene rearrangements lead to genetic variation. Variation leads to polymorphisms, literally meaning many forms. Single nucleotide polymorphisms (SNPs) are the most common type of polymorphism. Single variations at the DNA level are common and lead to the assumption that every individual’s DNA is unique [14].

1.4 Species identification using mitochondrial gene loci

There is a small amount of DNA variation within a species, known as intraspecies variation. Interspecies variation, on the other hand, is the variation between species [15]. The locus of choice must have very little intraspecies variation, such that all members of the same species have the same DNA type. The locus also needs to have enough interspecies variation so that members of one species can be separated from members of closely related species.

In animals, gene loci on the mitochondrial genome fit these criteria best. A major reason for using mitochondrial DNA (mtDNA) loci is that there is no recombination of mtDNA. All maternal descendents will have the same mitochondrial DNA sequence, with the exception of mutations, and all loci will be linked. Genetic variation within coding regions is likely to be much less than non-coding, especially if the encoded protein serves a function. With all the coding sections of the mitochondrial genome coding for proteins or RNA molecules, it would be expected that there would be conservation of sequence as any change in the proteins or RNA molecules could affect the organism. Unlike the nucleus, mitochondria do not contain an error reading enzyme to repair DNA bases added incorrectly during DNA replication. Therefore, the amount of single base mutations is up to five times higher in mtDNA compared to nuclear DNA. The analysis of genes on the mitochondria also has the advantage of being present at very high copy number [14, 16, 17].
Species identification by using mitochondrial gene loci is becoming routine but has not been standardized to one single locus [17].

1.5 Mitochondrial DNA

Mitochondria are structures within cells that generate chemical energy. Although most DNA is packaged in chromosomes within the nucleus, mitochondria also have a small amount of their own DNA. This genetic material is known as mitochondrial DNA or mtDNA [18].

MtDNA is a circular structure and human mtDNA contains 16,569 base pairs [19]. Due to the presence of multiple mitochondrion and chloroplast organelles within a single cell, these markers are present in multiple copies. In comparison to nuclear DNA, where in diploid species a single nucleus carries just 2 copies of each marker, this increases the chance of success when analyzing trace evidence samples that typically contain relatively little cellular material. Another advantage of using mitochondrial DNA rather than nucleus DNA is that the mitochondrion also has a strong protein coat that protects the mtDNA from degradation by bacterial enzymes [20, 21].

Differences between mtDNA and nuclear DNA are shown in Table 1:

Table 1: Differences between nuclear DNA and mitochondrial DNA [19]

<table>
<thead>
<tr>
<th></th>
<th>Nuclear DNA</th>
<th>Mitochondrial DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Size of genome</strong></td>
<td>About 3,2 billion bp</td>
<td>16,569 bp</td>
</tr>
<tr>
<td><strong>Copies per cell</strong></td>
<td>2 (1 allele from each parent)</td>
<td>Can be &gt;1000</td>
</tr>
<tr>
<td><strong>Percent of total DNA content per cell</strong></td>
<td>99.75%</td>
<td>0.25%</td>
</tr>
<tr>
<td><strong>Structure</strong></td>
<td>Linear; packaged in chromosomes</td>
<td>Circular</td>
</tr>
<tr>
<td><strong>Inherited from</strong></td>
<td>Father and mother</td>
<td>Mother</td>
</tr>
<tr>
<td><strong>Chromosomal pairing</strong></td>
<td>Diploid</td>
<td>Haploid</td>
</tr>
<tr>
<td><strong>Unique</strong></td>
<td>Unique to individual (except of identical twins)</td>
<td>Not unique to individual (same as maternal relatives)</td>
</tr>
</tbody>
</table>
Human mitochondrial DNA contains 37 genes, which consists of 13 protein coding genes, 22 transfer RNAs (tRNA), 2 ribosomal RNAs (rRNA) and a control region [19].

The 22 tRNA genes code for an amino acid, see Table 2. The 22 tRNA genes on the mtDNA are relatively small (less than 100bp) and exhibit little variation. Greater variation occurs in the 13 genes involved in the oxidative phosphorylation process and the two rRNA genes. This greater variation is due to the function of the transcribed molecule and the length of the molecule.

The 13 protein coding genes consist of ND1, ND2, COI, COII, AT8, AT6, COIII, ND3, ND4L, ND4, ND5, ND6 and finally Cytochrome b. These genes provide instructions for making enzymes involved in oxidative phosphorylation. Oxidative phosphorylation is a process that uses oxygen and simple sugars to create adenosine triphosphate (ATP), the cell's main energy source [22].

In mtDNA is also an 1122 base pairs control region (also called D-loop) that contains the origin of replication for one of the mtDNA strands but does not code for any gene products and is therefore referred sometimes as the ‘non-coding’ region. It contains controlling signals for replication and transcription of mtDNAs [23]. It is the only area in the mitochondrial DNA that can develop DNA mutations without affecting the role of the DNA. At this area occurs greater intraspecies variation compared to the coding regions of the mitochondrial DNA. In this control region are two hyper variable regions. These hyper variable regions are used in human identification due to the high levels of intra-species variation. The amount of intraspecies variation precludes their use in species identification [19, 24].

1.5.1 Snake mtDNA

Snake mtDNA is very similar to human mtDNA, but there are sometimes gene duplications and translocations. Most remarkable is the duplicate of the control region, which is present in all Alephinophidia snakes and not in the Scolecophidia snakes (blind snakes and thread snakes). Many species that have the duplicate region also show gene translocations around the second control region. Because of the variation in length and/or numbers of the control regions, the complete mitochondrial genomes from different species range in size. There are also a lot of gene translocations which differs per family, species or genus etc [25].

An example is the translocation from the tRNA$_{Leu}$ gene. In typical vertebrate mtDNA the tRNA$_{Leu}$ is located between the 16S rRNA and ND1, in the snake organization from the little file snake (Acrochordus granulates), the boa constrictor (Boa constrictor), the red pipe snake (Cylindrophis ruffus), the akamata (Dinodon semicarinatus), the ball python (Python regius), and the sunbeam snake (Xenopeltis unicolor) (which are all derived from different families) this gene is located
between the tRNA$^{\text{Ile}}$ and tRNA$^{\text{Gln}}$ genes. This gene location is shown in Figure 3, which is an overview of mtDNA from vertebrates (A) and snakes (B) [23].

Figure 3: Organization of the vertebrate mtDNA (A) and the organization of mtDNA from the little file snake, the boa constrictor, the red pipe snake, the akamata, the ball python, and the sunbeam snake (B) [23].

1.6 Phylogenetics

Phylogenetics is the study of relatedness among groups of organisms, using information about the differences between species and the rate of genetic changes. Trees are used as visual displays that show evolutionary events [26].

1.6.1 Phylogenetic tree

A phylogeny tree is a graphical representation of the relationship between different organisms that share a common ancestor. The phylogeny tree represents in what order species, or individuals, diverged from a common ancestor. There are many types of data that can be used to construct a phylogenetic tree. A phylogeny tree consists of a collection of nodes (internal and external) and branches to connect them. The internal nodes represent common ancestors; the external nodes, which are at the ends of the branches, represent living organisms (Operational Taxonomic Units (OTUs)). The lengths of the branches usually represent an elapsed time; sometimes it represents the number of molecular changes (for example mutations) that have taken place between the two nodes. This is calculated from the degree of differences when sequences are compared. Sometimes, the lengths are irrelevant and the tree represents only the order of evolution. Finally, the trees can be classified as rooted trees and unrooted trees as shown in Figure 4 [27].
An unrooted tree simply represents phylogenetic but does not provide an evolutionary path. In an unrooted tree, an external node represents a contemporary organism. Internal nodes represent common ancestors of some of the external nodes. In Figure 4, the unrooted tree shows the relationship between organisms A, B, C & D and does not tell us anything about the series of evolutionary events that led to these genes. There is also no way to tell whether or not a given internal node is a common ancestor of any 2 external nodes [28].

In a rooted tree, one of the internal nodes is used as an out-group, and becomes the common ancestor of all the other external nodes. The out-group therefore enables the root of a tree to be located and the correct evolutionary pathway to be identified [28].

1.6.2 Software

Two different softwares will be used during this research for align sequences and making up different trees. The first one, Geneious, is bioinformatics software which makes it possible to search, organize and analyze genomic and protein information. MEGA (in full and give the web address) is used for conducting sequence alignments and inferring phylogenetic trees.

The trees used in this study are UPGMA based. UPGMA stands for Unweighted Pair Group Method with Arithmetic Mean, and is the simplest method of tree construction. It is a distance method and therefore needs a distance matrix. It relies on the rates of evolution (the number of observed nucleotide or amino acid substitutions) among different lineages to be approximately equal [28].
1.7 Primer information

Since there is likelihood for SNPs in the nucleotide sequences by the use in species identification, the sequence can contain differences at specific locations. In a degenerate primer, a single symbol will be used to designate a variety of possible nucleotides at a single position. Table 2 shows all the possible nucleotide codes [29].

*Table 2: IUPAC (International Union of Pure and Applied Chemistry) codes [29].*

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Adenine</td>
</tr>
<tr>
<td>C</td>
<td>Cytosine</td>
</tr>
<tr>
<td>G</td>
<td>Guanine</td>
</tr>
<tr>
<td>T</td>
<td>Thymine</td>
</tr>
<tr>
<td>U</td>
<td>Uracil</td>
</tr>
<tr>
<td>R</td>
<td>Purine (A or G)</td>
</tr>
<tr>
<td>Y</td>
<td>Pyrimidine (C, T or U)</td>
</tr>
<tr>
<td>M</td>
<td>C or A</td>
</tr>
<tr>
<td>K</td>
<td>T, U or G</td>
</tr>
<tr>
<td>W</td>
<td>T, U or A</td>
</tr>
<tr>
<td>S</td>
<td>C or G</td>
</tr>
<tr>
<td>B</td>
<td>C, T, U or G (not A)</td>
</tr>
<tr>
<td>D</td>
<td>A, T, U or G (not C)</td>
</tr>
<tr>
<td>H</td>
<td>A, T, U or C (not G)</td>
</tr>
<tr>
<td>V</td>
<td>A, C or G (not T or U)</td>
</tr>
<tr>
<td>N</td>
<td>Any base (A, C, G, T or U)</td>
</tr>
</tbody>
</table>
During this research primer sets for ND6 and Cytochrome b were designed. The sequence, length and annealing temperature of the primers are found in Table 3. For the tree design, focused was on ND6, an overview of this fragment and the location of the primers is shown in Figure 5. Figure 6 and 7 shows the locations of the all primers in cytochrome b.

ND6 is a gene which provides information for making a protein called NADH dehydrogenase 6. This protein is part of a large enzyme complex known as complex I, which is one of several enzyme complexes necessary for oxidative phosphorylation [30].

Cytochrome b is a component of respiratory chain complex III, also known as the bc1 complex. This complex is involved in electron transport and the generation of ATP [22].

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Ta</th>
<th>Location</th>
<th>Length fragment</th>
</tr>
</thead>
<tbody>
<tr>
<td>ND6 forward</td>
<td>TCA GCC AAC AAA CCT CAA CAG AAC T</td>
<td>59°C</td>
<td>ND5</td>
<td>ND5 tRNA^Glu</td>
</tr>
<tr>
<td>ND6 reverse</td>
<td>TTT CAG GCC GCA GGA CTC GGT A</td>
<td>63°C</td>
<td>tRNA^Glu</td>
<td>740 bp</td>
</tr>
<tr>
<td>Cyt b 1 forward</td>
<td>AAC ACW ACC GCC CCA ACM ACC A</td>
<td>58°C</td>
<td>ND6</td>
<td>ND6 mid cyt b</td>
</tr>
<tr>
<td>Cyt b 1 reverse</td>
<td>GGT CRT TGA TTG CGA ACC CKC C</td>
<td>58°C</td>
<td>mid cyt b</td>
<td>618 bp</td>
</tr>
<tr>
<td>Cyt b 2 forward</td>
<td>CGC YTT CTT CGG WTA YGT CCT CCC</td>
<td>61°C</td>
<td>mid cyt b</td>
<td></td>
</tr>
<tr>
<td>Cyt b 2 reverse</td>
<td>GGG KTT RRC CCA GCT TTG GTT T</td>
<td>56°C</td>
<td>tRNA^Thr</td>
<td>814 bp</td>
</tr>
</tbody>
</table>

![Figure 5: Primer location and the length of the ND6 gene.](image)
1.8 Aims of this research

During this research the carpet python, *Morelia spilota*, will be investigated. This snake species is CITES listed and therefore identification of snake species is an important task. As there is no specific test yet, Sherryn Ciavaglia, a PhD student of Flinders University, is conducting research into the investigation and design of python DNA markers. A part of her research is looking at the interspecies and intraspecies variation in mitochondrial DNA loci to see if a particular (part of a) gene can be used for identification. I will look at the Cytochrome b and the ND6 gene during my internship and compare these genes.

Another thing I will look at is the phylogenetic relationship between the subspecies of the *Morelia spilota*. Unpublished research showed that there are probably no seven different subspecies in Australia, but less. I will extract samples from different areas in Australia and compare these to the taxonomy which is based in morphology.

Several techniques will be used. The method starts with different DNA extractions from the samples. Mitochondrial DNA loci will be amplified, purified and sequenced, followed by comparison of the resulting sequence with reference sequence data from different species.
2. METHODS AND MATERIALS

2.1 Extraction

For extractions from scale clippings and tissue samples with a low DNA yield, the Isolation of Total DNA from Nail Clippings and Hair protocol is used from the QIAamp® DNA Investigator Kit (QIAGEN).

All samples which were preserved in ethanol were first sub-headed in a heating block for approximately 10 minutes, and then measured and weighted. The samples were lysed by adding 300 µL of buffer ATL, 20 µL of proteinase K, and 20 µL of 1 M DTT. The tissue samples were lysed by adding 180 µL of buffer ATL and 20 µL proteinase K. After vortexing the samples were incubated at 56 °C with shaking at 900 rpm overnight. After the samples were spun down, 200 µL of buffer AL was added and pulse vortexed for 15 seconds. The samples were incubated at 70°C at 900 rpm for 10 minutes, spun down and 200 µL of ethanol (96-100%) was added to the samples. For the tissue sample, 200 µL was added. After vortexing and spinning down, the supernatant were transferred to the QIAamp MinElute column. The columns were spun down at 6000 x g for 1 minute and the flow-through was discarded. 500 µL of buffer AW1 was added, centrifuged at 6000 x g for 1 minute and the flow-through was discarded. Here after 700 µL of buffer AW2 was added to the columns, spun down and the flow-through was discarded again. The last clean up step was adding 700 µL of ethanol (96-100 %), the samples were spun down and the flow-through was discarded again. The columns were spun down were another 3 minutes and put on a heating block at 70°C for 3 minutes to make sure the membrane was dry. Finally, the DNA was eluted by adding 50-200 µL of buffer ATE to the columns and they were spun down at 6000 x g for 1 minute.

For extractions from blood, skin and tissue samples with a high DNA yield, the Isolation of Total DNA from Nail Clippings and Hair protocol is used from the QIAamp® DNA Mini Kit (QIAGEN).

All samples which were preserved in ethanol were first dried in a heating block for approximately 10 minutes, measured and weighted. The tissue samples were lysed by adding 180 µL of buffer ATL and 20 µL of proteinase K. The skin samples were lysed by adding 600 µL of buffer ATL, 40 µL of 1M DTT and 40 µL proteinase K and for the blood sample 180 µL of PBS, 20 µL proteinase K and 200 µL AL. After vortexing the samples were incubated at 56°C with shaking at 900 rpm overnight. After the samples were spun down, 200 µL of buffer AL was added to the tissue samples and pulse vortexed for 15 seconds. For the skin sample, 600 µL of buffer AL was added. The samples were incubated at 70°C at 900rpm for 10 minutes, spun down and 200 µL of ethanol (96-100%) was added to the tissue and blood samples. For the skin samples, 600 µL was added. After vortexing and spinning down, the
supernatant was transferred to the QIAamp DNA Mini Column. The columns were spun down at 6000 x g for 1 minute and the flow-through was discarded. 500 μL of buffer AW1 was added, centrifuged at 6000 x g for 1 minute and the flow-through was discarded. Hereafter 500 μL of buffer AW2 was added to the columns, spun down and the flow-through was discarded again. Finally, the DNA was eluted by adding 50-200 μL of buffer AE to the columns and they were spun down at 6000 x g for 1 minute.

2.2 PCR

PCR samples were prepared by adding 6 μL dH₂O, 10 μL of Mango Mix (Bioline), 1μL forward primer and 1μL reverse primer to 2μL of DNA. Finally the PCR was set up with the right PCR program and started.

In this research five different primer sets will be used: A primer set to amplify the ND2 region, another set for the ND4 region, ND6 region, a primer set that amplify the first half of Cytochrome b and another primer set that amplifies the second half of Cytochrome b. These primer sets will be called in the rest of this report as followed: Primer set ND2, primer set ND4, primer set ND6, primer set cytb1 and primer set cytb2.

2.3 Gel electrophoresis

A 2 % agarose (Applichem) was made by adding 100 mL of TE buffer (Tris and EDTA) (Biorad) to 2 grams of agarose. (A 1 % gel was made by adding 100 mL of TE buffer to 1 gram of agarose). This solution was heated until the agarose was dissolved. 5μL of ethidiumbromide (Biorad) was added to the gel when it was cooled down to approximately 60 °C. As soon as the gel was cooled down and hardened, it was transferred to an electrophoresis tank and covered with TE buffer. The samples were prepared by adding 2μL of loading buffer (Biorad) to 5μL of DNA sample. The samples were added in the wells of the gel and ran at 100 V for about 30 minutes. An UV transiluminator (Biorad, Gel Doc™ EZ Imager) was used to visualize the bands.

2.4 Purification

The DNA purification was done by using the QIAquick Gel Extraction Kit (QIAGEN).

First of all, the DNA fragments from the agarose gel were excised, weighted and put in a microfuge tube. Three volumes of buffer QG was added to 1 volume of gel (100 mg = 100 μL) and the samples were incubated at 50°C for 10 minutes. 1 volume of isopropanol was added to sample, mixed and the samples were applied to the QIAquick column and centrifuged for 1 minute. Another 500 μL of buffer
QG was added to the column and centrifuged at 6000 x g for 1 minute. 750 µL of buffer PE was added to the column to wash the DNA and the columns were centrifuged at 6000 x g for 1 minute again. The flow-through was discarded and the columns were spun down for another minute. Finally, the DNA was eluted by adding 30-50 µL of buffer EB (10mM Tris-Cl, pH8.5) or water to the column and centrifuged at 6000 x g for 1 minute.

2.5 Sequencing
After the samples were purified, they were put on a 2% gel to quantify them. For sequencing, which was done by AGRF, Australian Genome Research Facility, the samples needed to be prepared in a total of 12 µL, of which 30 to 80 ng consisted of DNA, 10 pmol of primer and supplemented with H₂O.

2.6 Tree construction
The software Geneious was used to look at the quality of the sequences. If necessary, bases in the beginning and at the end of each sequence were removed. In MEGA, the sequences were aligned by Clustel W and different trees were designed; Maximum Likelihood Tree, Neighbor-Joining Tree, Minimum-Evolution Tree, UPGMA Tree and the Maximum Parsimony Tree.
3. RESULTS AND DISCUSSION

3.1 PCR and primer optimizing

The PCR settings for the ND6 and both Cytochrome b (cyt b) primer sets needed to be optimized before starting amplifying these regions. The annealing temperatures from these six primers were between 56°C and 63°C so a temperature range from 52°C to 65°C was investigated. During this reaction 30 cycles were applied. Figure 8 and 9 shows the results of this primer optimizing.

Figure 8: Primer set cyt b1 with six different temperatures and primer set ND6 with three different temperatures.
1: 1kb ladder – 2: 65°C cyt b1 – 3: 62.5°C cyt b1 – 4: 60.1°C cyt b1 – 5: 57.1°C cyt b1 – 6: 55°C cyt b1 – 7: 52°C cyt b1 – 8: 65°C ND6 – 9: 62.5°C ND6 – 10: 60.1°C ND6

Figure 9: Primer set cyt b2 with six different temperatures and primer set ND6 with three different temperatures.
1: 1kb ladder – 2: 65°C cyt b2 – 3: 62.5°C cyt b2 – 4: 60.1°C cyt b2 – 5: 57.1°C cyt b2 – 6: 55°C cyt b2 – 7: 52°C cyt b2 – 8: 57.1°C ND6 – 9: 55°C ND6 – 10: 52°C ND6

According to these gels, an annealing temperature of 56°C was used for primer set cyt b1 and primer set cyt b2. For primer set ND6 an annealing temperature of 55°C was used and 35 cycles instead of 30.

To establish the minimum amount DNA template in a PCR the following masses of DNA were used in an amplification and then ran on a gel: 60 ng, 6 ng, 0.6 ng and 0.06 ng (60 pg). It is visible in Figure 10 that 60 ng provides a band, but there is too much DNA so the reagents cannot work optimally. The
samples containing 6 ng and 0.6 ng resulted in a very bright band, which means too much DNA was present. After adding 0.06 ng of DNA, amplification has occurred but there is not much DNA present.

Figure 10: Different amounts of DNA were added to a PCR to determine the best amount of DNA for amplification. Lane 1: 1kb ladder – 2: 60 ng – 3: 6 ng – 4: 0.6 ng – 5: 0.06 ng

3.2 Extraction of different snake samples

Before any research could started, samples were picked up from the South Australia museum in Adelaide as this is a repository for many samples from all over Australia. Figure 11 and Table 4 shows an overview of all samples extracted.

Figure 11: An overview of all extracted samples
Table 4: An overview of all different samples which were extracted. The samples were diluted in different amounts of buffer; also the concentration and purity are shown in this Table.

<table>
<thead>
<tr>
<th>Sample number</th>
<th>DNA diluted in µL buffer</th>
<th>Concentration in ng/µL</th>
<th>Purity, 260/280 ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Liver samples</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65857</td>
<td>40</td>
<td>19.9</td>
<td>1.27</td>
</tr>
<tr>
<td>65858</td>
<td>40</td>
<td>39.3</td>
<td>1.55</td>
</tr>
<tr>
<td>65859</td>
<td>40</td>
<td>33.8</td>
<td>1.34</td>
</tr>
<tr>
<td>65850</td>
<td>40</td>
<td>10.5</td>
<td>1.50</td>
</tr>
<tr>
<td>67641</td>
<td>600</td>
<td>15.0</td>
<td>1.91</td>
</tr>
<tr>
<td>65866</td>
<td>600</td>
<td>0.9</td>
<td>2.80</td>
</tr>
<tr>
<td>65867</td>
<td>40</td>
<td>35.7</td>
<td>1.37</td>
</tr>
<tr>
<td>65870</td>
<td>40</td>
<td>17.9</td>
<td>1.43</td>
</tr>
<tr>
<td>65869</td>
<td>40</td>
<td>72.2</td>
<td>1.52</td>
</tr>
<tr>
<td>70157</td>
<td>50</td>
<td>216.9</td>
<td>1.88</td>
</tr>
<tr>
<td>93375</td>
<td>50</td>
<td>1384.9</td>
<td>2.06</td>
</tr>
<tr>
<td><strong>Other tissue samples</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>69347</td>
<td>600</td>
<td>3.3</td>
<td>1.62</td>
</tr>
<tr>
<td>81663</td>
<td>600</td>
<td>0.8</td>
<td>-23.83</td>
</tr>
<tr>
<td>81307</td>
<td>40</td>
<td>13.1</td>
<td>1.83</td>
</tr>
<tr>
<td>81320</td>
<td>40</td>
<td>52.5</td>
<td>1.87</td>
</tr>
<tr>
<td>81662</td>
<td>200</td>
<td>1.2</td>
<td>5.68</td>
</tr>
<tr>
<td>81029</td>
<td>200</td>
<td>63.5</td>
<td>1.94</td>
</tr>
<tr>
<td>121831</td>
<td>200</td>
<td>31.0</td>
<td>1.93</td>
</tr>
<tr>
<td>81030</td>
<td>50</td>
<td>253.7</td>
<td>1.88</td>
</tr>
<tr>
<td>81028</td>
<td>50</td>
<td>480.3</td>
<td>1.92</td>
</tr>
<tr>
<td>32052</td>
<td>50</td>
<td>312.4</td>
<td>1.95</td>
</tr>
<tr>
<td>72828</td>
<td>50</td>
<td>258.0</td>
<td>2.10</td>
</tr>
<tr>
<td>81388</td>
<td>50</td>
<td>34.4</td>
<td>1.96</td>
</tr>
<tr>
<td>31935</td>
<td>50</td>
<td>288.0</td>
<td>2.08</td>
</tr>
<tr>
<td>86801</td>
<td>50</td>
<td>144.0</td>
<td>2.08</td>
</tr>
<tr>
<td>79106</td>
<td>50</td>
<td>59.4</td>
<td>1.65</td>
</tr>
<tr>
<td><strong>Skin samples</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>76225</td>
<td>200</td>
<td>9.4</td>
<td>1.79</td>
</tr>
<tr>
<td>83875</td>
<td>200</td>
<td>1.9</td>
<td>1.75</td>
</tr>
<tr>
<td>102726</td>
<td>100</td>
<td>17.8</td>
<td>1.51</td>
</tr>
</tbody>
</table>
### Scale clippings

<table>
<thead>
<tr>
<th>ID</th>
<th>Count</th>
<th>Length</th>
<th>Width</th>
<th>Peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>76530</td>
<td>200</td>
<td>5.5</td>
<td></td>
<td>1.72</td>
</tr>
<tr>
<td>118880</td>
<td>200</td>
<td>10.4</td>
<td></td>
<td>1.90</td>
</tr>
<tr>
<td>118877</td>
<td>100</td>
<td>6.8</td>
<td></td>
<td>1.86</td>
</tr>
<tr>
<td>118879</td>
<td>100</td>
<td>7.0</td>
<td></td>
<td>2.32</td>
</tr>
<tr>
<td>76529</td>
<td>100</td>
<td>3.1</td>
<td></td>
<td>1.23</td>
</tr>
<tr>
<td>12180</td>
<td>100</td>
<td>5.1</td>
<td></td>
<td>1.63</td>
</tr>
<tr>
<td>76268</td>
<td>100</td>
<td>17.1</td>
<td></td>
<td>1.80</td>
</tr>
<tr>
<td>81659</td>
<td>100</td>
<td>6.6</td>
<td></td>
<td>1.55</td>
</tr>
<tr>
<td>102726</td>
<td>100</td>
<td>17.8</td>
<td></td>
<td>1.51</td>
</tr>
<tr>
<td>76771</td>
<td>50</td>
<td>22.1</td>
<td></td>
<td>1.93</td>
</tr>
<tr>
<td>81240</td>
<td>50</td>
<td>20.6</td>
<td></td>
<td>1.88</td>
</tr>
<tr>
<td>81660</td>
<td>50</td>
<td>3.0</td>
<td></td>
<td>1.41</td>
</tr>
<tr>
<td>82987</td>
<td>50</td>
<td>39.5</td>
<td></td>
<td>1.92</td>
</tr>
<tr>
<td>106074</td>
<td>50</td>
<td>33.8</td>
<td></td>
<td>1.89</td>
</tr>
<tr>
<td>80324</td>
<td>50</td>
<td>10.0</td>
<td></td>
<td>1.76</td>
</tr>
<tr>
<td>81661</td>
<td>50</td>
<td>1.0</td>
<td></td>
<td>0.70</td>
</tr>
<tr>
<td>106072</td>
<td>50</td>
<td>2.6</td>
<td></td>
<td>4.71</td>
</tr>
<tr>
<td>66317</td>
<td>50</td>
<td>0.4</td>
<td></td>
<td>-0.87</td>
</tr>
<tr>
<td>76772</td>
<td>50</td>
<td>7.6</td>
<td></td>
<td>2.09</td>
</tr>
</tbody>
</table>
After the first set of samples were processed by DNA extraction, the concentration of most samples was very low. Therefore, the next extractions were eluted in half of the aforementioned volumes. As visible in this Table, the concentrations were obviously higher.

The function of mitochondrial DNA is generating energy in the cell. In liver samples and muscle there will be therefore much more mtDNA present than in for example skin- and scale clipping samples. For every 6 ng, there will be 2 copies of nuclear DNA. This does not count for mitochondrial DNA as the amount of mtDNA differs per cell. So a concentration of 20 ng/µL for a scale clipping can contain less mtDNA than a muscle sample with a concentration of 10 ng/µL.

3.3 Sequence quality

Different qualities of sequence results were obtained; this was dependent on the quality and quantity of the DNA used. The first couple sequences results obtained were not from good quality with different reasons. Figure 12 is an example of a successful sequencing reaction. A successful sequence reaction shows sharp and regular single colour peaks and no background and a smooth baseline. Figure 13 shows a high background / noisy sequencing signal. When there is an observation of small peaks underneath all peaks in a sequence, a low signal can be confirmed. Different causes
are possible; too little DNA template was added to the reaction, the primer binding to the template is not very efficient; the template was not purified well enough; there is more than one sequence present. Figure 14 shows a failed reaction, there is poor quality template or primer or the primer binding site is absent, deleted or mutated. Figure 15 shows a reaction where sequences were mixed from the beginning. The presence of more than one template in a sequencing reaction will lead to overlying sequences. Figure 16 shows an abrupt loss of signal. If high peaks stops abruptly at a certain position, there can we two causes; the template has a secondary structure or there is a high G/C content in the DNA template. Rich G/C regions can form hairpin structures that the sequencing polymerase cannot pass through. Finally, Figure 17 shows a top-heavy sequence. If there are very high peaks in the beginning followed by a sharp decrease in signal intensity, there is too much template or too much primer used in the sequencing reaction. The template or primers are depleted in the early rounds of cycle sequencing with creates a short fragment.

Figure 12: Successful reaction

Figure 13: High background/Noisy sequencing signal

Figure 14: Failed reaction
3.4 Species identification using ND6

To look at the intraspecies variation within *Morelia spilota* subspecies, twelve samples from different areas in Australia were extracted and sequenced. As most different subspecies are found on the east coast of Australia, most analyzed samples are from that area. Figure 18 shows the location of the samples used in this experiment.
3.4.1 Intraspecies variation

After sequencing the samples were aligned by the software Geneious and MEGA. An example of an alignment is shown in Figure 19. In this Figure the bases that vary are encircled. In the whole fragment that was sequenced, which was 560 bp long, there were 63 sites with variable sequence. So 11.25% of all bases vary in ND6, which means there is a lot of intraspecies variation within these samples.

By identification of a species, the sample will be compared to single samples existing in the databank so it is more important to know how much variety exists between two single samples. Table 5 shows the variety between a couple of samples for ND6. The average deviation between two samples is 4.0%.
Table 5: Amount and percentage of intraspecies variation in the ND6 gene of the carpet python.

<table>
<thead>
<tr>
<th></th>
<th>Brisbane</th>
<th>New South Wales</th>
<th>Darwin</th>
<th>North McKay</th>
<th>Western Australia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Brisbane</strong></td>
<td>X</td>
<td>10/657 1.52%</td>
<td>31/657 4.72%</td>
<td>7/657 1.07%</td>
<td>36/657 5.48%</td>
</tr>
<tr>
<td><strong>New South Wales</strong></td>
<td>X</td>
<td>X 29/657 4.41%</td>
<td>10/657 1.52%</td>
<td>40/657 6.09%</td>
<td></td>
</tr>
<tr>
<td><strong>Darwin</strong></td>
<td>X</td>
<td>X 30/678 4.42%</td>
<td>37/678 5.46%</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>North McKay</strong></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>35/657 5.33%</td>
<td></td>
</tr>
<tr>
<td><strong>Western Australia</strong></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X 35/657 5.33%</td>
<td></td>
</tr>
</tbody>
</table>

Figure 20 shows a part from the same alignment as Figure 19, but here is the DNA sequence translated into a protein sequence. Most base changes occur at the third base of a codon, which will not affect the resulting amino acid. A mutation of the first base will sometimes change the amino acid, whereas a mutation of the second base always produces an alternate amino acid. Most amino acid alterations are unnoticed, as the new altered amino acid falls in the same category (acid, basic, polar and non-polar) as the amino acid it was before. The green letter codes in Figure 20 are polar amino acids, yellow are non polar amino acids, blue are basic amino acids and red are acidic amino acids. Examples of this are encircled in black in Figure 20. Alteration of the amino acid between the different groups of proteins has much more effect, unless the amino acid plays little part in the activity of the protein. These examples are encircled in red.
Figure 20: Protein sequence of a part of ND6 of the carpet python. Yellow colored amino acids are non polar amino acids, green are polar-, blue are basic- and red are acidic amino acids. The black encircled amino acids show alterations but the category stays the same. The red encircled amino acids show alterations from an amino acid in another group. This can have effect on the activity of the protein.

3.4.2 Smallest fragment length needed

The fragment used in for this study was 560 bp long. If working with degraded samples, the shorter the fragment the better for the chance of obtaining good results. Here for, different parts of the ND6 gene with a lot of variation were used to produce a tree to see how small the fragment could be to obtain similar results as the tree which is produced based on the whole fragment. This tree is visible in Figure 21. A 90 bp long fragment is found which starts at base 290 and ends at base 380. Figure 22 shows the corresponding tree. Encircled are parts of the trees which correspond.
As visible in these two trees, all samples from the east coast group together. Within this group, the sample from Brisbane and South Cairns are very similar. Also the samples from New South Wales, Goyder Lagoon and East Adelaide, which are close to each other and a bit more inland, group together as well in both trees. The samples from north Australia (Melville Island, Darwin and North Western Australia) and Papua New Guinea form a group as well and the sample from South Western Australia differs a lot from the rest in both trees. The out-group is the Python Regius, which is another python family.

3.5 Species identification using Cytochrome b

The same samples and methods are used to analyze Cytochrome b for using this gen for species identification of the carpet python.

3.5.1 Intraspecies variation

Table 6 shows the amount and percentage of intraspecies variation between a couple of samples. The average deviation is 4.45% between two subspecies of the carpet python.
Table 6: Amount and percentage of intraspecies variation in Cytochrome b of the carpet python.

<table>
<thead>
<tr>
<th></th>
<th>Brisbane</th>
<th>New South Wales</th>
<th>Darwin</th>
<th>North McKay</th>
<th>Western Australia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brisbane</td>
<td>X</td>
<td>13/1114 1.17%</td>
<td>54/1114 4.85%</td>
<td>8/1114 0.72%</td>
<td>70/1114 6.28%</td>
</tr>
<tr>
<td>New South Wales</td>
<td>X</td>
<td>X</td>
<td>56/1114 5.03%</td>
<td>15/1114 1.35%</td>
<td>76/1114 6.82%</td>
</tr>
<tr>
<td>Darwin</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>56/1114 5.03%</td>
<td>74/1114 6.64%</td>
</tr>
<tr>
<td>North McKay</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>74/1114 6.64%</td>
</tr>
<tr>
<td>Western Australia</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

3.5.2 Smallest fragment length needed

For this gene, there was information available on GenBank from other Morelia species, which is analyzed within this study as well.

The phylogenetic tree which was produced from the entire Cytochrome b gen can be found in Figure 23. A fragment of 275bp was found at the end of Cytochrome b (from base 800 to 1075) which is most similar to the tree obtained from the entire gene. This tree can be found in Figure 24.
Snake species identification using mitochondrial gene loci – Ine van Velzen

Figure 23: UPGMA tree based on the entire cytochrome b gene.
As shown in these trees, it is obvious that all different *Morelia* species do differ a lot from each other. Most of these species group on the same way in both trees except of the *Morelia carinata*.

The *Morelia bredli* coincide with the *Morelia spilota*. This will be further discussed in paragraph 3.6 entitled ‘Taxonomy of the carpet python’.

Within the *Morelia spilota* the samples are more similar to each other, as expected. The samples from the east coast (pink encircled) are very similar to each other and again, the sample from south Western Australia do differ the most from the rest of the samples.
3.6 Taxonomy of the carpet python

Different trees (Maximum Likelihood-, Neighbour Joining-, Minimum Evolution-, UPGMA- and the Maximum Parsimony tree) were generated which were very similar to each other, but the tree which fits best with the taxonomy known so far, was the UPGMA tree. Figure 25 shows the phylogenetic tree based on ND6, Figure 26 shows the tree based on cytochrome b. Each subspecies is encircled by another color which stands for another subspecies. From Goyder Lagoon, which is north east of South Australia and Papua New Guinea, it is unknown what species it was expected to be.

Figure 25: UPGMA tree of the carpet python based on ND6. Every colour represents a different subspecies as known in the taxonomy.
Figure 26: UPGMA tree of the carpet python based on cytochrome b. Every colour represents a different subspecies as known in the taxonomy.

The python regius is taken as out-group as it is a python from another family so this one would differ enough from the carpet python species.

The sample from south Western Australia differs the most from the rest of the samples. It is very obvious that the samples from north Australia and Papua New Guinea differ from the samples from the east coast. The samples from the east coast are genetically very similar to each other. The two different subspecies from the east coast (yellow and red encircled) do not group together correctly. It is possible that these subspecies are so similar together that they are not really different subspecies anymore, the taxonomy is defined many years ago and it can change. Another possibility is that you can never be sure if the location given to the sample is the real location where the sample is from. Finally, more samples are needed for a more specific and more adequate tree.
3.6.1 Comparing with results of previous research

As mentioned before, the carpet python, *Morelia spilota*, is divided into six subspecies according to current taxonomy. Unpublished research though, discovered that most of the north and east of Australia is genetically similar as would be expected if there was no separate subspecies. In the southeast part of Australia lives one of the subspecies, the *Morelia spilota imbricata*. If there are subspecies of the *Morelia spilota* it may be less than six subspecies. Finally, in central Australia lives the *Morelia bredli*, which is another species. This species appears to be very similar to the *Morelia spilota*, and therefore this species is also known as the *Morelia spilota bredli*.

Figure 27 is an overview of the classification according to this research.

![Figure 27: Map of Australia with the habitation of the *Morelia spilota* and the *Morelia imbricata*. The pink colored area in this map shows where the *Morelia spilota* lives, which are genetically all similar so no division into five different subspecies in that area. The green colored area in this map is where the *Morelia spilota imbricata* lives and finally, the yellow area is where the *Morelia Bredli* would live, another species.](image)

The two different phylogenetic trees created during this research are displayed again in Figure 28 (based on ND6) and Figure 29 (based on cytochrome *b*), coloured according this new classification.
Figure 28: UPGMA tree based on ND6 divided into another classification: samples of north and east Australia which group together and the sample from south west Australia which really differs from the rest.

Figure 29: UPGMA tree based on cytochrome b. The samples are colored according to previous research.
According to previous research, it was expected that in south Western Australia would live another snake subspecies anyway, the *Morelia spilota imbricata*, which matches these obtained results as this sample differs a lot from the rest. Furthermore, the samples from north Australia and Papua New Guinea do differ from the samples from the east coast. These last samples are all very similar to each other.

The *Morelia bredli*, which is included in the tree based on cytochrome b, falls in the group of the *Morelia spilota* samples. These results show that this species probably can be better seen as the *Morelia spilota bredli*, one of the subspecies of the *Morelia spilota*. 
4. CONCLUSION

Compared to ND6, in cytochrome b is in general more intraspecies variation. Although there is less variation within the samples from the east coast, the variation between the samples from Darwin, South Western Australia and the east coast is much greater.

Although there is not a good separation between the subspecies on the east coast of Australia, both genes could be used for species identification of the carpet python. There are very clear differences between locations in Australia and there is a lot of intraspecies variation between these subspecies.

According to the results obtained during the taxonomy study, it seems like the carpet python can be divided into four groups; east coast of Australia, north Australia (Morelia spilota variegata), central Australia (Morelia spilota bredli) and west-south Australia (Morelia spilota imbricata).
5. REFERENCES


[25] Ciavaglia, S., Powerpoint presentation about PhD Project: *Investigation and design of python DNA Markers for forensic application*


