CHAPTER 3 RESULTS AND DISCUSSIONS

The aim of the first part of this section of the thesis was to identify parts of the avian mitochondrial genome that are capable of identifying to species level any unknown sample. As part of this study it would also be tested to see if the same sections of DNA are capable of reconstructing the phylogenies from species level, through genus, Family to Order.

3.1 Mitochondrial protein sequences analysis

It would be expected that regions of each of the protein coding mitochondrial genes are conserved or variable depending on the function of the protein that they encoded. This part of the thesis starts with the protein alignment of the ND family genes and the COII and COIII genes compared to the cyt *b* and the COI gene loci, as these last two loci have been used in the identification of species and taxonomic studies [1-15] . Variation and conservation of the protein sequences should be useful for species testing and in the design of primers that are either universal for all the species to be tested or are species specific. In this study, amino acid sequences were obtained 33 species spanning 13 Orders listed on GenBank (Table 2.15). These included: Galliformes, Anseriformes, Falconiformes, Tinamiformes, Struthioniformes, Ciconiiformes, Pelecaniformes, Sphenisciformes, Charadriiformes, Passeriformes, Podicipediformes, Gruiformes and Gaviiformes.

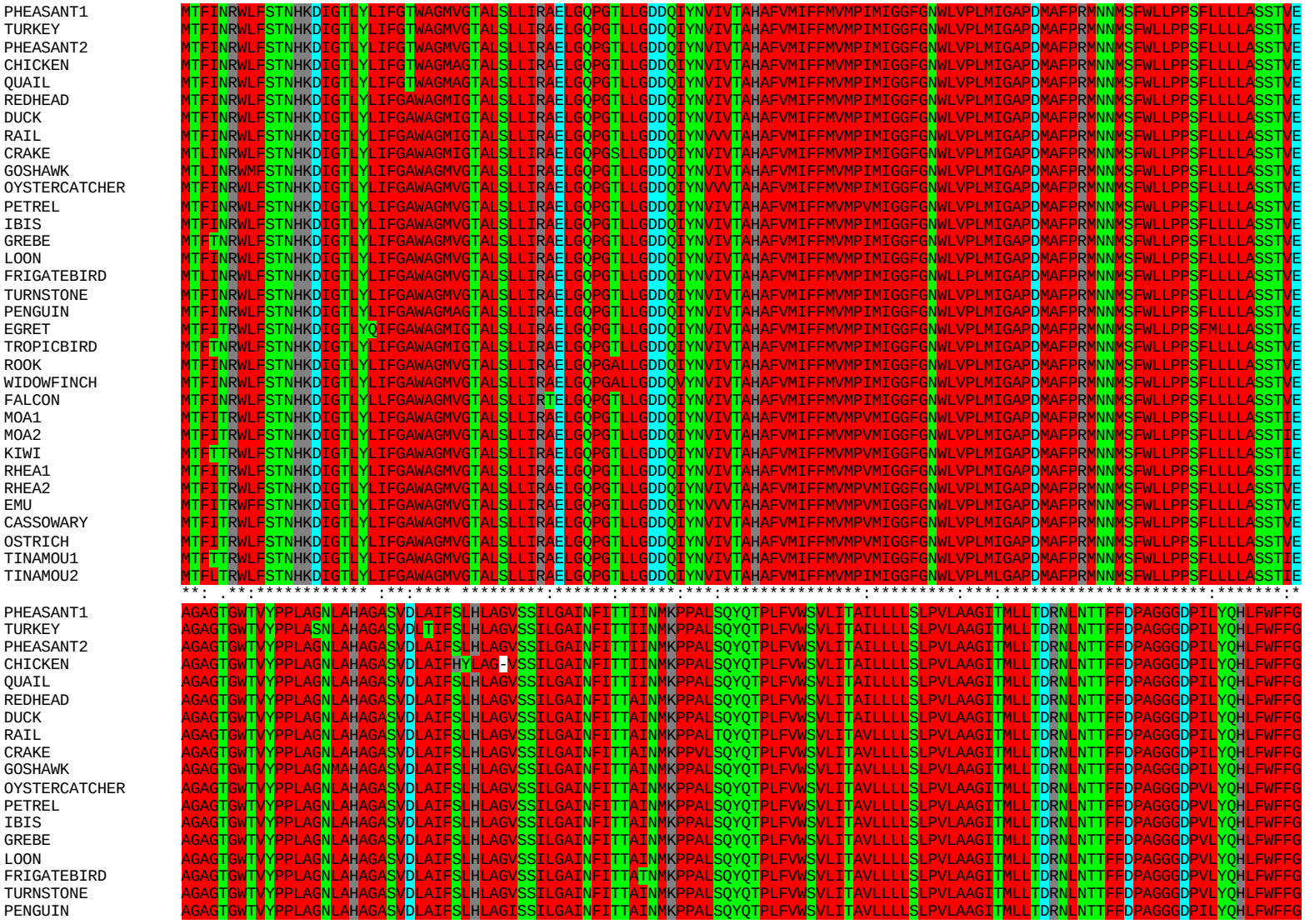
The amino acid length of each locus, percent homology for the total number of variable sites and the percentage variability within the complete amino acid sequence, which are shown in Table 3.1, were generated as part of this study.

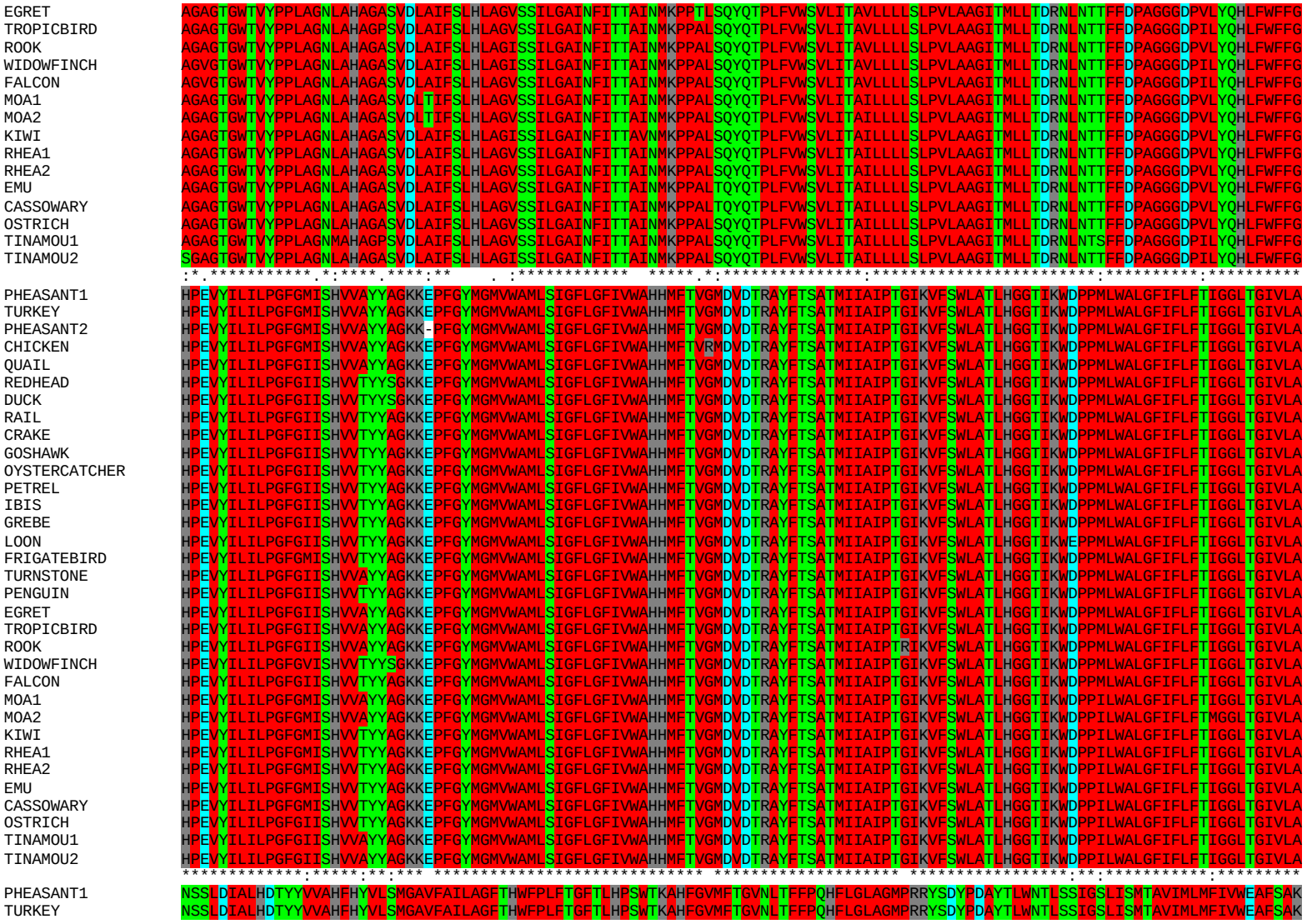
Table 3.1: showing protein length of each locus, percent homology, total number of variable sites and percent variable sites within the complete protein sequences of the ND family, COI, COII, COIII and cyt *b* loci based on 33 avian species chosen from Galliformes, Anseriformes, Falconiformes, Tinamiformes, Struthioniformes, Ciconiiformes, Pelecaniformes, Sphenisciformes, Charadriiformes, Passeriformes, Podicipediformes, Gruiformes and Gaviiformes.

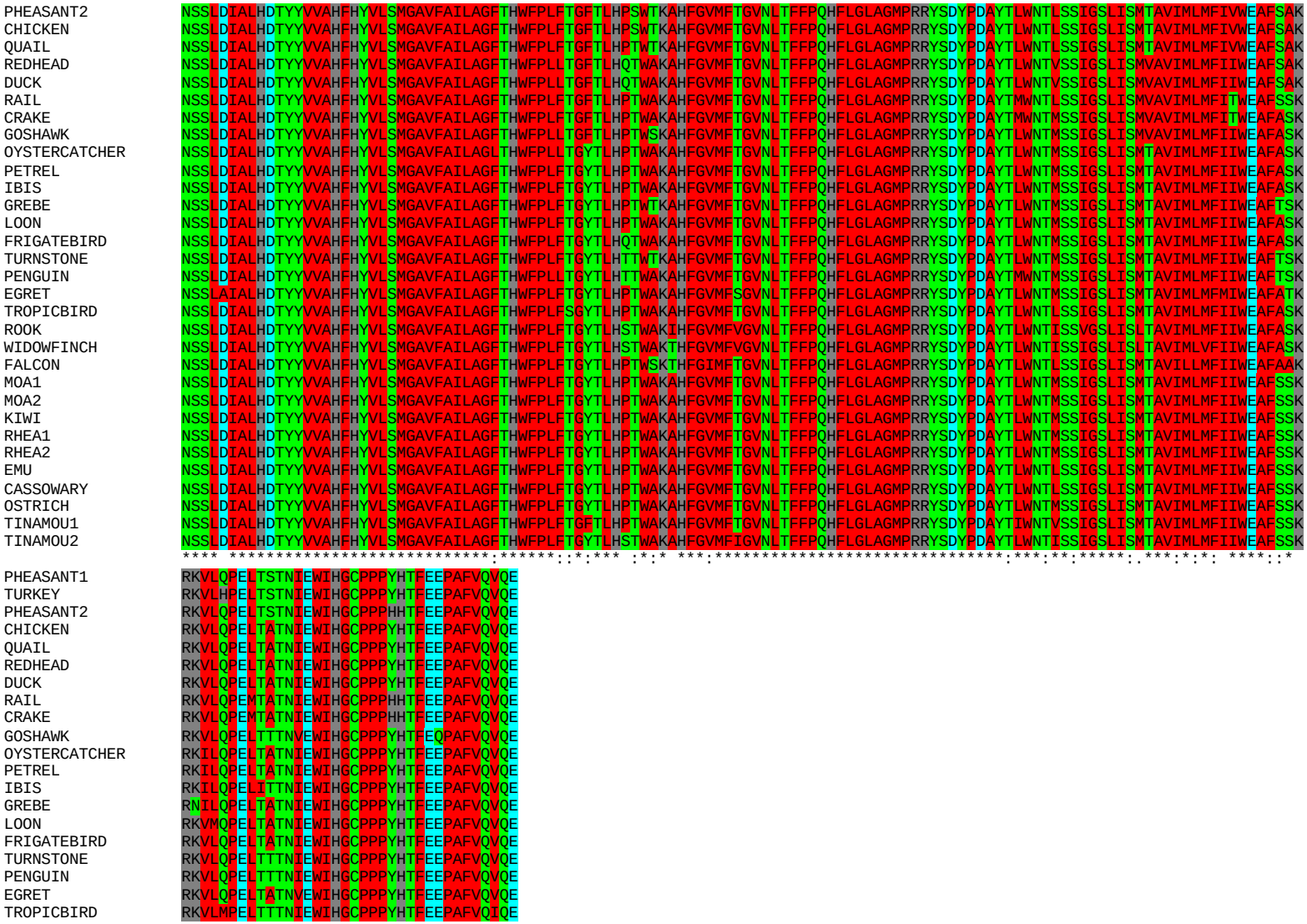
Loci	Length (amino acids)	% Homology based on 29 avian species	Total number of variable sites	% Variable sites within each amino acid sequences
ND1	325	89.6	119	36.62
ND2	346	78.2	210	60.69
ND3	116	84.6	43	37.07
ND4L	99	83.6	61	61.62
ND4	459	84.1	224	48.80
ND5	600	81.7	357	59.50
ND6	175	77.3	82	48.86
COI	516	96.7	85	16.47
COII	232	91.3	81	34.91
COIII	262	91.3	84	32.61
cyt *b*	350	90.5	125	35.71

The amino acid sequence of the the ND family, COI, COII, COIII and cyt *b* loci were selected from complete genomes of the 33 avian species. Based on those species the COI locus is the most highly conserved sequence with 96.7 % homology over the 516 amino acids; in comparison the cyt *b* gene has a comparable similarity of 90.5 % for the 350 amino acids. In contrast, the amino acid sequences of the ND2 and the ND5 genes are the most variable loci with 78.2 and 81.7 % similarity respectively. These two loci were selected for further analysis due to this variation at the amino acid level, and therefore potential variation at the DNA level.

The loci ND2 and ND5 were also compared in future analyses with the cyt *b* and COI loci as these two loci are the most frequently used in species testing and phylogenetics. Protein multiple alignment results from 33 species of the COI, cyt *b*, ND2 and ND5 genes can be found in Figures 3.1 through to 3.4; similar data for the COII, CIII, ND1, ND3, ND4, ND4L and ND6 loci are provided in Appendix A. All protein sequences were obtained from Genbank. The quality of each sequences were controlled by selecting only the coding sequences obtained from complete mitochondrial genomes and each coding sequences were checked to minimise any sequence errors. The translated protein sequences of these analysed DNA loci were subsequently used in this part of study.


PHEASANT1
TURKEY
PHEASANT2
CHICKEN
QUAIL
REDHEAD
DUCK
RAIL
CRAKE
GOSHAWK
OYSTERCATCHER
PETREL
IBIS
GREBE
LOON
FRIGATEBIRD
TURNSTONE
PENGUIN
EGRET
TROPICBIRD
ROOK
WIDOWFINCH
FALCON
MOA1
MOA2
KIWI
RHEA1
RHEA2
EMU
CASSOWARY
OSTRICH
TINAMOU1
TINAMOU2


EGRET
TROPICBIRD
ROOK
WIDOWFINCH
FALCON
MOA1
MOA2
KIWI
RHEA1
RHEA2
EMU
CASSOWARY
OSTRICH
TINAMOU1
TINAMOU2
PHEASANT1
TURKEY
PHEASANT2
CHICKEN
QUAIL
REDHEAD
DUCK
RAIL
CRAKE
GOSHAWK
OYSTERCATCHER
PETREL
IBIS
GREBE
LOON
FRIGATEBIRD
TURNSTONE
PENGUIN
EGRET
TROPICBIRD
ROOK
WIDOWFINCH
FALCON
MOA1
MOA2
KIWI
RHEA1
RHEA2
EMU
CASSOWARY
OSTRICH
TINAMOU1
TINAMOU2
PHEASANT1
TURKEY


PHEASANT2
CHICKEN
QUAIL
REDHEAD
DUCK
RAIL
CRAKE
GOSHAWK
OYSTERCATCHER
PETREL
IBIS
GREBE
LOON
FRIGATEBIRD
TURNSTONE
PENGUIN
EGRET
TROPICBIRD
ROOK
WIDOWFINCH
FALCON
MOA1
MOA2
KIWI
RHEA1
RHEA2
EMU
CASSOWARY
OSTRICH
TINAMOU1
TINAMOU2
PHEASANT1
TURKEY
PHEASANT2
CHICKEN
QUAIL
REDHEAD
DUCK
RAIL
CRAKE
GOSHAWK
OYSTERCATCHER
PETREL
IBIS
GREBE
LOON
FRIGATEBIRD
TURNSTONE
PENGUIN
EGRET
TROPICBIRD

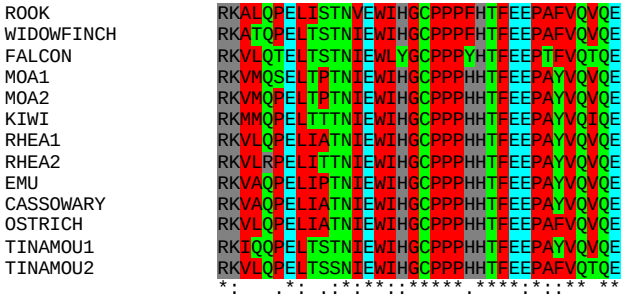
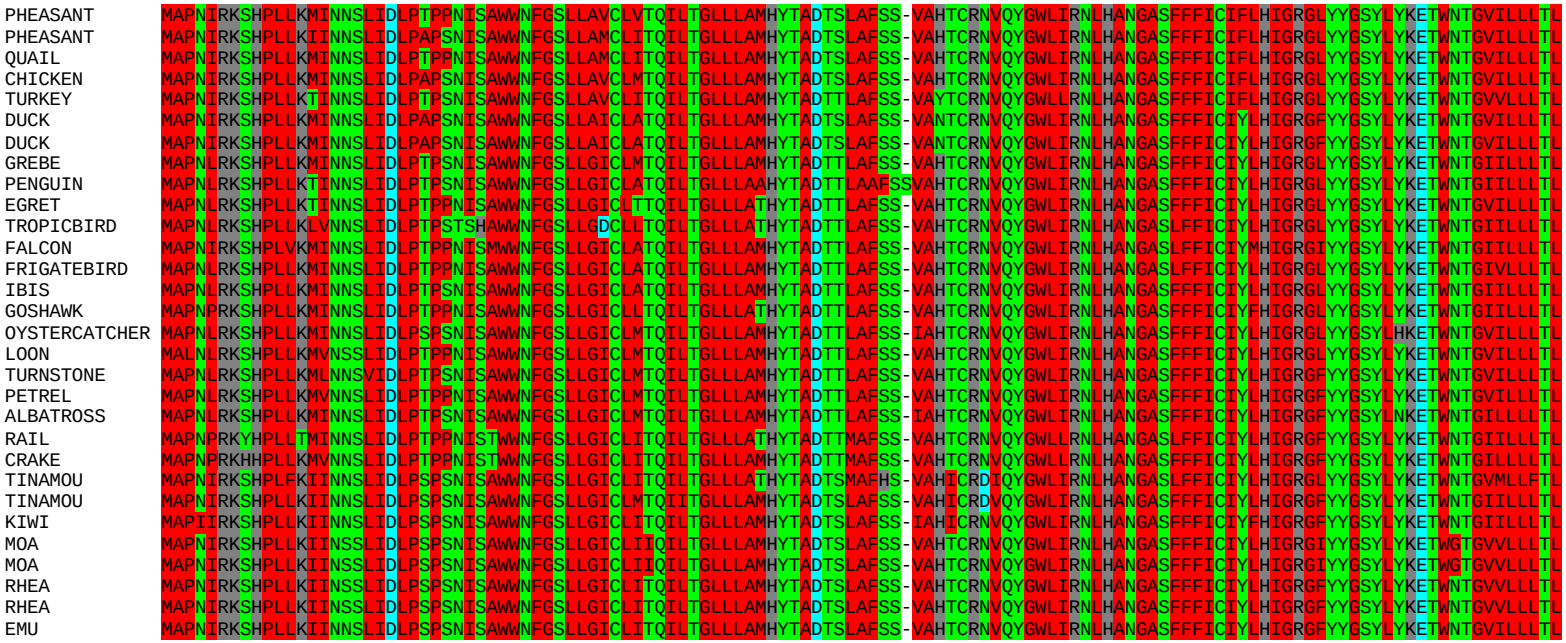
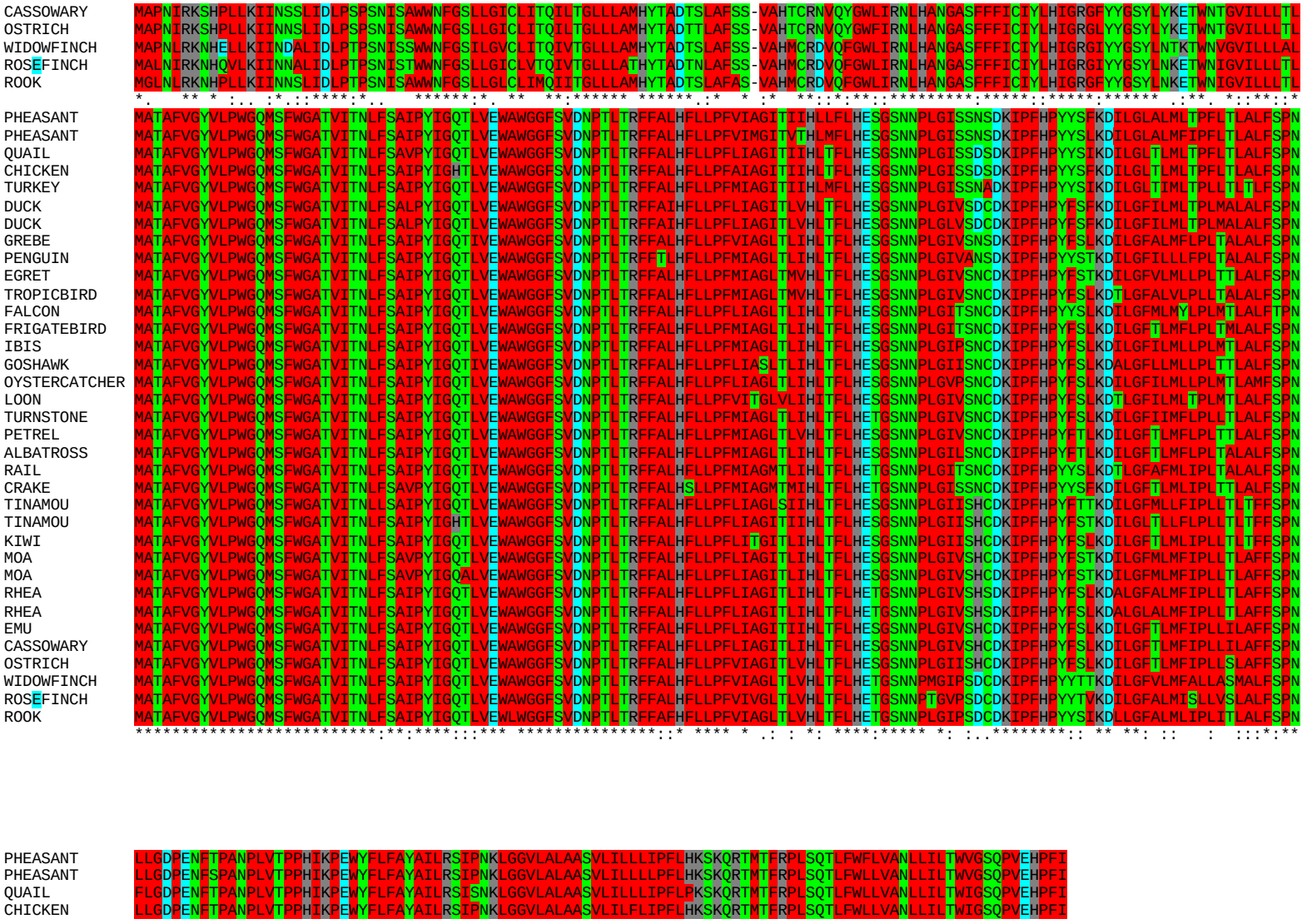


Figure 3.1: Protein alignment of COI gene in 33 different avian species using MEGA 4 program. The different colours are used to indicate the different groups of amino acid based on their side-chain properties; where polar is green, non-polar is red, acidic is blue and basic is grey.




CASSOWARY
OSTRICH
WIDOWFINCH
ROSEFINCH
ROOK
PHEASANT
PHEASANT
QUAIL
CHICKEN
TURKEY
DUCK
DUCK
GREBE
PENGUIN
EGRET
TROPICBIRD
FALCON
FRIGATEBIRD
IBIS
GOSHAWK
OYSTERCATCHER
LOON
TURNSTONE
PETREL
ALBATROSS
RAIL
CRAKE
TINAMOU
TINAMOU
KIWI
MOA
MOA
RHEA
RHEA
EMU
CASSOWARY
OSTRICH
WIDOWFINCH
ROSEFINCH
ROOK
PHEASANT
PHEASANT
QUAIL
CHICKEN

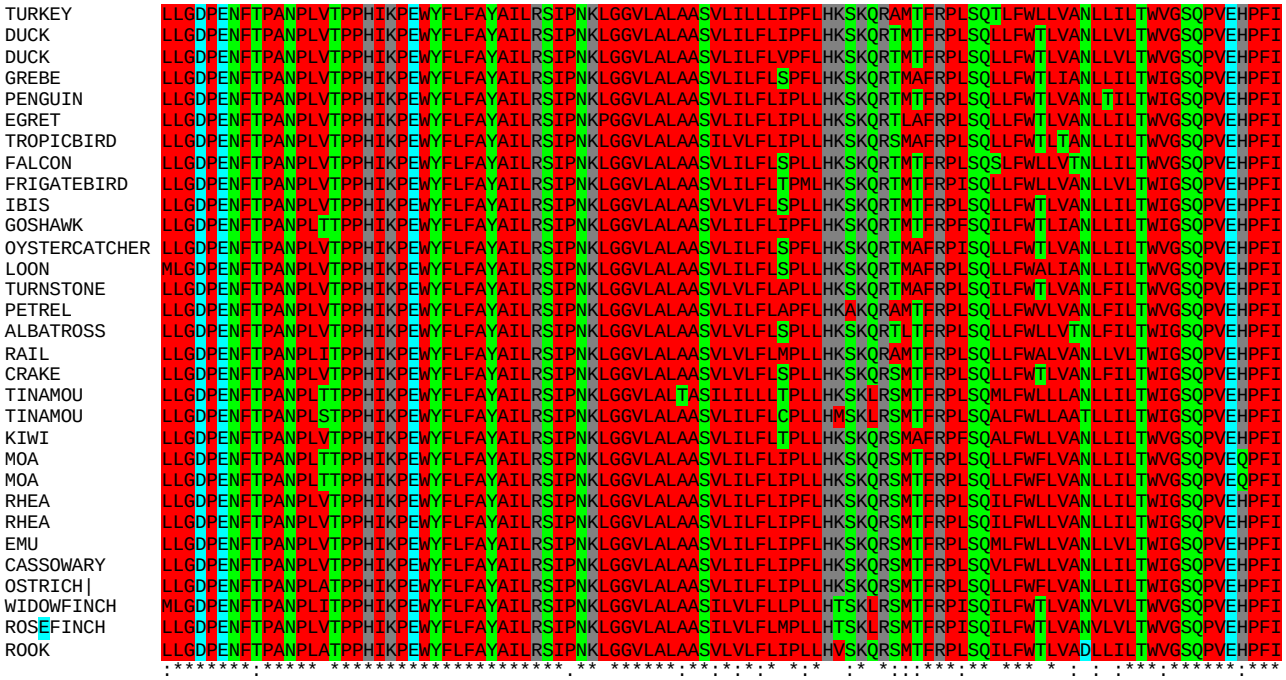
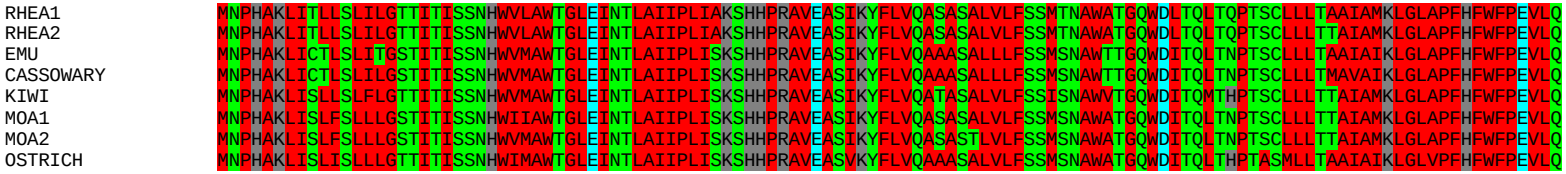
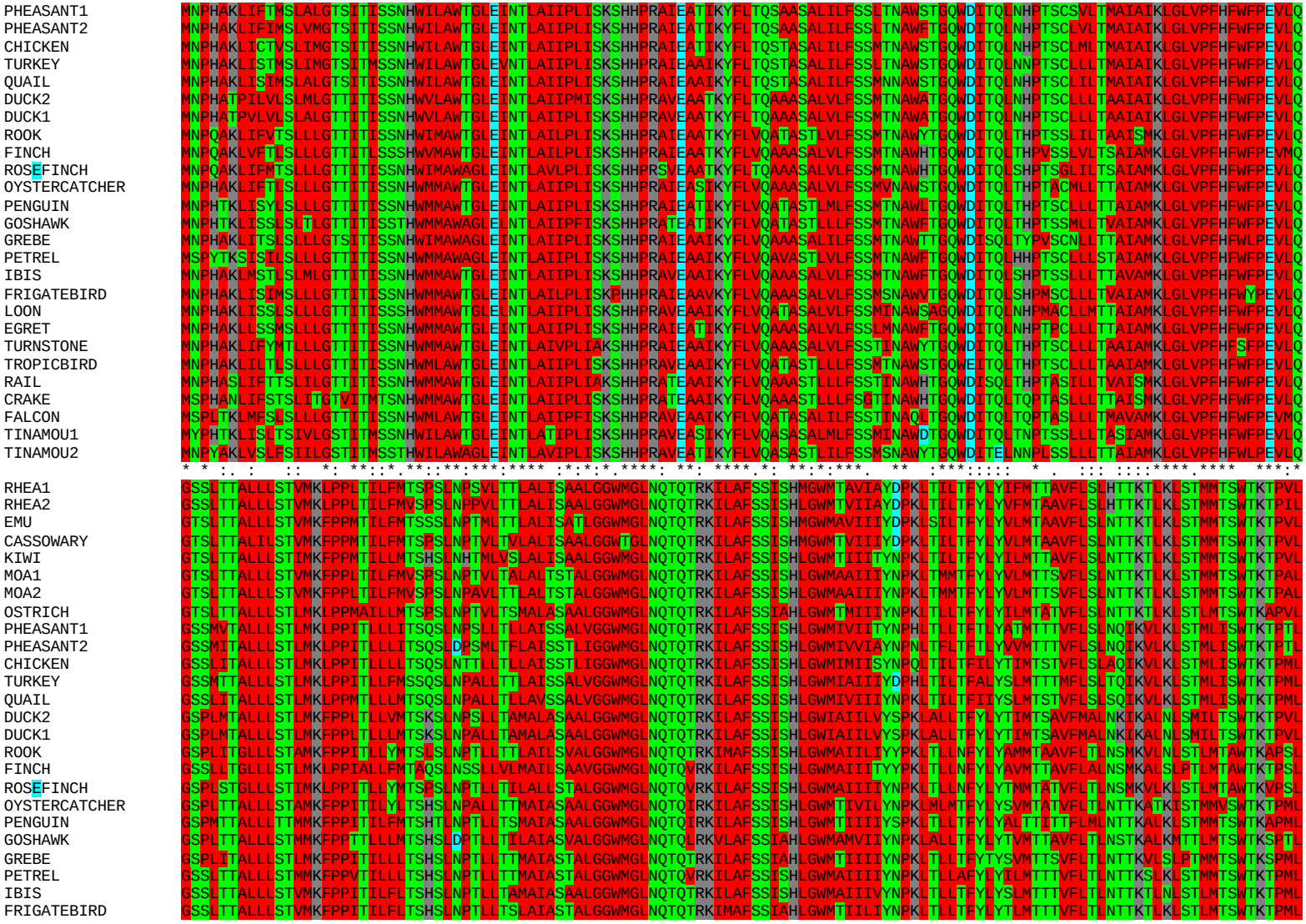


Figure 3.2: Protein alignment of cyt *b* gene in 33 different avian species using MEGA 4 program. The different colours are used to indicate the different groups of amino acid based on their side-chain properties; where polar is green, non-polar is red, acidic is blue and basic is grey.




PHEASANT1
PHEASANT2
CHICKEN
TURKEY
QUAIL
DUCK2
DUCK1
ROOK
FINCH
ROSEFINCH
OYSTERCATCHER
PENGUIN
GOSHAWK
GREBE
PETREL
IBIS
FRIGATEBIRD
LOON
EGRET
TURNSTONE
TROPICBIRD
RAIL
CRAKE
FALCON
TINAMOU1
TINAMOU2
RHEA1
RHEA2
EMU
CASSOWARY
KIWI
MOA1
MOA2
OSTRICH
PHEASANT1
PHEASANT2
CHICKEN
TURKEY
QUAIL
DUCK2
DUCK1
ROOK
FINCH
ROSEFINCH
OYSTERCATCHER
PENGUIN
GOSHAWK
GREBE
PETREL
IBIS
FRIGATEBIRD

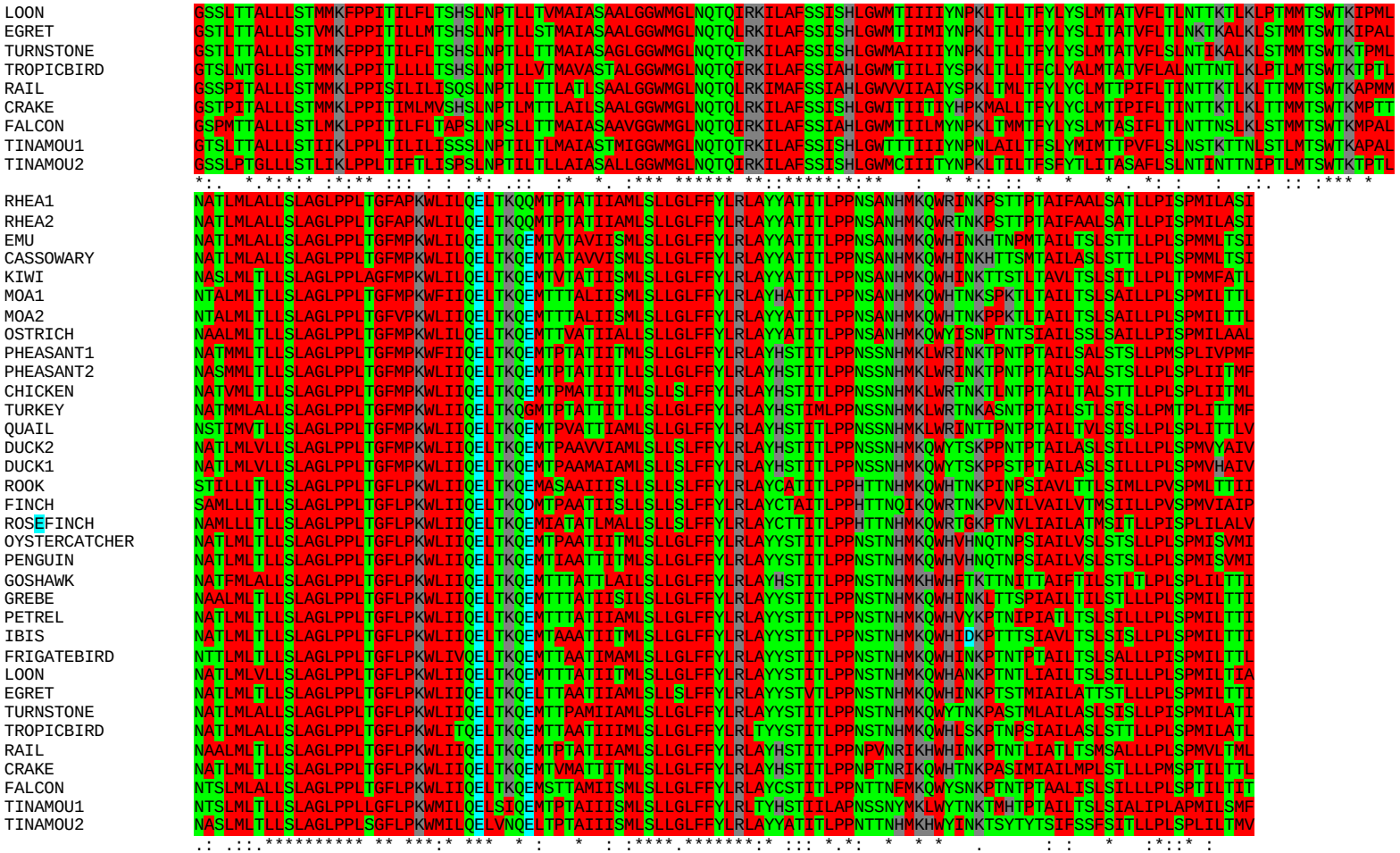
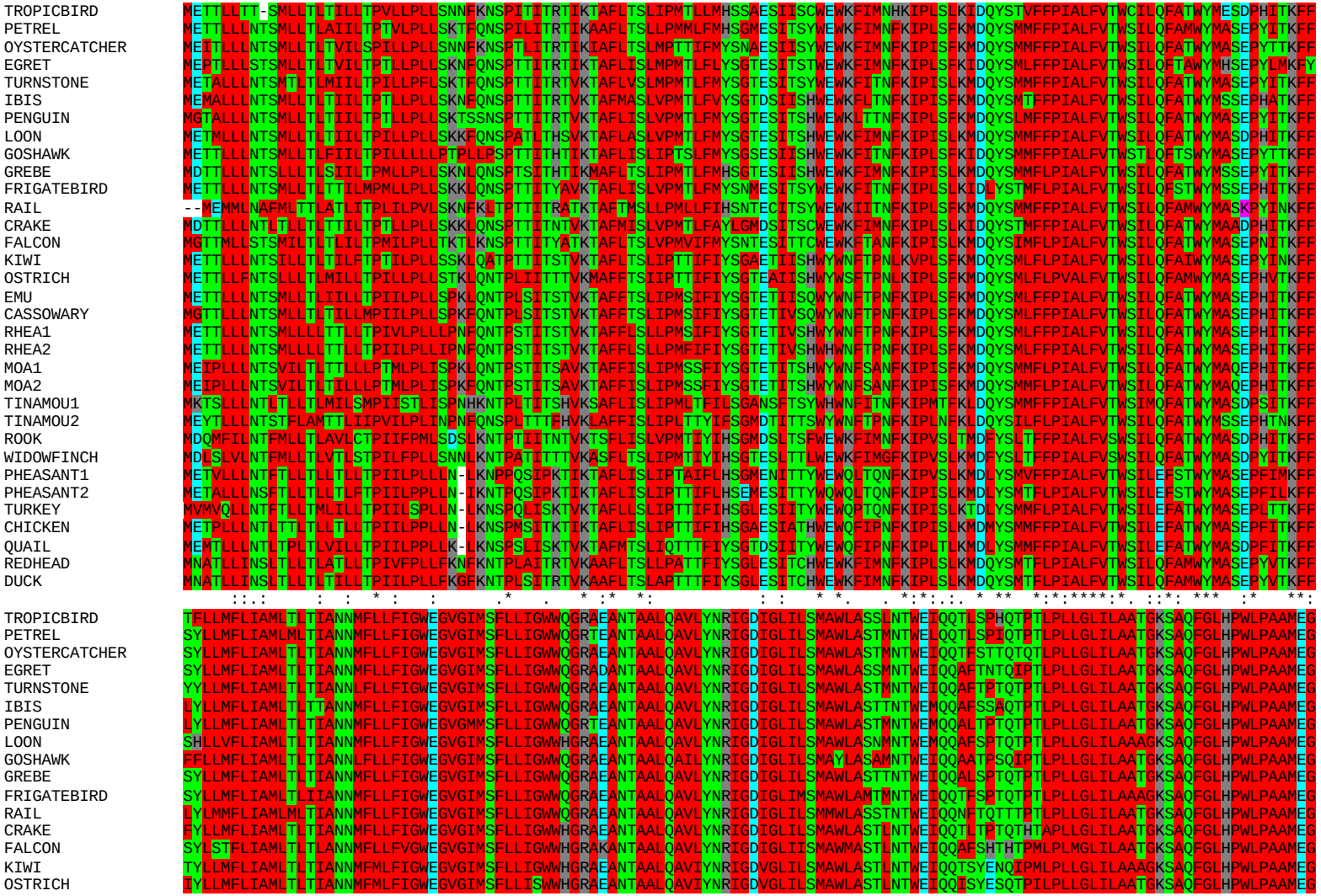
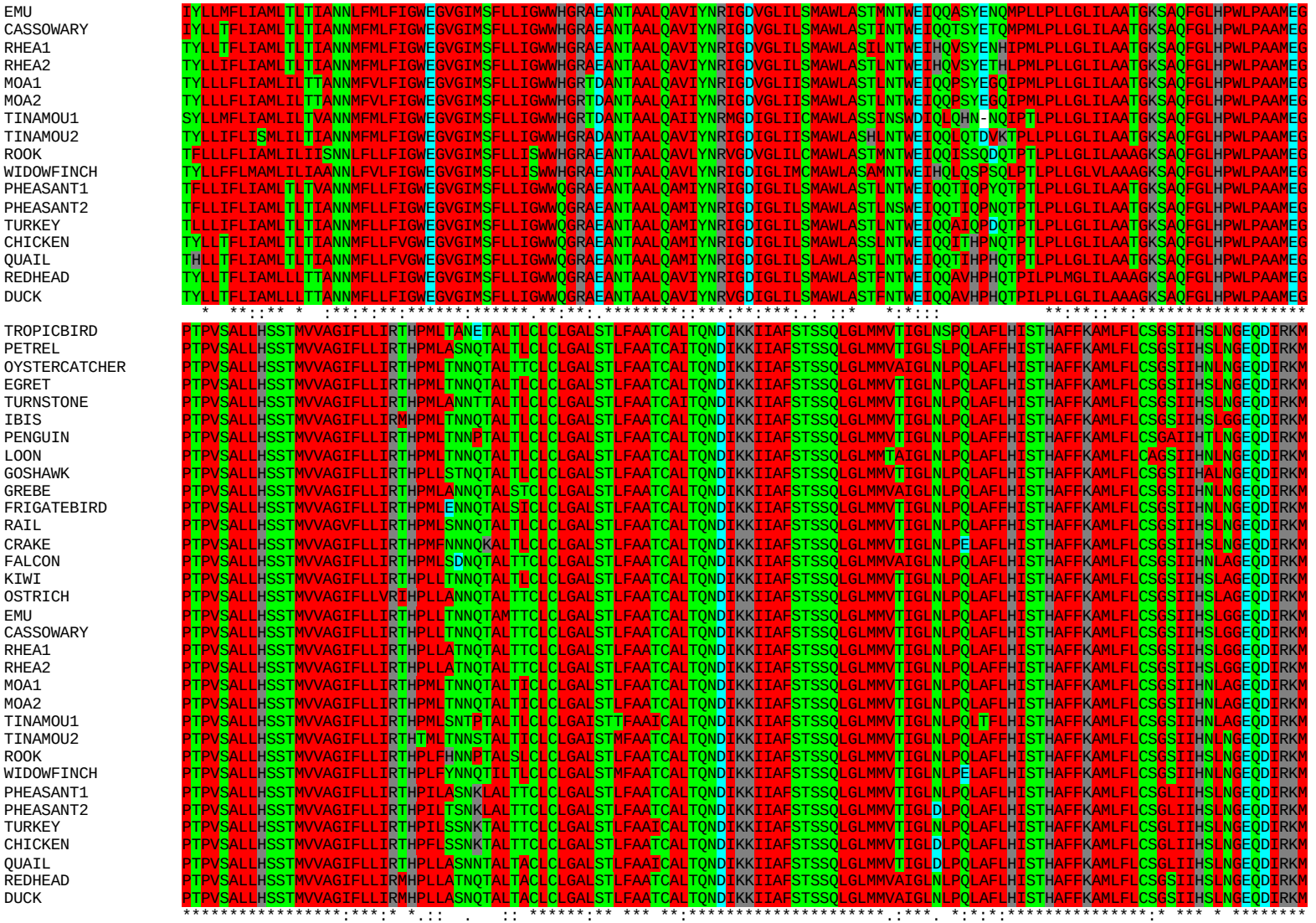


Figure 3.3: Protein alignment of ND2 gene in 33 different avian species using MEGA 4 program. The different colours are used to indicate the different groups of amino acid based on their side-chain properties; where polar is green, non-polar is red, acidic is blue and basic is grey.


TROPICBIRD
PETREL
OYSTERCATCHER
EGRET
TURNSTONE
IBIS
PENGUIN
LOON
GOSHAWK
GREBE
FRIGATEBIRD
RAIL
CRAKE
FALCON
KIWI
OSTRICH
EMU
CASSOWARY
RHEA1
RHEA2
MOA1
MOA2
TINAMOU1
TINAMOU2
ROOK
WIDOWFINCH
PHEASANT1
PHEASANT2
TURKEY
CHICKEN
QUAIL
REDHEAD
DUCK


EMU
CASSOWARY
RHEA1
RHEA2
MOA1
MOA2
TINAMOU1
TINAMOU2
ROOK
WIDOWFINCH
PHEASANT1
PHEASANT2
TURKEY
CHICKEN
QUAIL
REDHEAD
DUCK
TROPICBIRD
PETREL
OYSTERCATCHER
EGRET
TURNSTONE
IBIS
PENGUIN
LOON
GOSHAWK
GREBE
FRIGATEBIRD
RAIL
CRAKE
FALCON
KIWI
OSTRICH
EMU
CASSOWARY
RHEA1
RHEA2
MOA1
MOA2
TINAMOU1
TINAMOU2
ROOK
WIDOWFINCH
PHEASANT1
PHEASANT2
TURKEY
CHICKEN
QUAIL
REDHEAD
DUCK

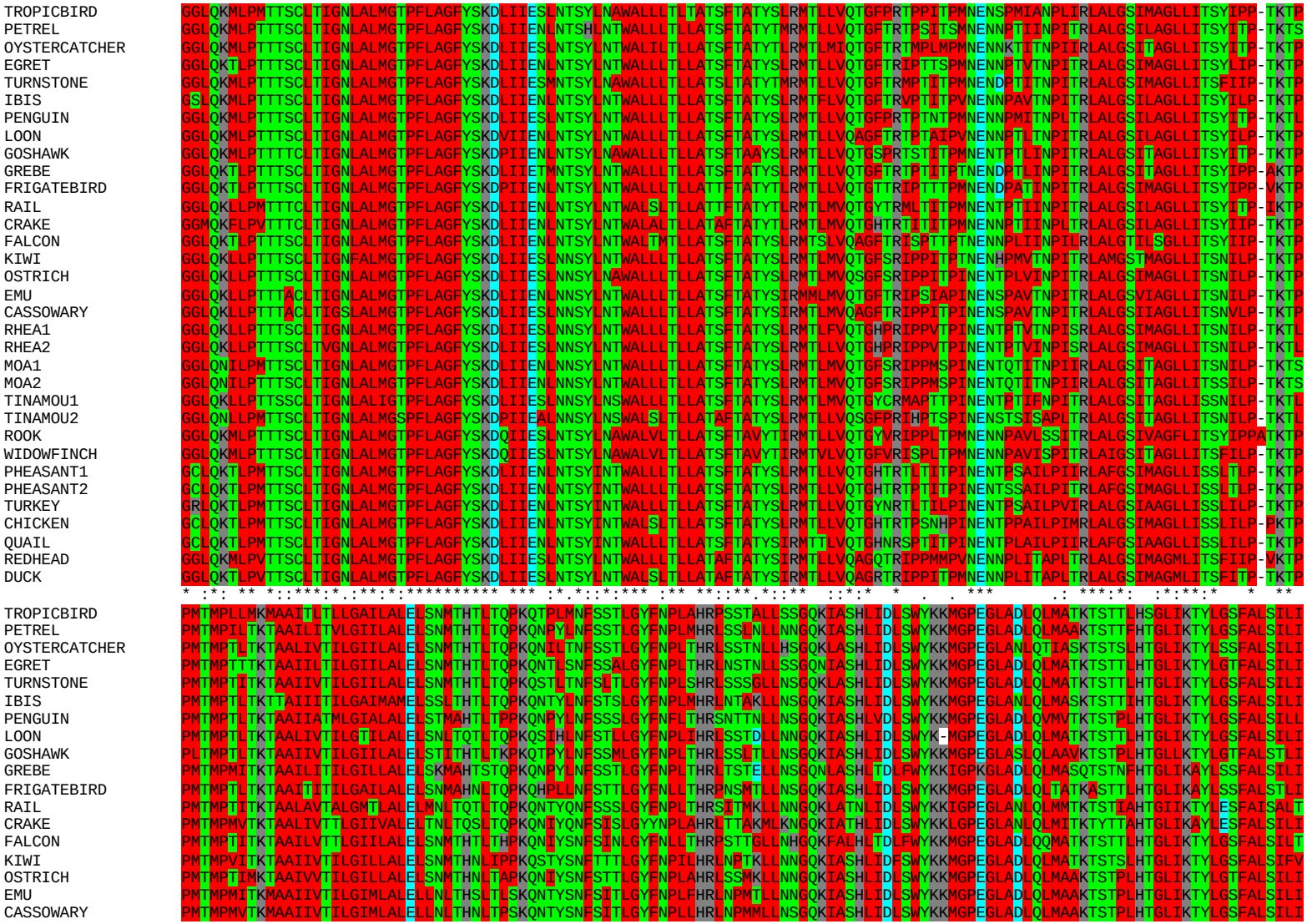

TROPICBIRD
PETREL
OYSTERCATCHER
EGRET
TURNSTONE
IBIS
PENGUIN
LOON
GOSHAWK
GREBE
FRIGATEBIRD
RAIL
CRAKE
FALCON
KIWI
OSTRICH
EMU
CASSOWARY
RHEA1
RHEA2
MOA1
MOA2
TINAMOU1
TINAMOU2
ROOK
WIDOWFINCH
PHEASANT1
PHEASANT2
TURKEY
CHICKEN
QUAIL
REDHEAD
DUCK

Figure 3.4: Protein alignment of ND5 gene in 33 different avian species using MEGA 4 program. The different colours are used to indicate the different groups of amino acid based on their side-chain properties; where polar is green, non-polar is red, acidic is blue and basic is grey.

The DNA Barcoding region is approximately a 648 bp region positioned near the 5' terminus of the COI gene [9, 16-20] , while the 3’ terminus of this gene has not been used previously for species identification. This locus is shown in Figure 3.5. The partial sequences of the cyt *b* gene from both 5’ and 3’ end have been reported previously for avian species testing and phylogeny study [2, 21-25]. Both COI and cyt *b* have been widely used in species testing of many species [2, 10, 16, 22-39] including avian species, while the ND2 and ND5 loci have been rarely used in species identification; this is especially the case in avian species identification.

Figure 3.5 identifies the areas of ND2, ND5, COI and cyt *b* used most commonly in species identification. This figure also indicates the priming sites used. It is noticeable that there is little correlation between the area examined in species testing and the areas of greatest variation at the amino acid level.

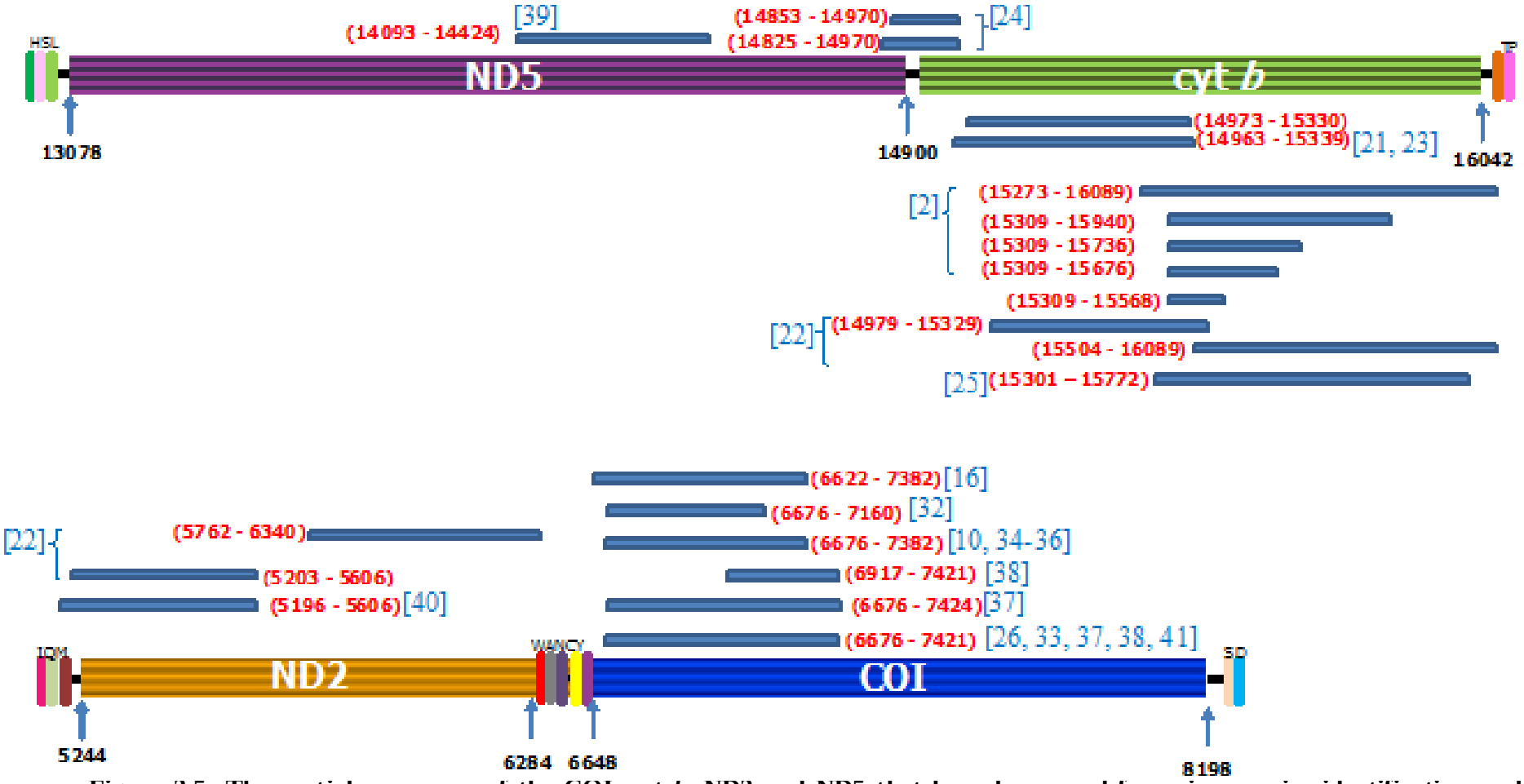


Figure 3.5: The partial sequences of the COI, cyt *b*, ND2 and ND5 that have been used for avian species identification and evolutionary studies. The numbers indicate the base positions according to the mtDNA sequence of *Gallus gallus* (AY235571). Single letters indicate the tRNA gene for each amino acid: H is tRNA-His, S is tRNA-Ser, L is tRNA-Leu, T is tRNA-Thr, P is tRNA-Pro, I is tRNA-Ile, Q is tRNA-Gln, M is tRNA-Met, W is tRNA-Trp, A is tRNA-Ala, N is tRNA-Asn, C is tRNA-Cys, Y is tRNA-Tyr, and D is tRNA-Asp.

The genetic distance of the COI, cyt *b*, ND2 and ND5 amino acid sequences were calculated at every amino acid position using MEGA5 program to show amino acid variation of each loci (Figure 3.6).

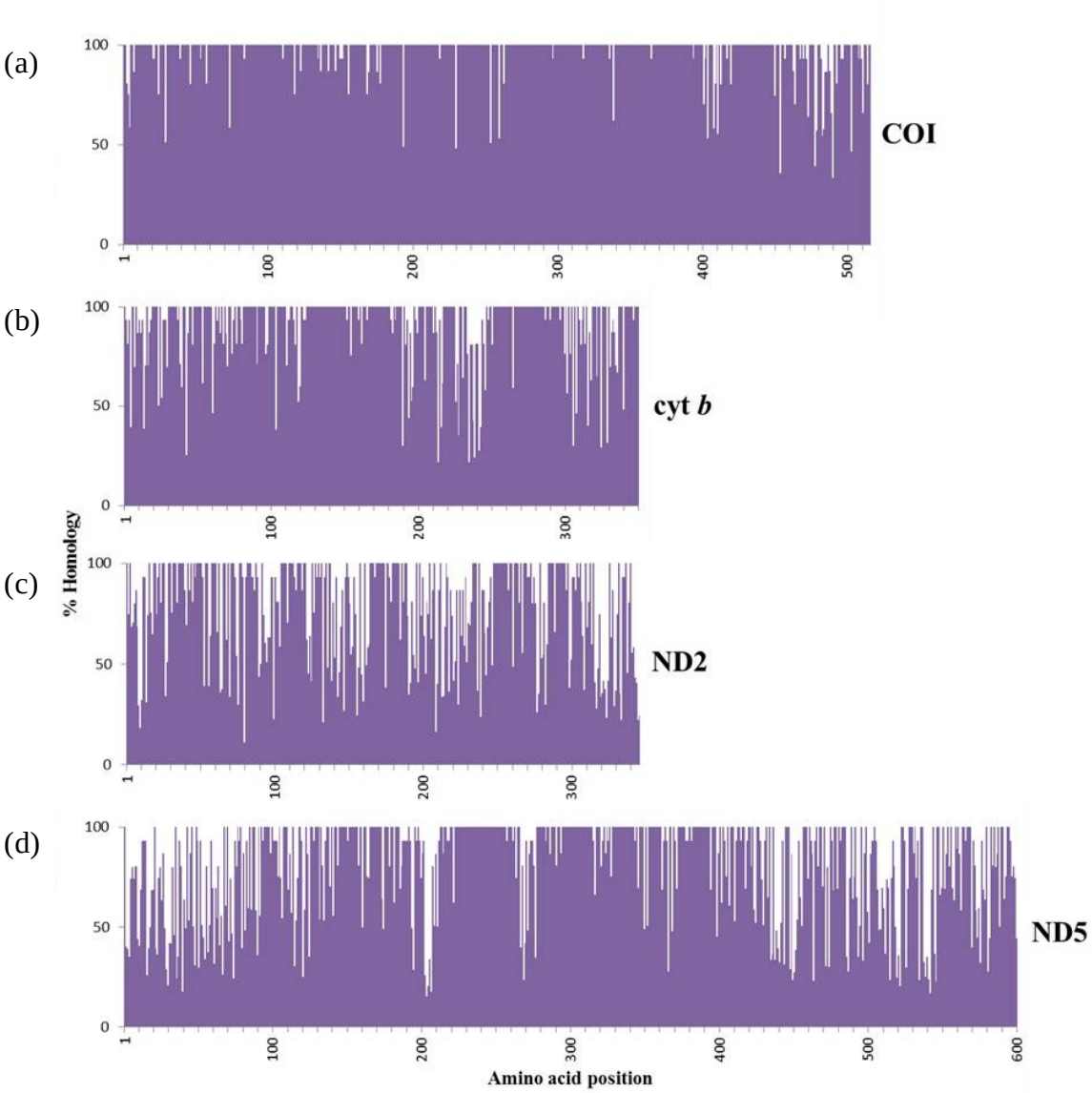


Figure 3.6: Genetic variation at the protein level of the COI (a), cyt *b* (b), ND2 (c) and ND5 (d). The X-axis represents an amino acid position along the four loci and the Y-axis represents the percentage of homology between amino acid of each locus based on 33 avian species, 33 genera, 29 Families and 13 Orders.

The COI locus was found to be the most highly conserved gene of the four loci studied (Figure 3.6a). There are only a few regions showing variation in the amino acid sequence dispersed along the locus with greatest variation from approximately the amino acid at position 400 to the 3' end. This region appears to exhibit the highest variation within this gene, although it is not used as part of the Barcoding locus. The cyt *b* gene encodes a structural protein with some parts of the protein embedded in the mitochondrial membrane and some parts folded outside the inner and outer membrane. If there is a correspondence between the functionality and structure of the cyt *b* gene it would be expected to find about eight regions of highly conserved amino acid domains. The parts of the peptide that fold outside an inner and outer mitochondrial membrane expected exhibit greater variation than the conserved membrane bound amino acids. These regions of higher variability in the human cyt *b* gene are at amino acid positions 1-33, 55-76, 98-113, 158-178, 200-222, 251-288, 310-323 and 340-350 as shown in Figure 3.7. In *Gallus gallus*, the cyt *b* amino acid sequence at amino acid positions 1-30, 54-81, 90-124,152-166, 182-217, 226-265, 297-318 and 329-336 correspond to regions of greater variability as shown in Figure 3.6b. The ND2 and the ND5 genes exhibit lesser conservation compared to the cyt *b* and COI. According to the Figure 3.6, there is less variation within the middle part of the ND5 gene between positions 220-260 with domains of greater variability at the 5' end at amino acid from positions 1-220 and at the 3' from the amino acid positions 260-600 as shown in Figure 3.6d. By contrast regions of high variability occur constantly throughout the ND2 gene with no clear areas of conservation (Figure 3.6c).

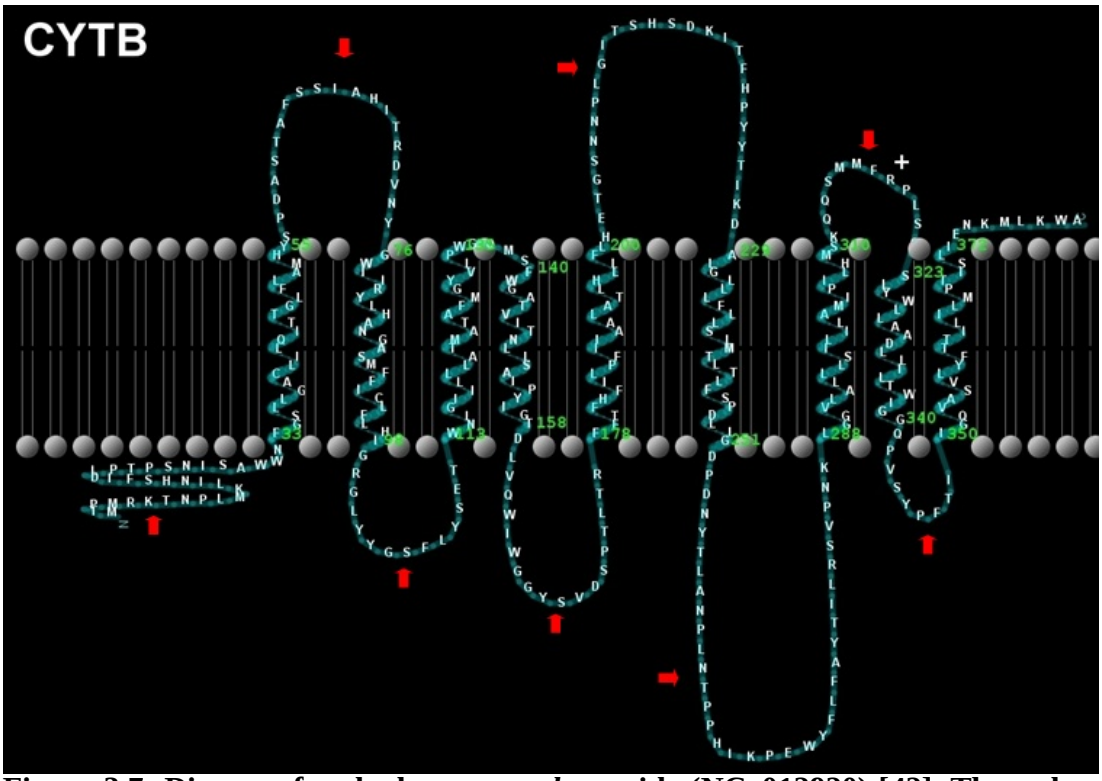


Figure 3.7: Diagram for the human cyt *b* peptide (NC_012920) [42]. The red arrows indicate the parts of the peptide that fold outside an inner and outer mitochondrial membrane which are expected exhibit greater variation than the conserved membrane bound amino acids. The numbers shows the position of the amino acids based on the human sequence.

In summary, the results from Figure 3.1 - 3.4 and 3.6 may be of value in the identification of conserved and variable regions of these four gene loci (ND2, ND5, COI and cyt *b*). If there is a correspondence between the structure and function of the encoded proteins there should be domains of variability and areas of conserved sequence. These conserved regions should be suitable for designing universal primers to amplify internal variable regions; being ideal for the purpose of species identification. In addition, the variable sites which are unique in only one particular species can be used for designing species specific primers as well.

3.2 Mitochondrial DNA sequences analysis

The mitochondrial genomes from a wide range of avian species were chosen from 19 Orders, 40 Families, 75 genera, and 102 different species with the aim of covering all the major avian taxonomic groups. This section of the dissertation examines the use of the entire avian mitochondrial DNA genome to determine whether using this entire sequence is capable of reconstructing the phylogenies as currently defined by taxonomic studies. Subsequently each individual gene locus is used to identify to species level and reconstruct the phylogeny with the fewest anomalies. Finally sections of varying size of particular genes are used for the same purposes.

3.2.1 Phylogenetic reconstruction using mitochondrial genomes

It would be expected that if the taxonomic and genetic alignments were concordant, then MEGA should place all the 102 species in their corresponding taxonomic cluster. The phylogenetic reconstruction using complete mitochondrial DNA sequences (Figure 3.8) resulted in only a few Orders being divided; these included members of the Order Gruiformes and Ciconiiformes. Species representative of the Galliformes, Anseriformes, Struthioniformes, Tinamiformes, Falconiformes and Passeriformes all cluster together; this would be expected as they are members of the same Order.

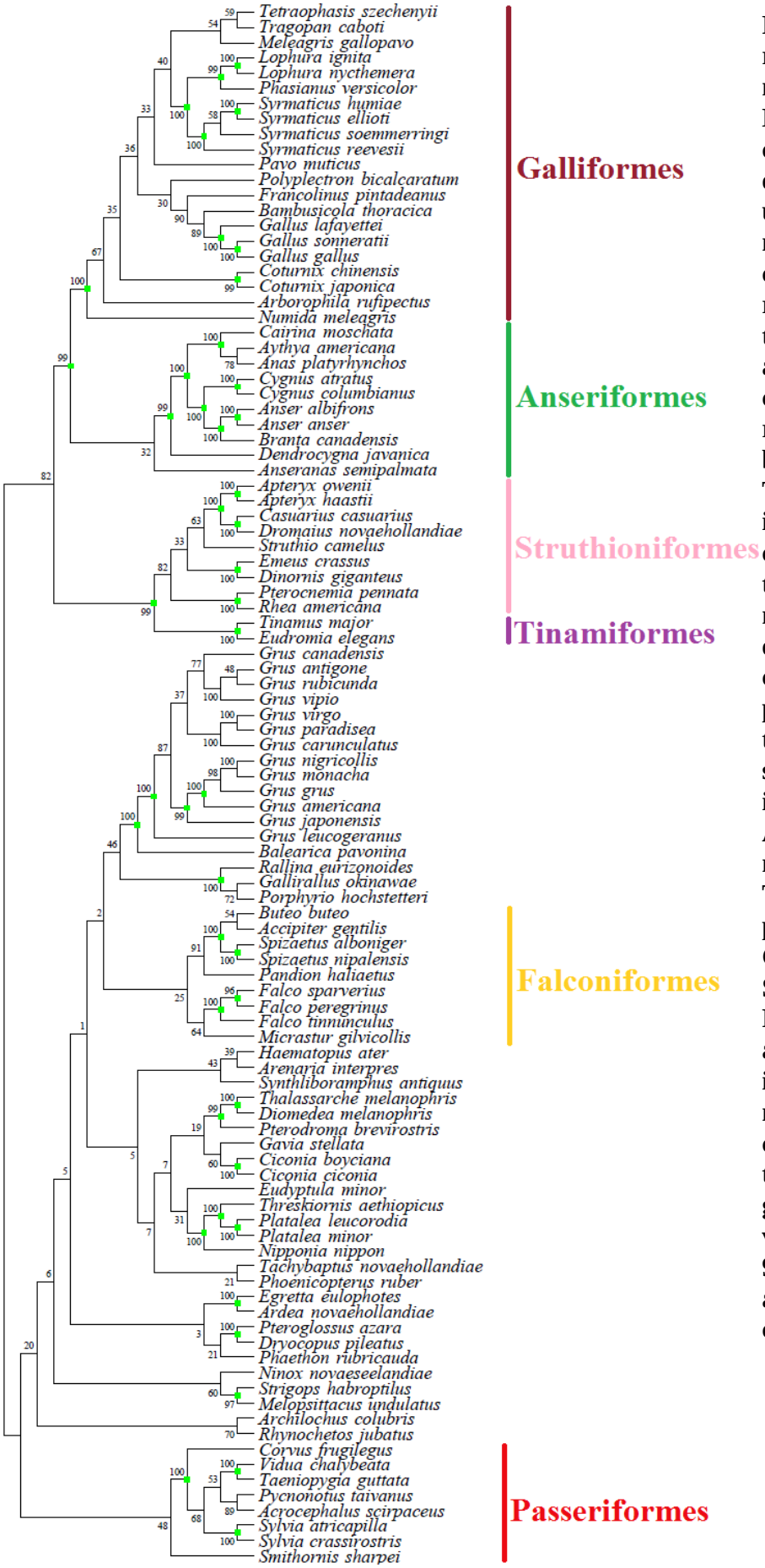


Figure 3.8: Phylogenetic tree reconstruction of the complete mitochondrial genome sequences. Evolutionary analyses were conducted in MEGA5 [43]. The evolutionary history was inferred using the Neighbor-Joining method [44]. The bootstrap consensus tree inferred from 1000 replicates [45] is taken to represent the evolutionary history of the taxa analysed [45]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [45]. The evolutionary distances were computed using the Kimura 2-parameter method [46] and are in the units of the number of base substitutions per site. The analysis involved 102 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 9166 positions in the final dataset. The Galliformes, Anseriformes, Struthioniformes, Tinamiformes, Falconiformes and Passeriformes are clustered together as expected if the genetic data of the whole mitochondrial genome matches the current taxonomic groups at the taxonomic level of Order. The green dots indicate the bootstrap value for a given interior branch is 95% or higher, then the topology at that branch is considered correct [47].

Typically the DNA samples that are encountered in forensic testing are highly degraded or at trace amounts. Loci within the mitochondrial DNA have been used in forensic analysis because of the benefits of high copy number [48-50] and that the mitochondrial genome is within a protective double-membrane. It is unlikely that in a forensic context the whole of the mitochondrial genome will be available for testing, nor is it likely that the time and resources will be taken to sequence the entire mitochondrial genome using current technology. Rather sections of the mitochondrial genome such as gene sequences or, sections of genes, are used; this includes a section of cyt *b* gene [22-25, 34, 36-39] or the COI gene [10, 26-33, 35, 36]. The smallest parts of mitochondrial DNA that can identify species and reconstruct the phylogenetic tree will be ideal for forensic applications.

The DNA sequences of each mitochondrial locus were isolated from the same 102 species as used in the complete mitochondrial genome comparison. Each of the individual gene sequences were extracted to perform multiple alignments from which genetic distances using the program MEGA 5 were calculated. The results are shown in Figure 3.9. The genes with the highest nucleotide similarity were all the tRNA genes; these genes exhibited similarities of over 90 % and are highlighted in yellow. The nucleotide similarity between the two ribosomal RNA genes (12s rRNA and 16s rRNA) was 87.9% for both; these gene loci are considered to be slowly evolving [51-56] and therefore more likely than other loci to have a high degree of nucleotide similarity. The rRNA genes are highlighted in green as are the cyt *b*, COI, COII, COIII along with a few tRNA genes; all of these loci have nucleotide similarities between 81- 90 %. The six

members of the ND family of genes exhibited similarities ranging from 71-80 % and are coloured pink. The other loci in pink are the tRNA-Glu, the ATP synthase6 and ATP synthase8 loci.

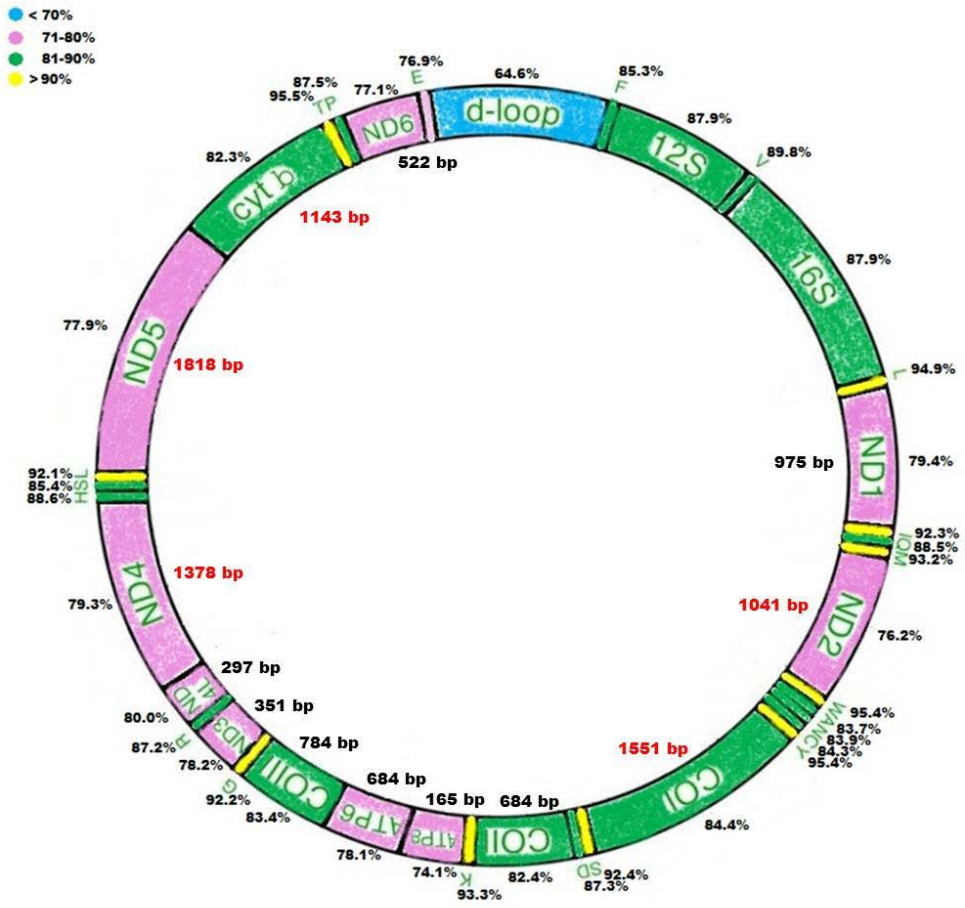


Figure 3.9: The percentage of nucleotide similarity for each of the thirteen avian mitochondrial genes, two rRNA genes and the 22 tRNA gene sequences based on 102 avian species. The similarity is provided next to each locus. The size of each gene is provided. The tRNA loci are coloured in yellow (except the tRNA-Glu is coloured in pink), loci with similarities of between 81 – 90 % are coloured in green, those loci coloured in pink have similarities of between 71 – 80%, and the D-loop is coloured in blue to note the nucleotide similarity of 64.6%.

3.2.2 Phylogenetic reconstruction using complete mitochondrial loci

For further gene sequence alignment it is necessary to have DNA sequences from a wide range of avian species corresponding to all of the major taxonomic groups; these fall within the 102 species selected. The criterion for species testing is that the locus must show sufficient genetic variation between each species and little variation between each member of the same species. In order to determine which gene performs this task the best, complete gene sequences were used for genetic alignment. The COI, cyt *b*, ND2, ND4 and ND5 genes were selected for species identification and phylogenetic tree reconstruction according to their low percentage of homology between different species as shown in the previous result of protein sequence analysis section (Table 3.1). In this part, the ND4 locus was added for this analyse with the others four loci (ND2, ND5, cyt *b*, COI) due to its length and high nucleotide similarity. The length of each locus needs to be of sufficient size to permit species identification; the tRNA genes are predominantly too short to allow this type of identification when comparing closely related species. These loci (COI, cyt *b*, ND2, ND4 and ND5) have nucleotide similarities between 71-90 %. Other mitochondrial loci, including the ND1, ND3, ND4L, ND6, COII, COIII, ATP6 and ATP8, fall into this range of nucleotide similarity (the green and the pink loci highlighted in Figure 3.9). These additional loci were analysed (and these data can be seen in Appendix B). The loci not part of this part of the study were the two ribosomal RNA genes and the tRNA encoding loci. The tRNA genes were excluded as the loci were typically less than 100 bp and unlikely to allow sufficient DNA sequence for species identification.

The two rRNA gene loci were not considered further, as due to their slow rate of evolution it was unlikely that two closely related avian species would have different DNA sequences at these two loci. The complete gene sequences of the COI, cyt *b*, ND2, ND4 and ND5 genes were aligned and the phylogenetic tree of each locus was reconstructed using NJ the method with the Kimura 2-parameter model using MEGA 5 program. The phylogenetic trees are shown in Figures 3.10 - 3.14. In all cases the complete gene sequence is greater than 1,000 bp; 1041 bp for ND2, 1377 bp for ND4, 1818 bp for ND5, 1143 bp for cyt *b*, 1551 bp for COI. These sections of the gene loci are of comparable lengths.

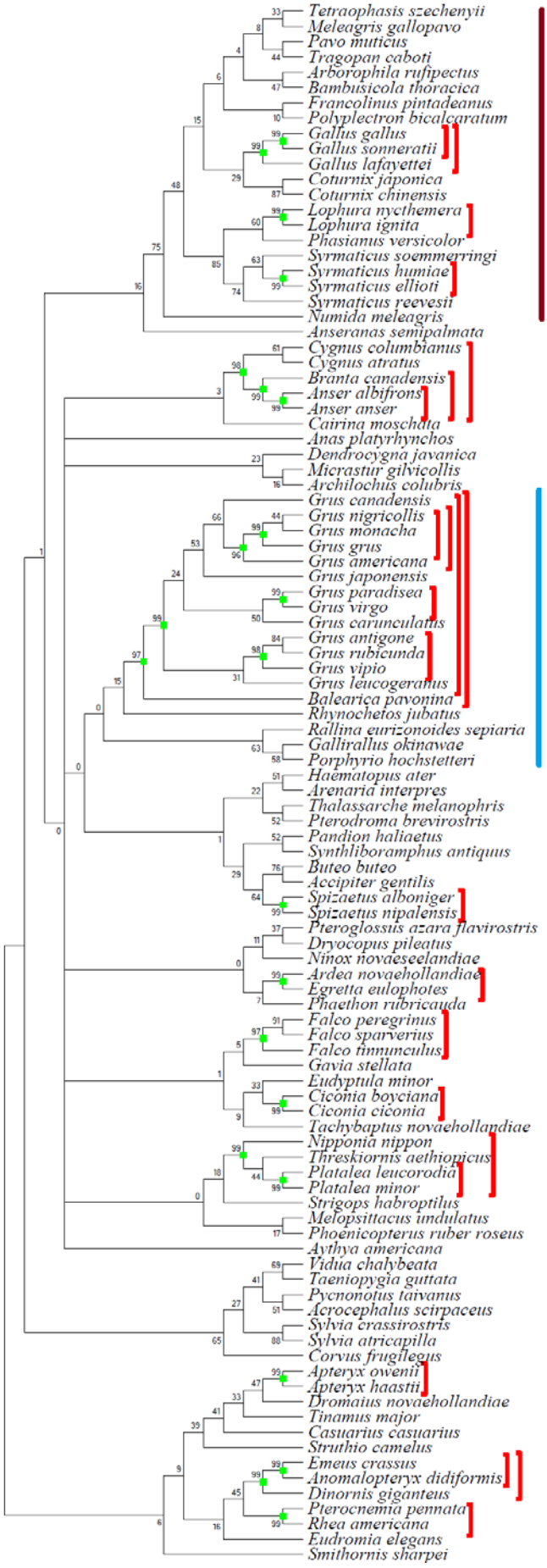


Figure 3.10: Phylogenetic tree reconstruction of the complete gene sequences of the COI gene. Evolutionary analyses were conducted in MEGA5 [43]. The evolutionary history was inferred using the Neighbor-Joining method [44]. The bootstrap consensus tree inferred from 1000 replicates [45] is taken to represent the evolutionary history of the taxa analyzed [45]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [45]. The evolutionary distances were computed using the Kimura 2-parameter method [46] and are in the units of the number of base substitutions per site. The analysis involved 102 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 1238 positions in the final dataset. The Galliformes and Gruiformes are clustered together as expected if the genetic data of the complete gene sequences of the COI gene matches the current taxonomic groups at the taxonomic level of Order. The green dots indicate the bootstrap value for a given interior branch is 95% or higher, then the topology at that branch is considered correct.

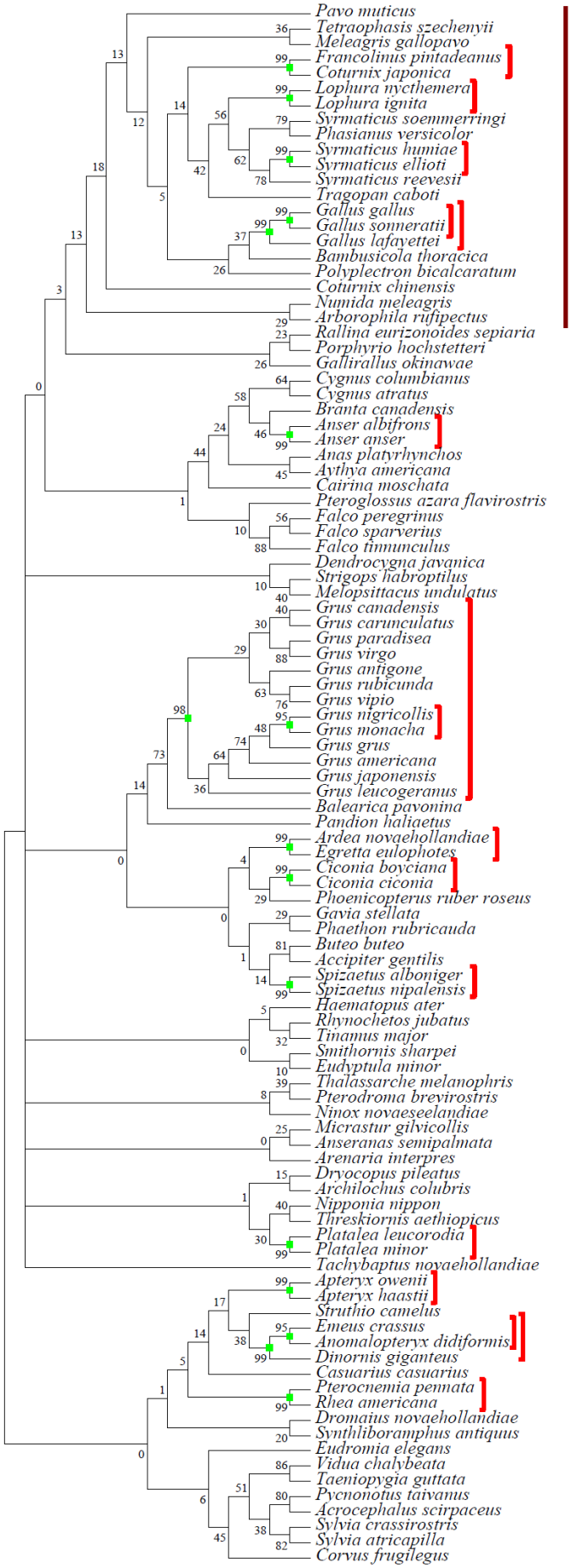


Figure 3.11: Phylogenetic tree reconstruction of the complete gene sequences of the cyt *b* gene. Evolutionary analyses were conducted in MEGA5 [43]. The evolutionary history was inferred using the Neighbor-Joining method [44]. The bootstrap consensus tree inferred from 1000 replicates [45] is taken to represent the evolutionary history of the taxa analyzed [45]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [45]. The evolutionary distances were computed using the Kimura 2-parameter method [46] and are in the units of the number of base substitutions per site. The analysis involved 102 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 770 positions in the final dataset. The Galliformes is clustered together as expected if the genetic data of the complete gene sequences of the cyt *b* gene matches the current taxonomic groups at the taxonomic level of Order. The green dots indicate the bootstrap value for a given interior branch is 95% or higher, then the topology at that branch is considered correct [47].

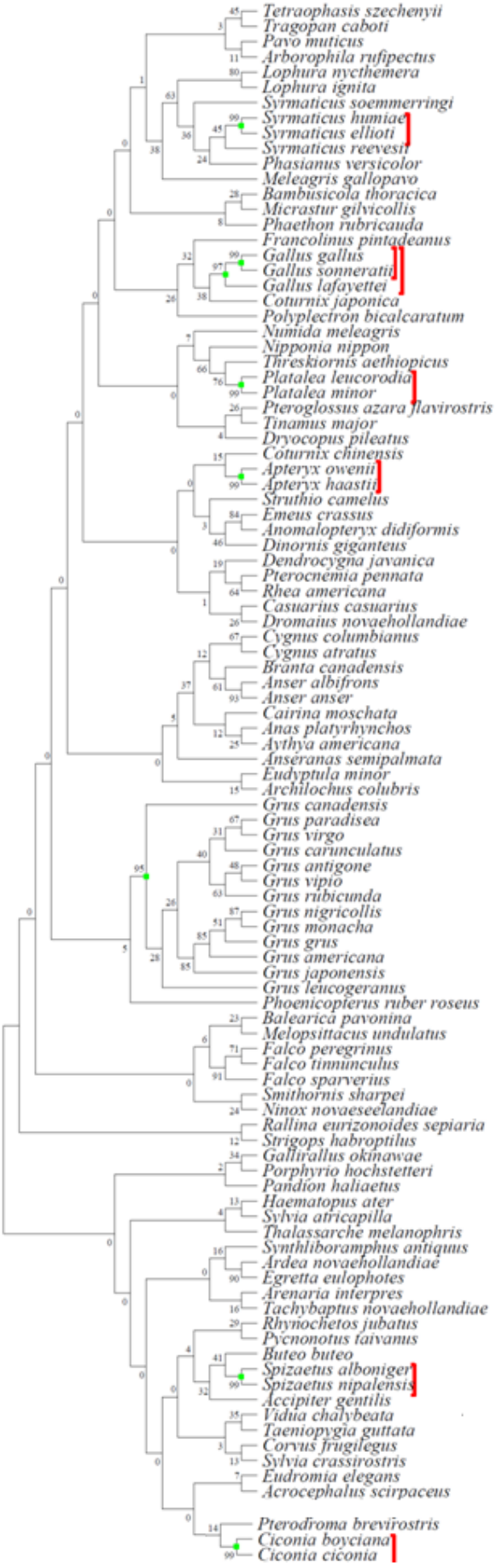


Figure 3.12: Phylogenetic tree reconstruction of the complete gene sequences of the ND2 gene. Evolutionary analyses were conducted in MEGA5 [43]. The evolutionary history was inferred using the Neighbor-Joining method [44]. The bootstrap consensus tree inferred from 1000 replicates [45] is taken to represent the evolutionary history of the taxa analyzed [45]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [45]. The evolutionary distances were computed using the Kimura 2-parameter method [46] and are in the units of the number of base substitutions per site. The analysis involved 102 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 460 positions in the final dataset. All Orders are not clustered together as expected if the genetic data of the complete gene sequences of the ND2 gene. The green dots indicate the bootstrap value for a given interior branch is 95% or higher, then the topology at that branch is considered correct [47].

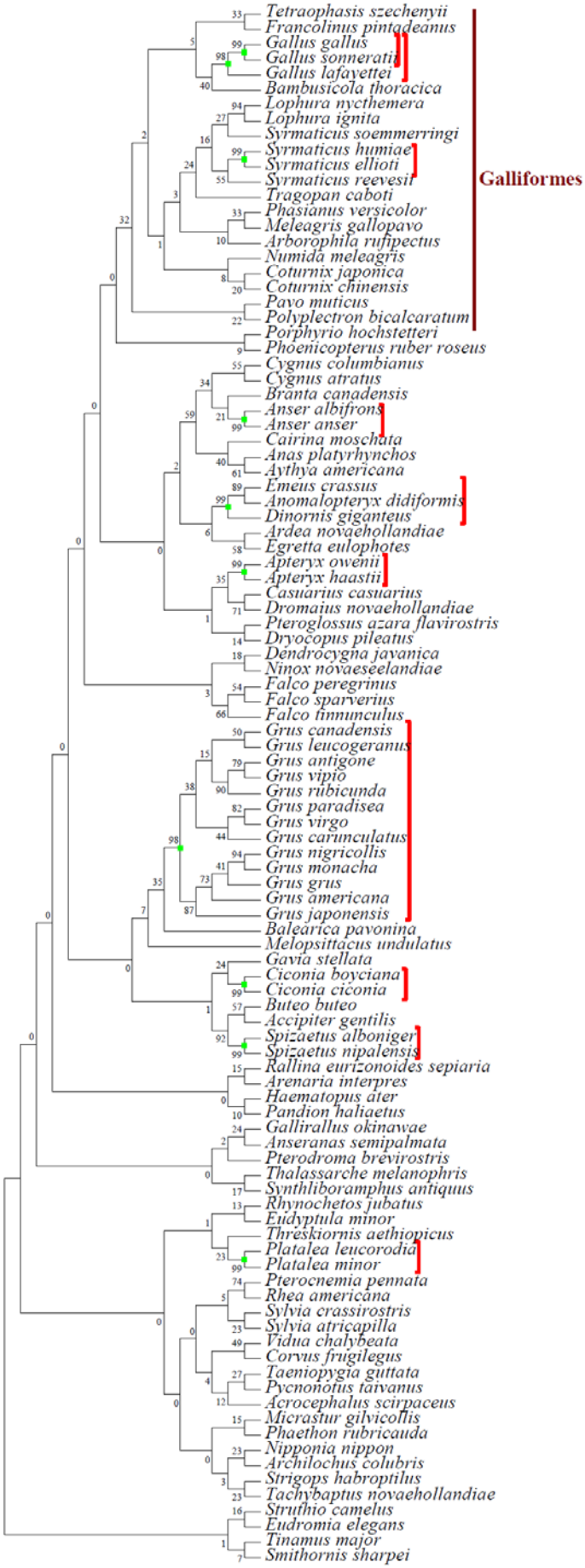


Figure 3.13: Phylogenetic tree reconstruction of the complete gene sequences of the ND4 gene. Evolutionary analyses were conducted in MEGA5 [43]. The evolutionary history was inferred using the Neighbor-Joining method [44]. The bootstrap consensus tree inferred from 1000 replicates [45] is taken to represent the evolutionary history of the taxa analyzed [45]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [45]. The evolutionary distances were computed using the Kimura 2-parameter method [46] and are in the units of the number of base substitutions per site. The analysis involved 102 nucleotide sequences. Codon positions included were 1st, +2nd, +3rd, +Non-coding. All positions containing gaps and missing data were eliminated. There were a total of 652 positions in the final dataset. The Galliformes are clustered together as expected if the genetic data of the complete gene sequences of the ND4 gene matches the current taxonomic groups at the taxonomic level of Order. The green dots indicate the bootstrap value for a given interior branch is 95% or higher, then the topology at that branch is considered correct [47].

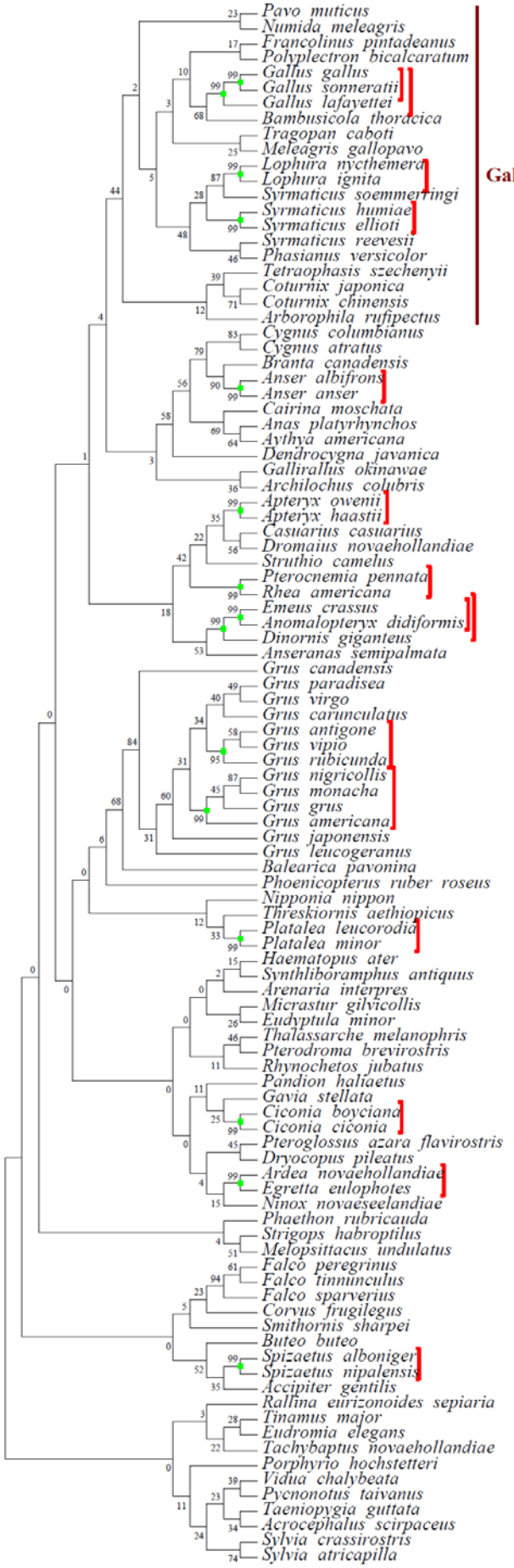


Figure 3.14: Phylogenetic tree reconstruction of the complete gene sequences of the ND5 gene. Evolutionary analyses were conducted in MEGA5 [43]. The evolutionary history was inferred using the Neighbor-Joining method [44]. The bootstrap consensus tree inferred from 1000 replicates [45] is taken to represent the evolutionary history of the taxa analyzed [45]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [45]. The evolutionary distances were computed using the Kimura 2-parameter method [46] and are in the units of the number of base substitutions per site. The analysis involved 102 nucleotide sequences. Codon positions included were 1st, +2nd, +3rd, +Non-coding. All positions containing gaps and missing data were eliminated. There were a total of 908 positions in the final dataset. The Galliformes are clustered together as expected if the genetic data of the complete gene sequences of the ND5 gene matches the current taxonomic groups at the taxonomic level of Order. The green dots indicate the bootstrap value for a given interior branch is 95% or higher, then the topology at that branch is considered correct [47].

The multiple alignments of the complete sequences of the COI gene correctly grouped species within the Orders Galliformes and Gruiformes, while the others three loci (except the ND2) only correctly grouped the Order Galliformes. The following experiment was dividing into 100 bp and 450 bp segments along each gene, as shown in Figure 3.15. While 100 bp is very small, the 450 bp sections were designed to overlap with the previous section by 350 bp to allow for maximum coverage of any conserved and variable regions of these loci.

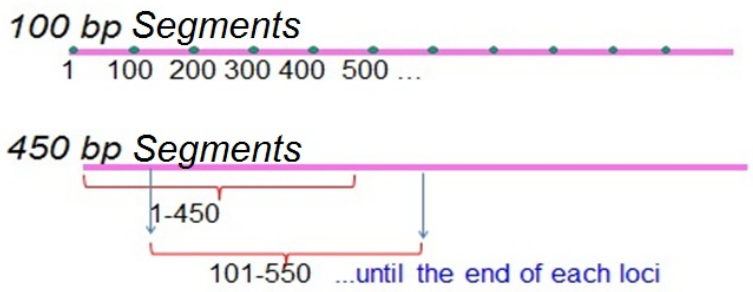


Figure 3.15: The gene loci chosen were divided into 100 bp and 450 bp segments from which the sequences were extracted and used in alignments using the MEGA 5.

All 100 bp and 450 bp segments from the five genes were aligned and phylogenetic trees reconstructed using the MEGA 5. The segments that can identify species and reconstruct a phylogenetic tree with fewest anomalies were parts of the ND2 gene (base positions 1 - 450) and the ND5 gene (base positions 101 - 550). Both segments are from 5' end of the ND2 and the ND5 genes as shown in Figure 3.16a and 3.16b. The sizes of these amplicon are more typical size used in forensic application.

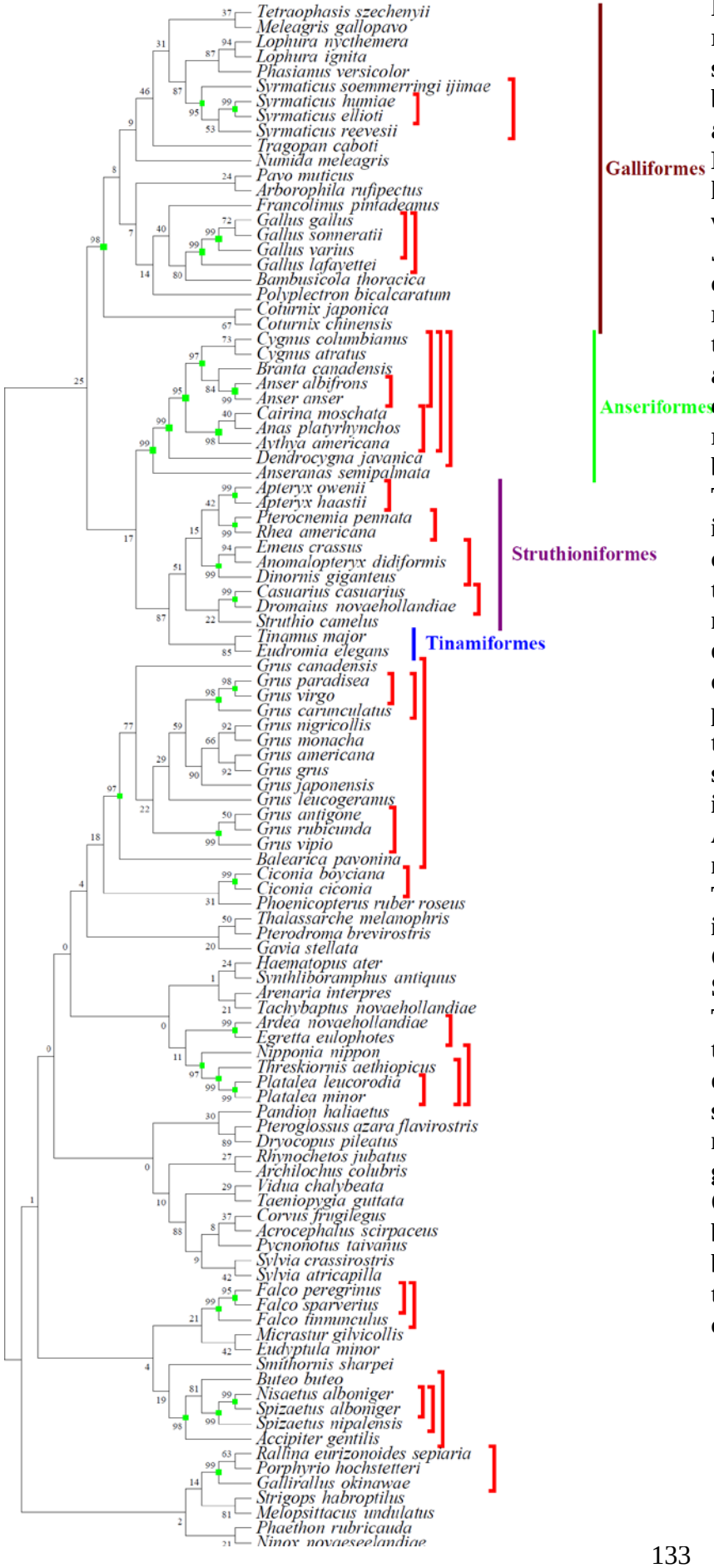


Figure 3.16a: Phylogenetic tree reconstruction of the 450 bp segment from the ND2 gene at base positions 1-450. Evolutionary analyses were conducted in MEGA5 [43]. The evolutionary history of the 1-450 ND2 fragment was inferred using the Neighbor-Joining method . The bootstrap consensus tree inferred from 1000 replicates [45] is taken to represent the evolutionary history of the taxa analyzed [45]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [45]. The evolutionary distances were computed using the Kimura 2-parameter method [46] and are in the units of the number of base substitutions per site. The analysis involved 104 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 450 positions in the final dataset. The Galliformes, Anseriformes, Struthioniformes and Tinamiformes are clustered together as expected if the genetic data of the complete gene sequences of the ND2 gene matches the current taxonomic groups at the taxonomic level of Order. The green dots indicate the bootstrap value for a given interior branch is 95% or higher, then the topology at that branch is considered correct [47].

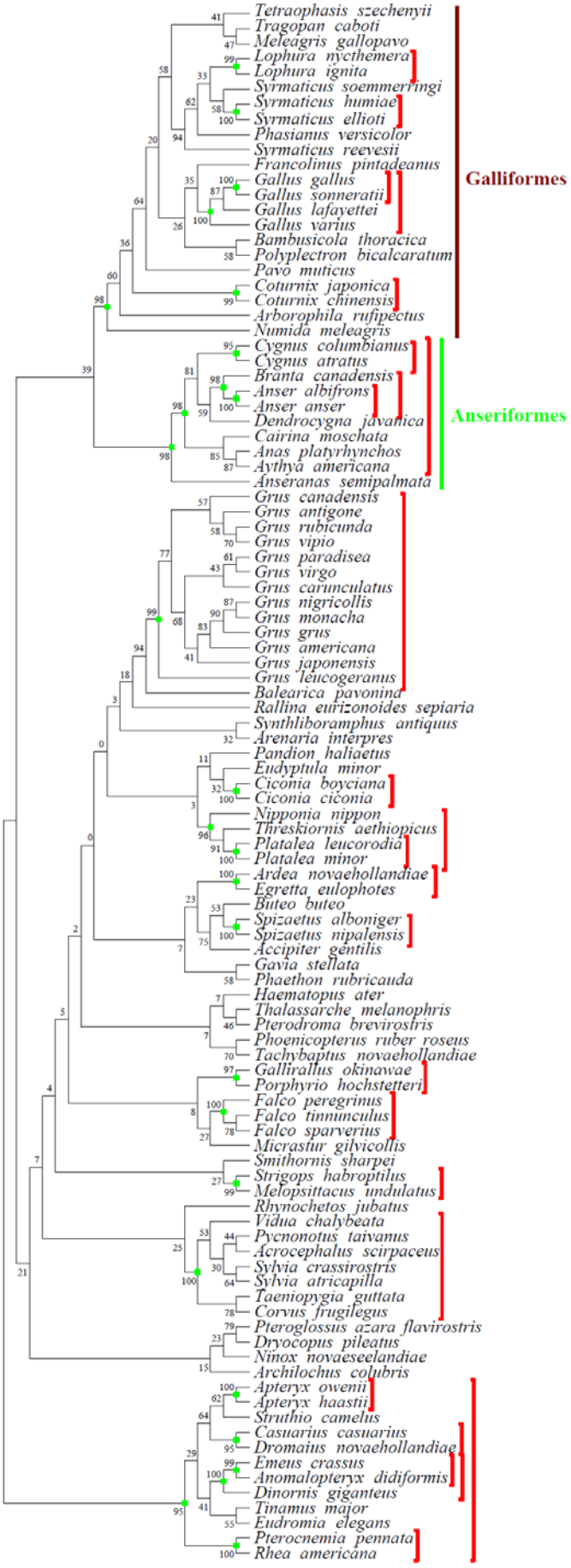


Figure 3.16b: Phylogenetic tree reconstruction of the 450 bp segment from the ND5 gene at base positions 101-550 bp. Evolutionary analyses were conducted in MEGA5 [43]. The evolutionary history of the 101-550 bp ND5 fragment was inferred using the Neighbor-Joining method . The bootstrap consensus tree inferred from 1000 replicates [45] is taken to represent the evolutionary history of the taxa analyzed [45]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [45]. The evolutionary distances were computed using the Kimura 2-parameter method [46] and are in the units of the number of base substitutions per site. The analysis involved 102 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 450 positions in the final dataset. The Galliformes and Anseriformes are clustered together as expected if the genetic data of the complete gene sequences of the ND5 gene matches the current taxonomic groups at the taxonomic level of Order. The green dots indicate the bootstrap value for a given interior branch is 95% or higher, then the topology at that branch is considered correct [47].

Additionally, the 450 bp of the ND5 gene was found to group correctly the 102 avian species into their appropriate taxonomic groups (Figure 3.17). This is unexpected as much emphasis is placed on the use of COI, as used in Barcoding and cyt *b* gene, the locus used most commonly in mammalian taxonomy [36-39]. All phylogenetic trees of the others 100 bp and 450 bp segments from the five loci are shown in Appendix C.

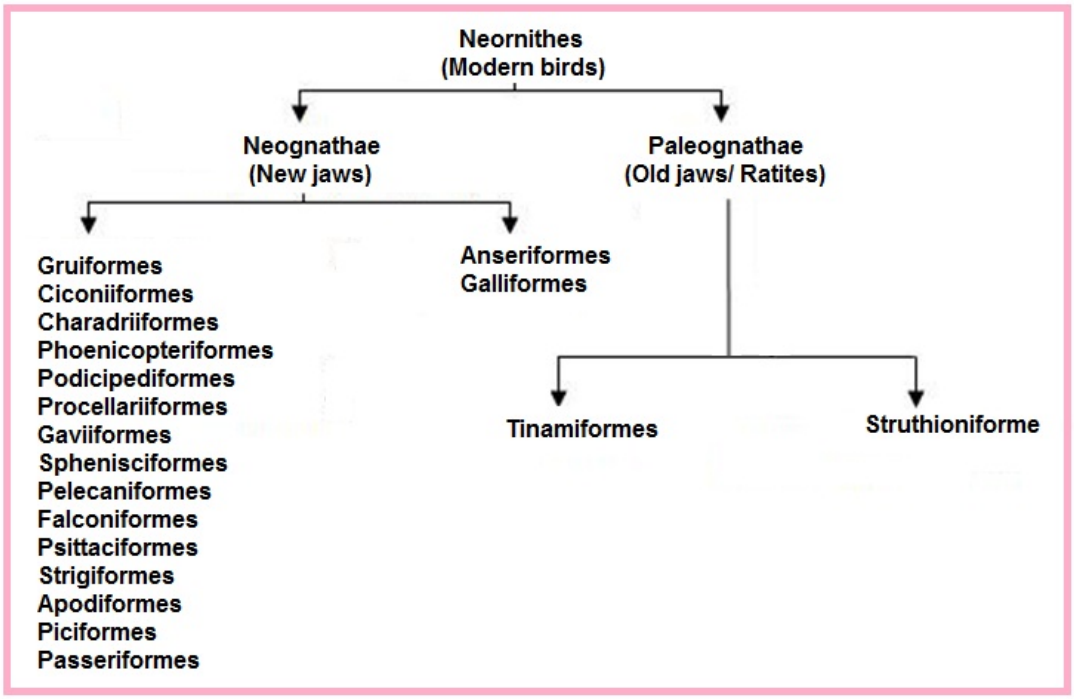


Figure 3.17: Higher avian taxonomy including subclass Neognathae and subclass Paleognathae [57]. All avian species were split into a group that includes all the flightless birds (emus and ostriches etc) and all other birds that evolved flight. The gene sequences of both the ND2 and ND5 loci were able to split the avian species at this higher taxonomic level as well as at lower taxonomic levels. In this study, the 19 Orders were classified as show in this diagram.

3.3 Species identification and phylogenetic tree reconstruction in the members of the same genera

The comparison of avian species using the loci described in the previous sections illustrates the use of these mitochondrial loci in genetic linkage over a wide range of avian species. In order to determine if similar homologies were noted when using members of closely related species, a number of members of the parrot, cockatoo, finch family were tested.

3.3.1 Finches - a potential model group of species

Finches are common birds present over much of the northern hemisphere and are taxonomically of the Order Passeriformes, and Family Fringillidae. These species were chosen originally as bird collectors from around the UK had kindly donated blood samples as part of a related study and permission was granted for their use in this project. They represented many closely related species with a recently divergent genetic history.

The DNA sequences for the five loci selected were not available on GenBank for all the finch samples provided, therefore additional sequencing of the provided samples was necessary to augment the data available. The section sequenced in the first instance represented a section of the ND2 gene. As the R1 and F1 primers have a degenerate sequence, there will be uncertainty at these bases.

This problem can be solved by using F1 and dnR2 which lie outside of the gene and generate the partial sequence of tRNA-Met, the whole ND2 gene and partial tRNA-Trp, the approximately size is 1130 bp.

3.3.1.1 Intra-species variation within finch species

Five different finch species including *Fringilla montifringilla*, *Fringilla coelebs*, *Carduelis chloris*, *Carduelis carduelis* and *Carduelis spinus* were selected for intra-species variation analysis of the ND2 gene. Additional DNA sequences for the cyt *b* and the COI loci were obtained from the GenBank DNA database. The locus ND5 was not used in this analysis (including the inter-species variation of the finch species in the next topic) due to the limited amount of genetic data for this locus from the *Fringilla* and *Carduelis* genera. Additionally there were no ND5 sequence data from the *Fringilla* sp. and there were only two ND5 sequences of the *Carduelis* sp. from *C. spinus* and *C. sinica* available on the database at the time of analysis. The numbers of sequences from each species are shown in Table 3.2

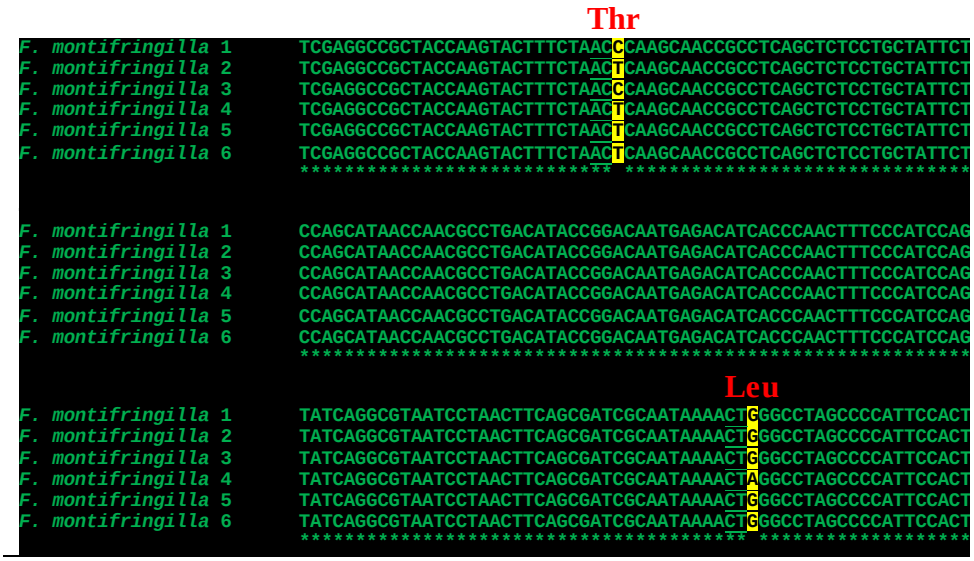
Table 3.2: A list of the number of each finch species used in this study to analyse intra-species variation

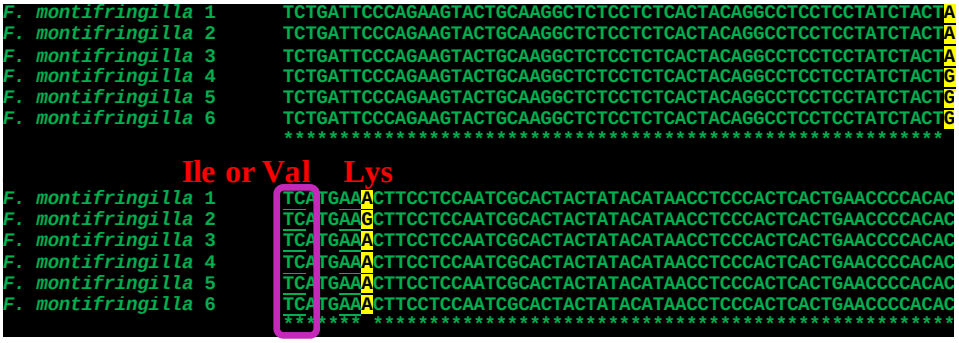
Scientific name	Common name	Number of samples		
		From this study	From the Database	
		ND2	cyt *b*	COI
F. montifringilla	Bramblefinch	6	4	6
F. coelebs	Chaffinch	5	6	6
C. chloris	European greenfinch	5	6	6
C. carduelis	Goldfinch	6	4	6
C. spinus	Eurasian sisken	6	5	6

All the ND2 partial sequences from these finch species and were aligned using the ClustalW program. The variation sites within this locus of each species and the encoded amino acid from those variation sites are shown in Table 3.3.

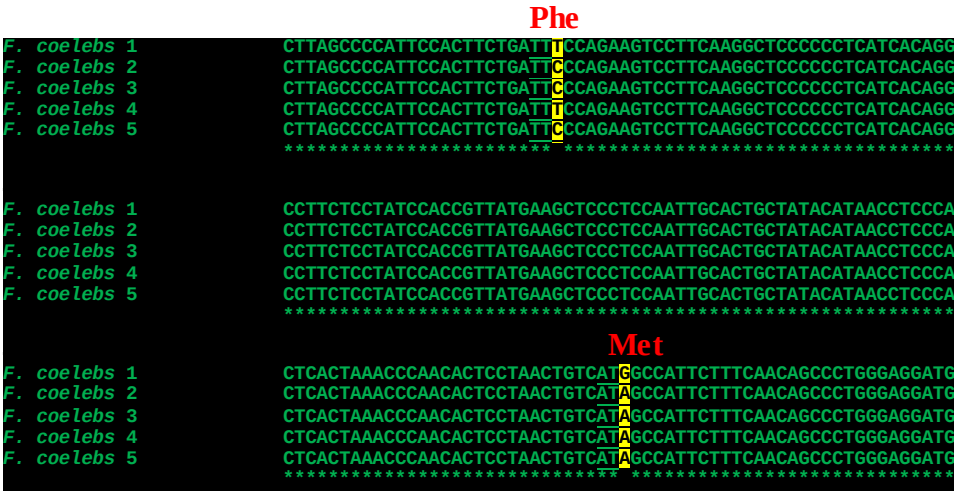
Table 3.3: Multiple alignment result of the ND2 partial sequences from the five finch species. The yellow colour indicates the variable bases within species. The amino acids (using the three letter code) that are encoded from the variable sites are shown in red above each variable site. The numbers in the bracket indicate the base positions on the ND2 gene from each species. The pink boxes indicate the positions where the amino acid has been changed.

1. *F. montifringilla* (181-480)

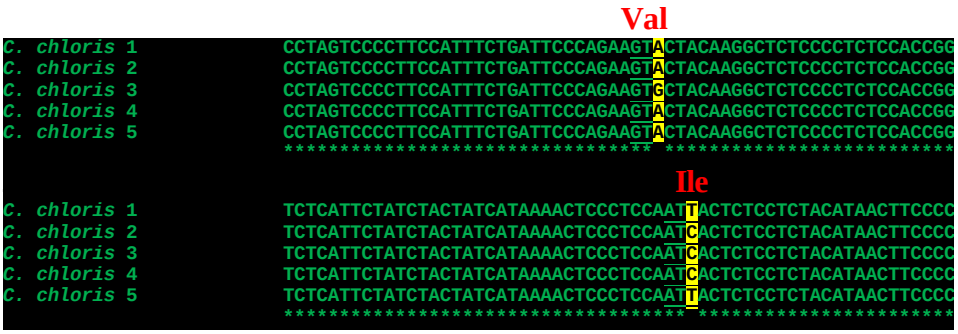




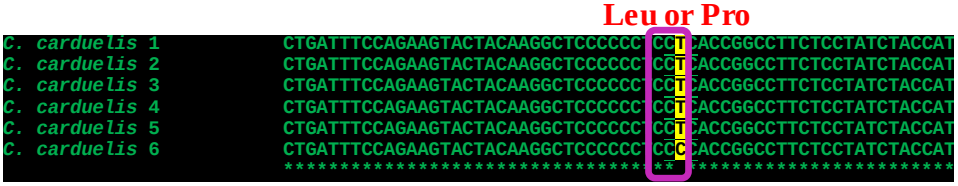
2. *F. coelebs* (321-500)



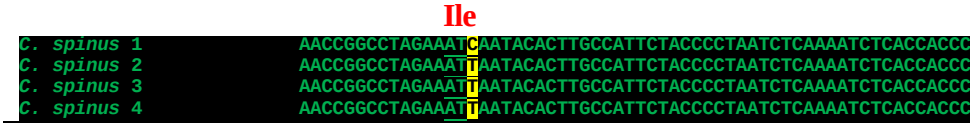
3. *C. chloris* (321-440)



4. *C. carduelis* (321-380)



5. *C. spinus* (91-450)



The base changes within each finch species are predicted not to result in a change of amino acid encoded, therefore the genetic variations within this particular section of the ND2 gene of the five finch species are predominantly synonymous; this is except for two positions which are circled in pink as shown in Table 3.3. These changes occur in the bramblefinch (*F. montifringilla*) at base position 422 and goldfinch (*C. carduelis*) at base position 355, the amino acid can be Ile or Val and Leu or Pro, respectively, which all are in non-polar amino acid group. The mean distances of the DNA sequences from ND2, cyt *b* and COI loci within *F. montifringilla*, *F. coelebs*, *C. chloris*, *C. carduelis* and *C. spinus* were calculated using the MEGA 5 program. The overall mean of a standard comparison of each species was calculated using Kimura 2-parameter model, 1000 bootstrap repetitions. The result and % homology of the sequences within the each species are shown in Table 3.4.

Table 3.4: Percent homology of the partial sequences of the ND2, cyt *b* and COI loci within *F. montifringilla*, *F. coelebs*, *C. chloris*, *C. Carduelis* and *C. spinus* using the MEGA 5 program. The numbers in the bracket indicate the base positions on each gene and the length (bp) of the sequences from each species. The number (n) of sequences used in this study is indicated in red.

Species	% Homology		
	From this study	From the database	
	ND2	cyt *b*	COI
F. montifringilla	n = 6 **99.6 % ± 0.2** (56-515, 459 bp)	n = 4 **98.3 % ± 0.3** (100-740, 641 bp)	n = 6 **99.9 % ± 0.1** (62-708, 647 bp)
F. coelebs	n = 5 **99.8 % ± 0.2** (55-513, 459 bp)	n = 6 **98.9 % ± 0.2** (136-1137, 1002 bp)	n = 6 **99.9 % ± 0.1** (86-708, 617 bp)
C. chloris	n = 5 **99.8 % ± 0.1** (66-540, 475 bp)	n = 6 **99.7% ± 0.1** (136-740, 605 bp)	n = 6 **99.9 % ± 0.1** (61-708, 648 bp)
C. carduelis	n = 6 **99.9 % ± 0.1** (94-508, 415 bp)	n = 4 **97.8% + 0.3** (100-1023, 924 bp)	n = 6 **99.8 % ± 0.1** (92-708, 617 bp)
C. spinus	n = 6 **99.7 % ± 0.1** (61-526, 466 bp)	n = 5 **99.9 % ± 0.1** (100-639, 540 bp)	n = 6 **100% ± 0.0** (73-741, 669 bp)

The cyt *b* partial sequences alignment result shows the highest genetic distance compared to the ND2 and COI loci for all species tested except *C. Spinus*. The COI Barcoding region exhibits the highest conservation for any these loci.

3.3.1.2 Inter-species variation between finch species

The ND2 sequences of the nine finch species from this study and four other closely related species (giving a total of 12 species) from the DNA database were aligned using the MEGA program to compare to the partial sequence of the COI and cyt *b* loci. A list of the species used for inter-species variation is shown in Table 3.5.

Table 3.5: A list of finch species used in this study and the accession number of the sequences obtained from the database.

Scientific name	Common name	Accession number		
		ND2	cyt *b*	COI
F. montifringilla	Bramblefinch	From this study	AF002897	GU571899
F. coelebs	Chaffinch		GU592667	GQ481926
C. chloris	European greenfinch		AF284070	GU571791
C. carduelis	Goldfinch		AF284069	GU571789
C. spinus	Eurasian sisken		DQ792779	GQ481504
P. pyrrhula	Bullfinch		HQ284624	GU572069
C. coccothraustes	Hawfinch		DQ792780	GU571829
C. cannabina	Eurasian linnet		L76298	GU571307
C. flammea cabaret	Lesser redpoll		L76386	DQ433427
C. sinica	Oriental greenfinch	NC_015196	NC_015196	NC_015196
C. flavirostris	Twite	FJ547506	U83199	GU571794
C. pinus	Pine siskin	AF447269	EF530031	DQ434513

The mean distances of DNA sequences from the ND2, cyt *b* and COI loci were calculated using the MEGA 5 program. The overall mean of a standard comparison of loci were calculated using Kimura 2-parameter model, 1000 bootstrap repetitions. The alignment of the partial sequences using the ND2 gene from 12 different finch species (at base positions 34 -756, 723 bp) resulted in a similarity score of 86.4% $\pm$ 0.9. The alignment of a partial sequence of the cyt *b* gene (at base positions 136-740, 605 bp) showed a similarity score of 90.8% $\pm$ 0.8. The COI Barcode locus at base positions 98 -708 (611 bp) resulted in a similarity score of 89.7% $\pm$ 0.7. There were nine different species within the *Carduelis* genus therefore the mean distances of the three loci were calculated additionally. The results are 88.8 % $\pm$ 0.8 % homology for the ND2 gene, 92.8% $\pm$ 0.7% for the cyt *b* gene and 91.4% $\pm$ 0.7% for the COI gene.

In summary, the partial sequence of the ND2 gene at 5' end terminus at base positions 34 -756 being 723 bp exhibited the highest variation between close genetic relatives of different finch species and also less intra-species variation. In contrast, the other two loci (the cyt *b* and the COI loci) exhibited less inter-species variation between the members of the same genus. In addition, the cyt *b* showed lower % homology within the same species compare to the ND2 and the COI loci. The conclusions of intra- and inter-species variation study in finches are shown in Table 3.6.

Table 3.6: The table of conclusion showing percent homogy within the species (intra-species variation) and between species (inter-species variation) of the 12 finches.

Scientific name	% homology within the same species			% homology within the same genus (*Carduelis* sp.)			% homology between species		
	ND2	cyt *b*	COI	ND2	cyt *b*	COI	ND2	cyt *b*	COI
F. montifringilla	**99.6**	**98.3**	**99.9**	**88.8**	**92.8**	**91.4**	**86.4**	**90.8**	**89.7**
F. coelebs	**99.8**	**98.9**	**99.9**						
C. chloris	**99.8**	**99.7**	**99.9**						
C. carduelis	**99.9**	**97.8**	**99.8**						
C. spinus	**99.7**	**99.9**	**100**						
P. pyrrhula									
C. coccothraustes									
C. cannabina									
C. cabaret									
C. sinica									
C. flammea									
C. flavirostris									
C. pinus									

The data in Table 3.6 suggest that the partial sequences at 5' end of the ND2 gene are the most suitable for species identification in finch species and may also be suitable for all avian species identification and phylogenetic tree reconstruction studies.

3.3.1.3 Phylogenetic tree reconstruction for the finches using partial sequence of the ND2, cyt *b* and the COI loci

Phylogenetic tree reconstructions of the partial sequences at 5' end of the ND2 (57-508, 452 bp), cyt *b* (136-639, 504 bp) and the COI (98-708, 611 bp) loci from the 10 finch species were constructed using statistical Neighbor-joining with Kimura 2-parameter model in the MEGA 5 program. The out-group in the Figure 3.18-3.20 is a moa which is confidently excluded from the in-group used in this part of the study.

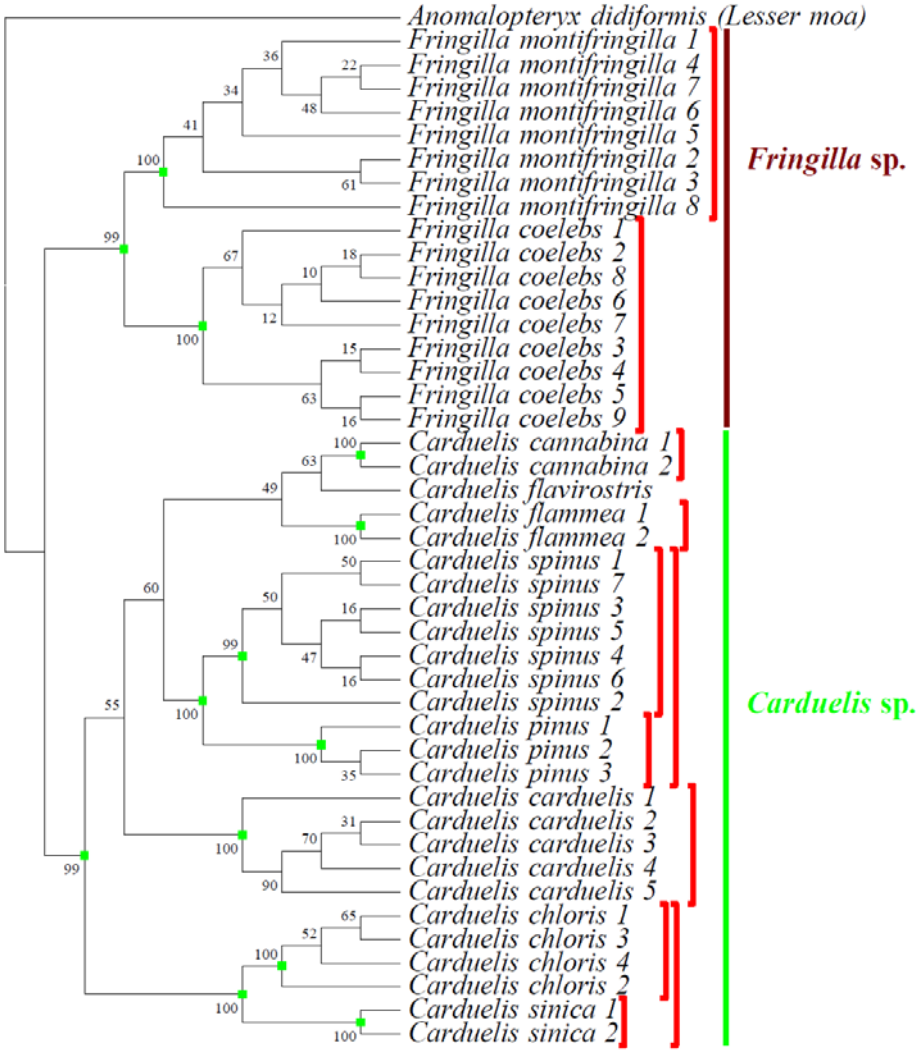


Figure 3.18: Phylogenetic tree reconstruction of the partial sequences at 5' terminus (57-508, 452 bp) of the ND2 gene of 10 different species from *Carduelis* and *Fringilla* genera. The out group was a moa. Evolutionary analyses were conducted in MEGA5 [43].The evolutionary history was inferred using the Neighbor-Joining method [44] . The bootstrap consensus tree inferred from 1000 replicates [45] is taken to represent the evolutionary history of the taxa analyzed [45]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [45]. The evolutionary distances were computed using the Kimura 2-parameter method [46] and are in the units of the number of base substitutions per site. The analysis involved 44 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 452 positions in the final dataset. The two genera are clustered together as expected if the genetic data of the 452 bp at 5' end of the ND2 gene matches the current taxonomic groups at the taxonomic level of genus. The green dots indicate the bootstrap value for a given interior branch is 95% or higher, then the topology at that branch is considered correct [47].

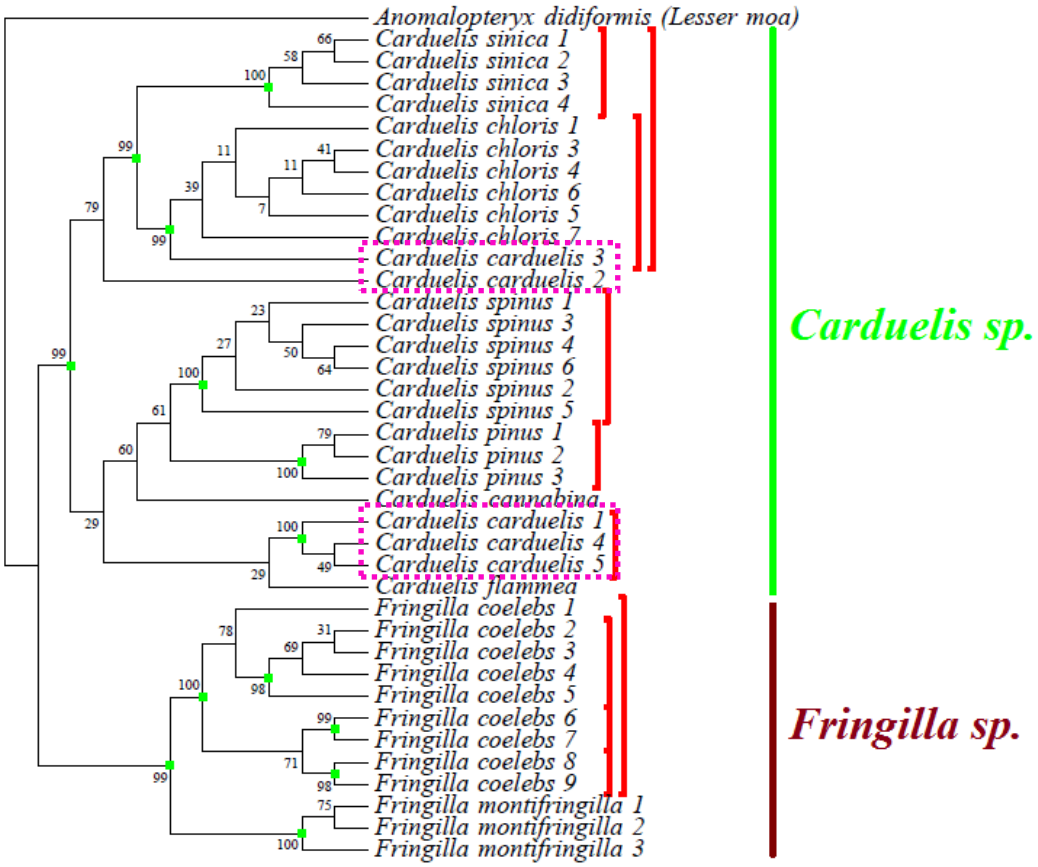


Figure 3.19: Phylogenetic tree reconstruction of the partial sequences at 5' terminus (136-639, 504 bp) of the cyt *b* gene from 10 different species. The out group was a moa. Evolutionary analyses were conducted in MEGA5 [43]. The evolutionary history was inferred using the Neighbor-Joining method [44]. The bootstrap consensus tree inferred from 1000 replicates [45] is taken to represent the evolutionary history of the taxa analyzed [45]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [45]. The evolutionary distances were computed using the Kimura 2-parameter method [46] and are in the units of the number of base substitutions per site. The analysis involved 39 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 503 positions in the final dataset. The two genera are clustered together as expected if the genetic data of the 504 bp at 5' end of the cyt *b* gene matches the current taxonomic groups at the taxonomic level of genus. The green dots indicate the bootstrap value for a given interior branch is 95% or higher, then the topology at that branch is considered correct [47].

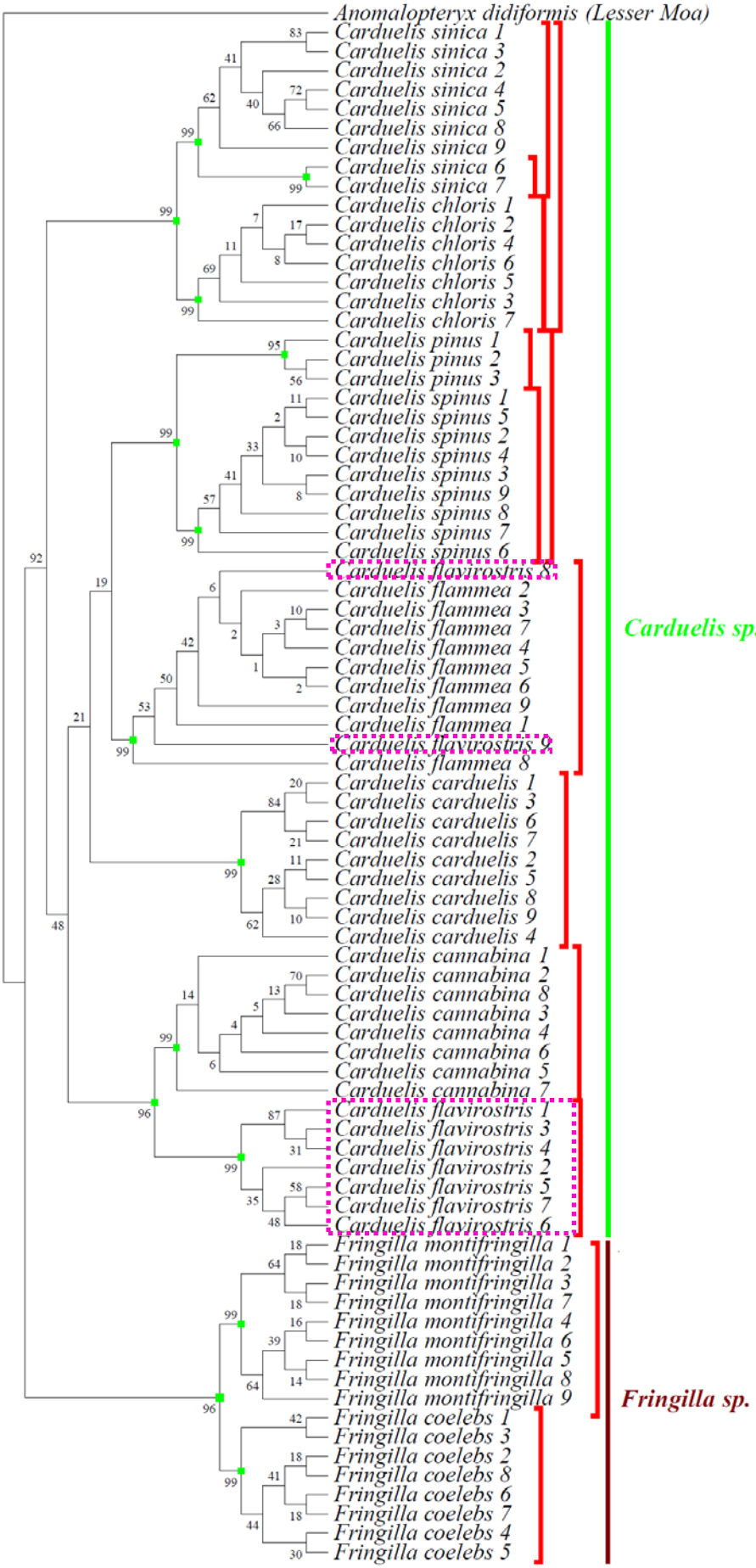


Figure 3.20: Phylogenetic tree reconstruction of the partial sequences at 5' terminus (98-708, 611 bp) of the COI gene from 10 different species. The out group was a moa. Evolutionary analyses were conducted in MEGA5 [43]. The evolutionary history was inferred using the Neighbor-Joining method [44]. The bootstrap consensus tree inferred from 1000 replicates [45] is taken to represent the evolutionary history of the taxa analyzed [45]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [45]. The evolutionary distances were computed using the Kimura 2-parameter method [46] and are in the units of the number of base substitutions per site. The analysis involved 81 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 610 positions in the final dataset. The two genera are clustered together as expected if the genetic data of the 611 bp at 5' end of the COI gene matches the current taxonomic groups at the taxonomic level of genus. The green dots indicate the bootstrap value for a given interior branch is 95% or higher, then the topology at that branch is considered correct [47].

The results in Figure 3.18-3.20 show that the use of the partial sequence from 5' region of the ND2, cyt *b* and COI loci for phylogenetic tree reconstruction can distinguish the different between species and place the members of the finch species into their appropriate taxonomic groups. At the 5' terminus, the cyt *b* patrial sequences incorrectly placed *C. carduelis* and the COI patrial sequences incorrectly placed *C. flavarostris* (indicated by the pink rings in the Figure 3.19 and 3.20, respectively).

In conclusion, the data suggest that these sections of the ND2 (base positions 112-513, 402 bp), cyt *b* (base positions 136-639, 504 bp) and the COI loci (base positions 98-708, 611 bp) can reconstruct the phylogeny and provide a means of species identification for species of members of the same genera. However, the partial sequence of the ND2 at the 5' terminus was superior for both species identification and phylogenetic tree reconstruction in closely related finch species because of the less intra-species variation and the highest inter-species variation compare to the cyt *b* and the COI loci.

3.3.2 Species Testing of parrot and cockatoo Species

Parrot and cockatoo species belong to the order Psittaciformes. The true parrots are members of the Psittacidae family and the cockatoos are members of the Cacatuidae family. This study used 452 bp of the 5' terminus of the ND2 gene (at base positions 58-509) and ND5 gene (at base positions 101-552) to determine if they can be used in the identification of members of the Psittacidae and Cacatuidae families.

3.3.2.1 Intra-species variation within parrot and cockatoo species

A list of the species used in this study is shown in Table 3.7. DNA sequences from these species were aligned using the ClustalW program. The bases that showed variation within these loci for each of the species used, and the amino acid encoded from those variation sites, are shown in Tables 3.8 and 3.9.

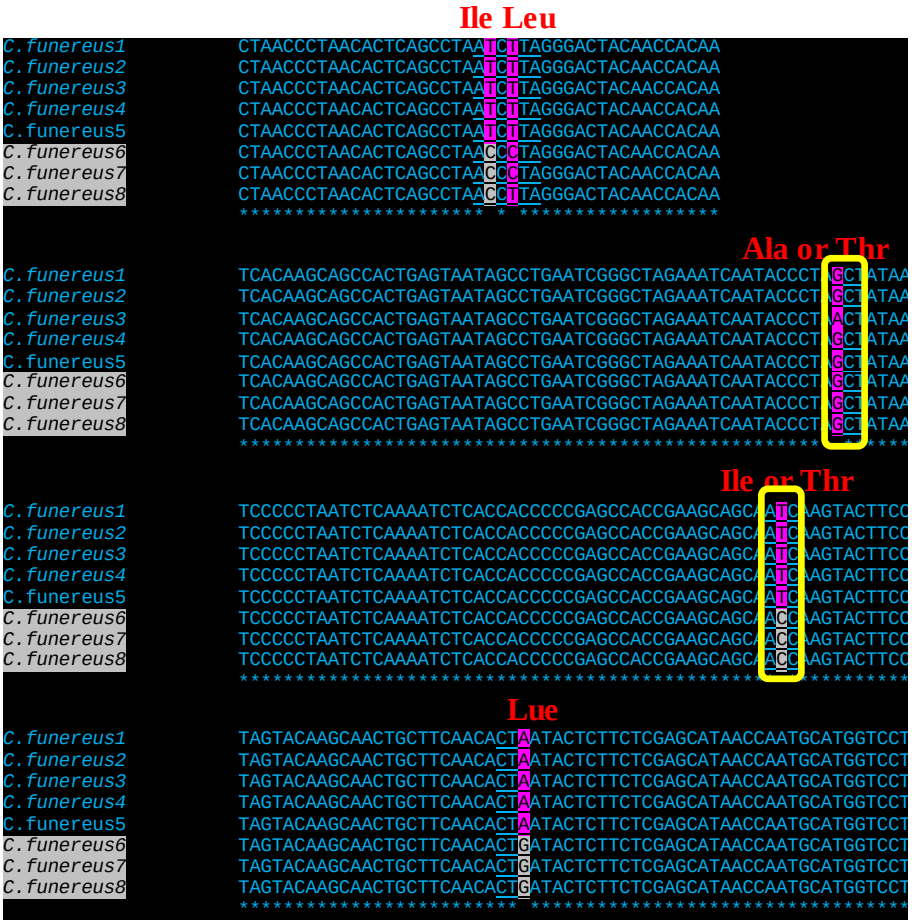
Table 3.7: A list of parrot and cockatoo species that formed the examination of a study of intra-species variation.

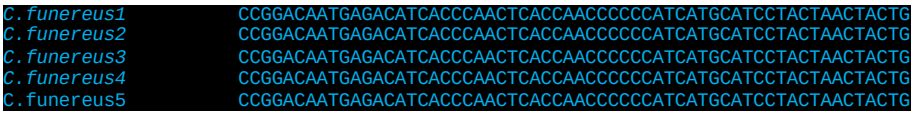
Species	Sub-species	Common name	Number of sequences			
			From this study		From the database	
			ND2	ND5	cyt *b*	COI
Calyptorhynchus banksii	*samueli-1* *samueli-2*	Red-tailed black cockatoo	6	6	2	2
	naso-1 *naso-2*					
	macrorhynchus-1 *macrorhynchus-2*					
Calyptorhynchus funereus	*latirostris-1* *latirostris-2* *latirostris-3*	White-tailed black cockatoo	7	7	2	2
	Funereus-1 *Funereus-2* *Funereus-3* *Funereus-4* *Funereus-5*	Yellow-tailed black cockatoo				
Psittacula alexandri	-	Moustached parakeet	5	6	5	Not on the database
Amazona ochrocephala	-	Yellow-crowned Amazon parrot	6	4	6	6

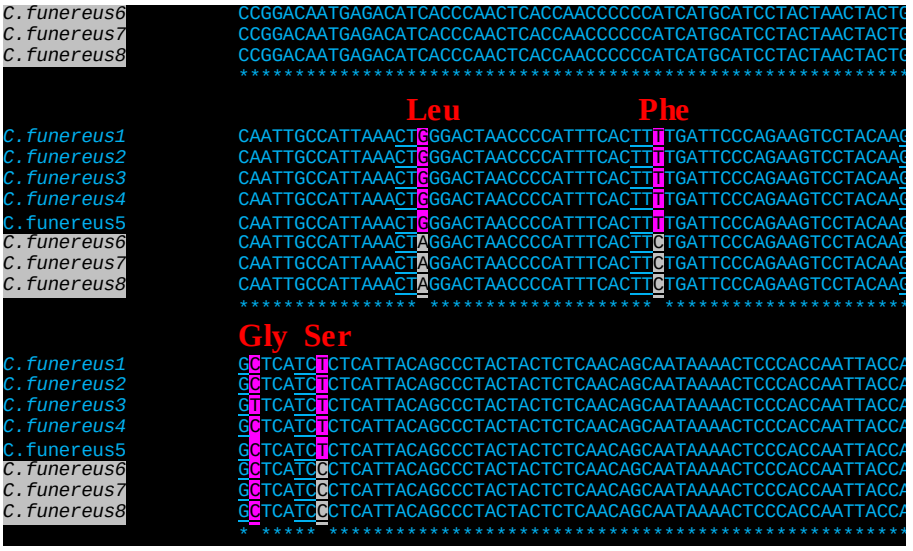
Table 3.8: Multiple alignment result of the ND2 partial sequences from the two parrots; *P. alexandri* and *A. ochrocephala* (data not shown, there is no genetic variation within *A. ochrocephala* species) and the two cockatoo species; *C. banksii* and *C. funereus*. The pink colour indicates the variable sites within species. The pink colour indicates the variable sites within species. The amino acids that are encoded by the variable sites are shown in red letter above each variable site. For the two sub-species of the *C. funereus*, the genetic variation sites of the *C. funereus latirostris* and this sub-species are highlighted in grey. The numbers in the bracket indicate the base positions on the ND2 gene from each species. The yellow boxes indicate the positions where the amino acid has been changed.

1. *C. banksii* (188-247)

2. *C.funereus* (20-421)







3. *P.alexandri* (298-480)

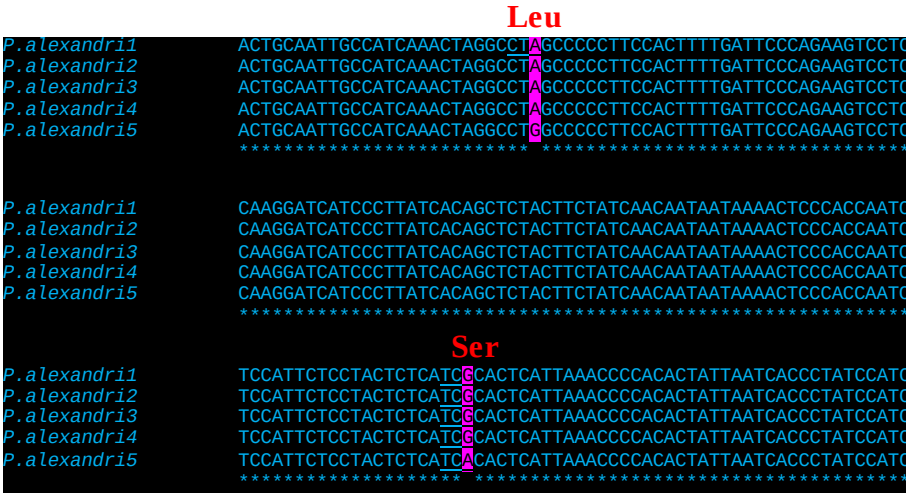
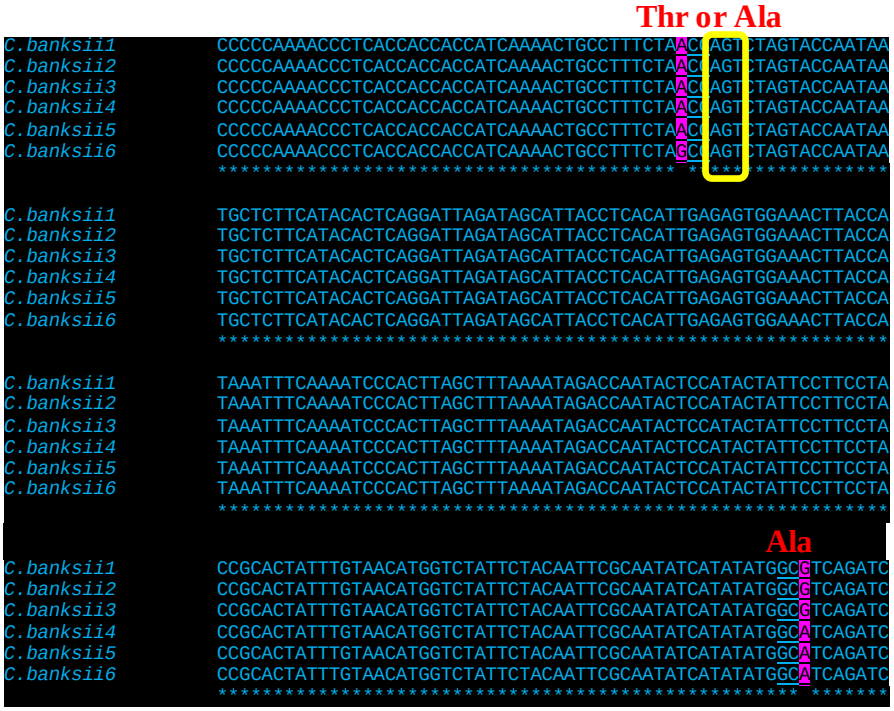
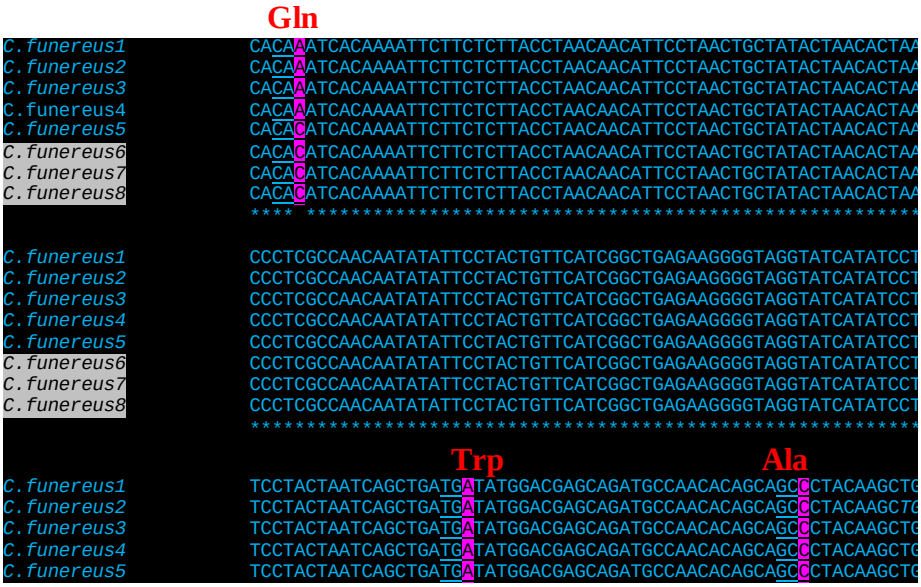


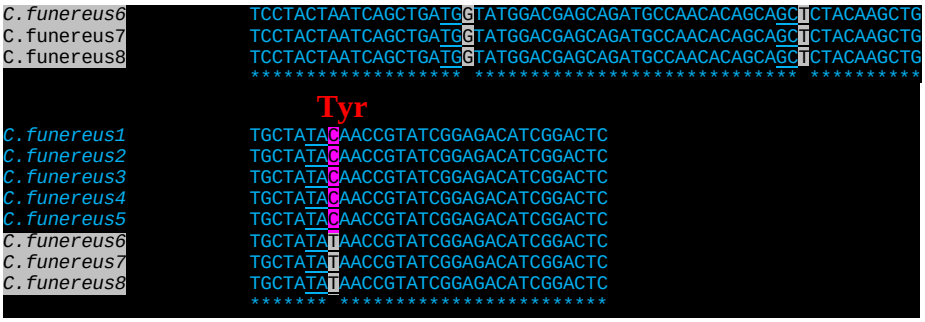
Table 3.9: Multiple alignment result of the ND5 partial sequences from the two parrots; *P. alexandri* and *A. ochrocephala* and the two cockatoos; *C. banksii* and *C. funereus*. The pink colour indicates the variable sites within species. The amino acids that are encoded by the variable sites are shown in red letter above each variable site. For the two sub-species of the *C. funereus*, the genetic variation sites of the *C. funereus latirostris* and this sub-species are highlighted in grey. The numbers in the bracket indicate the base positions on the ND5 gene from each species. The yellow boxes indicate the positions where the amino acid has been changed.

1. *C. banksii* (102-340)

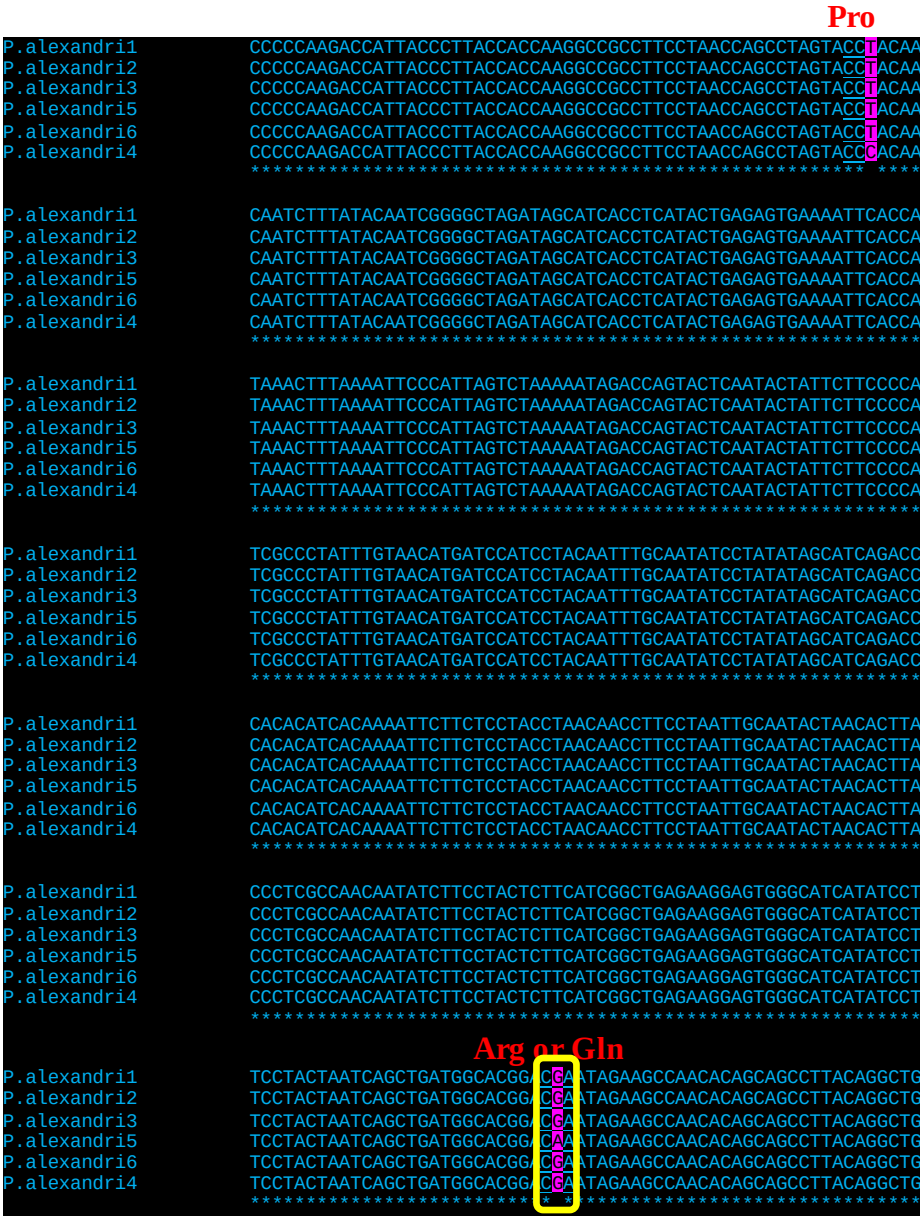


2. *C. funereus* (341-552)





3. *P. alexandri* (101-520)



4. *A. ochrocephala* (161-220)

The base changes within the ND2 and ND5 loci in the parrot and cockatoo species used in this study are mostly silent mutations. Only a few amino acids which are indicated in the yellow boxes, as shown in Table 3.8 and 3.9, are the result of changes generating a different amino acid within the same species. The side-chains of the amino acids that have changed are in the same group such as in the ND2 gene of *C. funereus* at base position 117 where alanine has changed to threonine. Another example is at base position 171 where isoleucine has changed to threonine, as shown in Table 3.8, and the side-chain of these amino acids are in hydrophobic neutral group. In the ND5 gene of *C. funereus* at base position 144 threonine has changed to alanine (hydrophobic neutral group) and at base position 488 of *P. alexandri* arginine has changed to glutamine, both of which are in hydrophillic neutral group. Although the side-chain of threonine and alanine are in the same group (hydrophobic neutral), threonine is a polar amino acid while alanine is a non-polar amino acid. An example where there is a change in the group of amino acids occurred between the two sub-species, *C. funereus latirostris* and *C. funereus funereus*, where the amino acid change within the ND2 locus at base position 117 from non-polar (Ala) in *C. f. funereus* to polar (Thr) in the white tailed phenotype, as shown in Table 3.8. Also in the ND5 locus at base position 144 of sample number 6 of *C. banksii* the amino acid changes from polar (Thr) to non-polar (Ala) (Table 3.9). These amino acid changes may affect the protein folding as the polar side-chains will fold outside the protein structure to interact with water while the nonpolar side-chains are normally folded inside of the protein structure. If these changes occur near an active site of the protein may also affect to the function of the protein.

The mean distance of the ND2, ND5, cyt *b* and COI loci within the parrot and cockatoo species was calculated using the MEGA 5 program. An overall mean of a standard comparison of each species was calculated using Kimura 2-parameter model using 1000 bootstrap repetitions. The percentage of homology of the ND2, ND5, cyt *b* and COI loci within the each species are shown in Table 3.10.

Table 3.10: Percentage of homology (%) within species of the two cockatoos: *C. banksii* and *C. funereus* and the two parrots: *P. alexandri* and *A. ochrocephala* using the MEGA 5 program. The partial sequences of the ND2 and ND5 genes are obtained from this study and the partial sequence of the cyt *b* gene and the COI DNA Barcoding region are obtained from the database. The numbers in the bracket indicate the base positions on each gene and the length (bp) of the sequences from each species. The number of sequences used in this study indicates in red.

Species	% Homology			
	From this study		From the database	
	ND2 (58-509, 452 bp)	ND5 (58-509, 452 bp)	cyt *b*	COI
C. banksii	six sequences **99.8 % ± 0.1**	six sequences **99.7 % + 0.2**	two sequences **99.8 % + 0.2** (413-860, 448 bp)	two sequences **100 % ± 0.0** (39-759, 721 bp)
C. funereus	seven sequences **99.4 % ± 0.3**	seven sequences **99.6 % + 0.2**	two sequences **100 % + 0.0** (413-860, 448 bp)	two sequences **99.9 % ± 0.1** (39-759, 721 bp)
P. alexandri	five sequences **99.8 % ± 0.1**	six sequences **99.9 % + 0.1**	five sequences **94.1% ± 0.6** (203-907, 705 bp)	**Not on the database**
A. ochrocephala	six sequences **100 % ± 0.0**	four sequences **99.9 % + 0.1**	six sequences **98.4% + 0.3** (100-793, 694 bp)	six sequences **98.4 % ± 0.4** (750-1257,508 bp)*

* At 3' terminus of the COI gene

The 452 bp section at 5' terminus of the ND2 locus between base positions 58-509, the 452 bp section within ND5 locus at base positions 101-552 and the 721 bp section within the COI loci at base positions 39-759, appear to contain less genetic variation within these avian species than the cyt *b* locus; with the only exception being within *C. funereus*.

The highest intra-species variation within the five sequences of *P. alexandri* is in the cyt *b* locus at base position 203-907; being 704 bp in length. The 508 bp section of the COI gene at base position 750-1257 (at 3' terminus) of *A. ochrocephala* shows the same percentage of homology to the 5' region of this gene with a similarity of 98.4% as shown in Table 3.10 . The 3' region (508bp) of the COI locus of *A. ochrocephala* used in this study is about 213 bp smaller than the 5' region (721 bp). This suggests that more variation sites occur at the 3' region than at the 5' region of this gene.

In summary, the segments that are ideal for species testing are the 452 bp section at 5' terminus of the ND2 gene between base positions 58-509 and ND5 at base positions 101-552 due to their low genetic distance within the same species (intra-species variation). The 721 bp section within the COI loci at base positions 39-759 shows the lowest intra-species variation but this highly conserve gene might not be able to classify closely related species. In contrast, for avian species identification and phylogenetic tree reconstruction purpose, a sequence of 704 bp within the cyt *b* locus, at base position 203-907, shows the greatest intra-species variation might places the same species into different taxonomic group.

3.3.2.2 Inter-species variation between parrots and cockatoos

Sections of the ND2 and ND5 at 5' terminus of each gene for 15 species (listed in Table 3.11) were aligned using MEGA 5 program. These data were compared to partial sequences of the cyt *b* gene and the COI Barcoding region obtained from the GenBank DNA database.

Table 3.11: A list of parrot and cockatoo species used in analysing inter-species variation as part of this study

Scientific name	Common name	Accession number				
		ND2	ND5	cyt *b*	COI At 3' terminus region	COI At 5' terminus region
Aprosmictus erythropterus	Red-winged Parrot	From this study	From this study	AB177959	EU621596	Not on the database
Alisterus amboinensis	Australian King Parrot			Not on the database	EU621594	Not on the database
Psephotus dissimilis	Hooded Parrot			Not on the database	HQ316868	Not on the database
Trichoglossus haematodus	Rainbow Lorikeet			AB177942	EU621667	Not on the database
Platycercus elegans	*Adelaide* Rosella			DQ467900	HQ316866	Not on the database
Polytelis anthopeplus	Regent Parrot			AF031918	HQ316867	Not on the database
Polytelis alexandrae	Princess Parrot			Not on the database	EU621649	Not on the database
Psittacula alexandri	Moustached parakeet			AB177970	Not on the database	Not on the database
Amazona ochrocephala	Yellow crowned Amazon			AY283468	AY301453	Not on the database
Callocephalon fimbriatum	Gang-gang Cockatoo			JF414312	Not on the database	JF414284
Eolophus roseicapillus	Galah Cockatoo			FJ498976	Not on the database	JF414294
Calyptorhynchus banksii	Red-tailed Black Cockatoo			JF414309	Not on the database	JF414281
Calyptorhynchus lathami	Glossy Black Cockatoo			JF414241	JF414241	JF414241
Calyptorhynchus funereus	Yellow-tailed Black Cockatoo			JF414307	Not on the database	JF414279
Calyptorhynchus latirostris	Short-billed Black Cockatoo			JF414243	JF414243	JF414243

The mean distances between 15 parrot and cockatoo species of the ND2, ND5, cyt *b*, COI loci were calculated using the MEGA 5 program. Overall mean of a standard comparison of loci were calculated using the Kimura 2-parameter model and 1000 bootstrap repetitions. The partial sequences alignment of the ND2 gene (at base positions 58-509, 452 bp) showed a similarity score of 83.0 % $\pm$ 1.3. The partial sequence alignment of the ND5 gene (at base positions 101-552, 452 bp) resulted in a similarity score of 81.8 % $\pm$ 1.4. The partial sequence alignment of the cyt *b* gene (at base positions 686-860, 175 bp) resulted in a similarity score of 86.3 % $\pm$ 1.9. The 5' end of the COI at base positions 39-759 (721 bp) showed a similarity of 87.7 % $\pm$ 1.0 and the 3' end of the COI at base positions 741-1272 (532 bp) resulted in a similarity score of 87.0 % $\pm$ 1.1.

In summary, the partial sequence of the ND2 and ND5 loci at their 5' termini exhibited highest variation between parrot and cockatoo species examined in this study but less intra-species variation. The 5' end of the COI loci was found to be more conserved than the 3' region; with the homology of the 5' region of the COI gene being only 0.7% higher than the segment used at the 3'end, although the length used for the analysis was 190 bp longer than the 3' region. It is noteworthy that the cyt *b* partial sequences used in inter-species variation analysis are only 175 bp but the result showed a homology score less than many longer sections of DNA used such as the 5' and 3' segments from the COI gene.

These data suggests that the partial sequences at 5' end of the ND2 and ND5 loci are the most suitable for species identification in parrot and cockatoo species and may also be suitable for other avian species for the purpose both of species identification and phylogenetic tree reconstructions.

3.3.2.3 Phylogenetic tree reconstruction for a range of parrot and cockatoo species using partial sequence of the ND2, ND5, cyt *b* and the COI loci

Phylogenetic tree reconstructions using partial sequences at the 5' end of the ND2 (58-509, 452 bp) , ND5 (101-552, 452 bp), cyt *b* (686-860, 175 bp), the 5' region of the COI (39-760, 722bp) and the 3' region of the COI gene (741-1272, 532 bp) were constructed using statistical Neighbor-joining with Kimura 2-parameter model in the MEGA 5 program; the species used were either parrot or cockatoo species. The trees are shown in Figures 3.21-3.25.

Figure 3.21: Phylogenetic tree reconstruction of the partial sequences at 5' terminus (58-509, 452 bp) of the ND2 gene of 14 different species from Psittacidae and Cacatuidae families. Evolutionary analyses were conducted in MEGA5 [43]. The evolutionary history was inferred using the Neighbor-Joining method [44]. The bootstrap consensus tree inferred from 1000 replicates [45] is taken to represent the evolutionary history of the taxa analyzed [45]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [45]. The evolutionary distances were computed using the Kimura 2-parameter method [46] and are in the units of the number of base substitutions per site. The analysis involved 15 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 449 positions in the final dataset. The two Families are clustered together as expected if the genetic data of the 452 bp at 5' end of the ND2 gene matches the current taxonomic groups at the taxonomic level of family. The green dots indicate the bootstrap value for a given interior branch is 95% or higher, then the topology at that branch is considered correct [47].

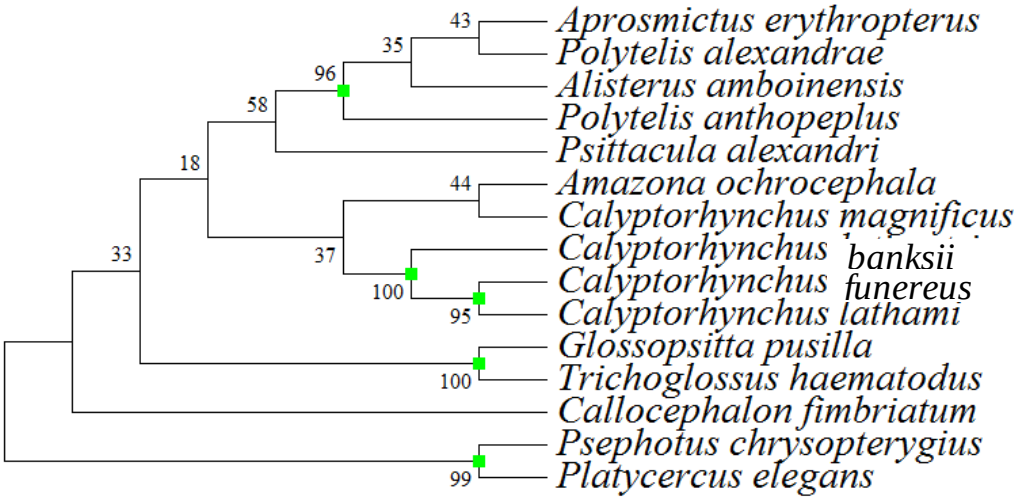


Figure 3.22: Phylogenetic tree reconstruction of the partial sequences at 5' terminus (101-552, 452 bp) of the ND5 gene of 14 different species from Psittacidae and Cacatuidae families. Evolutionary analyses were conducted in MEGA5 [43]. The evolutionary history was inferred using the Neighbor-Joining method [44]. The bootstrap consensus tree inferred from 1000 replicates [45] is taken to represent the evolutionary history of the taxa analyzed [45]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [45]. The evolutionary distances were computed using the Kimura 2-parameter method [46] and are in the units of the number of base substitutions per site. The analysis involved 15 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 448 positions in the final dataset. The two Families are not clustered together as expected if the genetic data of the 452 bp at 5' end of the ND5 gene matches the current taxonomic groups at the taxonomic level of family. The green dots indicate the bootstrap value for a given interior branch is 95% or higher, then the topology at that branch is considered correct [47].

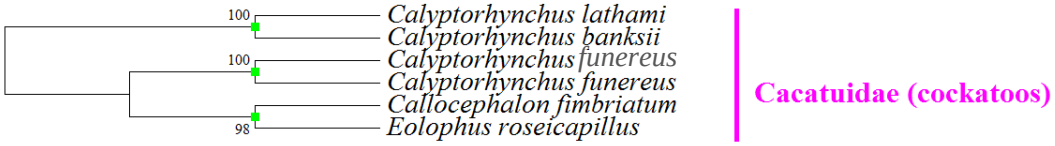


Figure 3.23: Phylogenetic tree reconstruction of the partial sequences at 5' terminus (39-760, 722bp) of the COI gene of 5 different species. According to there is no sequences at 5' end of this gene from Psittacidae family submitted on the database so that the tree was reconstruct from Cacatuidae family. Evolutionary analyses were conducted in MEGA5 [43]. The evolutionary history was inferred using the Neighbor-Joining method [44]. The bootstrap consensus tree inferred from 1000 replicates [45] is taken to represent the evolutionary history of the taxa analyzed [45]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [45]. The evolutionary distances were computed using the Kimura 2-parameter method [46] and are in the units of the number of base substitutions per site. The analysis involved 6 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 722 positions in the final dataset. The green dots indicate the bootstrap value for a given interior branch is 95% or higher, then the topology at that branch is considered correct [47].

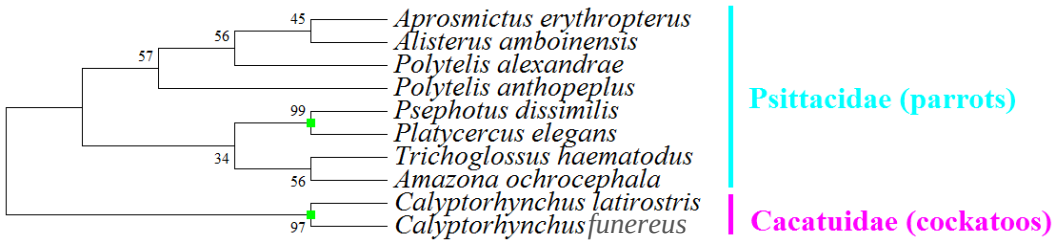


Figure 3.24: Phylogenetic tree reconstruction of the partial sequences at 3' terminus (741-1272, 532 bp) of the COI gene of 10 different species from Psittacidae and Cacatuidae families. Evolutionary analyses were conducted in MEGA5 [43]. The evolutionary history was inferred using the Neighbor-Joining method [44]. The bootstrap consensus tree inferred from 1000 replicates [45] is taken to represent the evolutionary history of the taxa analyzed [45]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [45]. The evolutionary distances were computed using the Kimura 2-parameter method [46] and are in the units of the number of base substitutions per site. The analysis involved 10 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 523 positions in the final dataset. The two Families are clustered together as expected if the genetic data of the 532 bp at 5' end of the COI gene matches the current taxonomic groups at the taxonomic level of family. The green dots indicate the bootstrap value for a given interior branch is 95% or higher, then the topology at that branch is considered correct [47].

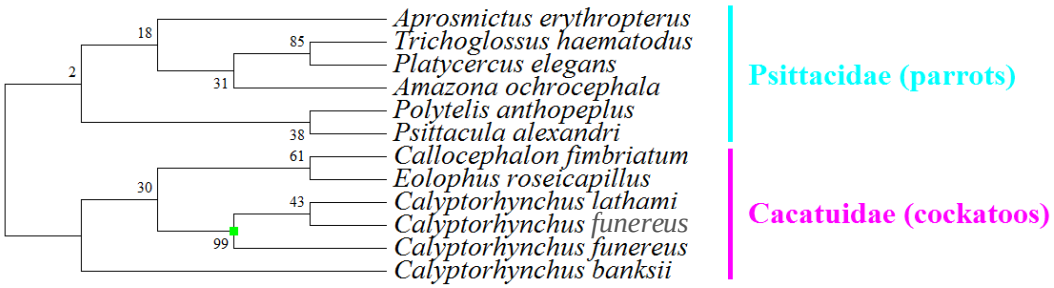


Figure 3.25: Phylogenetic tree reconstruction of the partial sequences at 3' terminus (686-860, 175 bp) of the cyt *b* gene of 11 different species from Psittacidae and Cacatuidae families. Evolutionary analyses were conducted in MEGA5 [43]. The evolutionary history was inferred using the Neighbor-Joining method [44]. The bootstrap consensus tree inferred from 1000 replicates [45] is taken to represent the evolutionary history of the taxa analyzed [45]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [45]. The evolutionary distances were computed using the Kimura 2-parameter method [46] and are in the units of the number of base substitutions per site. The analysis involved 12 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 175 positions in the final dataset. The two Families are clustered together as expected if the genetic data of the 175 bp at 5' end of the cyt *b* gene matches the current taxonomic groups at the taxonomic level of family. The green dots indicate the bootstrap value for a given interior branch is 95% or higher, then the topology at that branch is considered correct [47].

The results show that the use of the partial sequences of the ND2, ND5, cyt *b* and the 5' and 3' termini of the COI loci for phylogenetic tree reconstruction can distinguish the different species used in this study. The use of partial sequences of the ND2, 3' region of the COI and cyt *b* loci can distinguish between members of the Psittacidae (parrots) and Cacatuidae (cockatoos) families and place the members of the same family into their appropriate taxonomy group. Most clades that belong to the same species had a 100% bootstrap value. The exception was the partial sequences at 3' terminus of the cyt *b* gene where the *C. funereus* was placed on a branch next to *C. lathami* (43%).

In conclusion, the data suggest that these sections of the ND2, cyt *b* and the COI loci can reconstruct the expected phylogenetic tree and provide a means of species identification for members of the same genera. However, the partial sequence of the ND2 locus at base position 58-509 is superior for both species identification and phylogenetic tree reconstruction and in finches, parrots and cockatoo species due to the lower intra-species variation and the higher inter-species variation compared to the cyt *b* and the COI loci. The ND5 gene at base positions 101-552 showed less intra-species variation and higher inter-species variation than the cyt *b* and COI loci, it incorrectly placed the parrot group and the cockatoo group in the resulting phylogenetic tree.

3.4 Methods and models comparison for phylogenetic tree reconstruction of the parrot and cockatoo families

The trees generated can be different or similar based on the method and model that have been used for the calculation according to different assumptions. This part of the study aims to find the most suitable combination of method and model, using the MEGA 5 program, which generates the tree with fewest oddities using DNA sequence from members of the parrot and cockatoo families based on the 452 bp segment at 5' terminus of the ND2 and the ND5 loci. A comparison between all the methods and models provided in MEGA 5 and MrBayes programs were made. As expect, the tree created using MrBayes can cluster and placing the species to appropriate taxonomic group (Table 3.12).

Table 3.12: The comparison of the use of all methods and models provided in MEGA 5 and MrBayes programs.

Methods and models	ND2			ND5		
	Distinguishing species	Distinguishing out-group	Distinguishing parrots and cockatoos	Distinguishing species	Distinguishing out-group	Distinguishing parrots and cockatoos
MEGA program						
1. Maximum Likelihood						
- Jukes-Cantor model	✓	X	✓	✓	✓	X
- Kimura 2-parameter model	✓	X	✓	✓	✓	✓
- Tamura 3-parameter model	✓	X	✓	✓	✓	✓
- Hasegawa-Kishino-Yano model	✓	X	✓	✓	✓	✓
- Tamura-Nei model	✓	X	✓	✓	✓	✓
- General Time Reversible model	✓	X	✓	✓	✓	✓
2. Neighbor-Joining						
- No. of differences	✓	X	✓	✓	✓	✓
- p-distance	✓	X	✓	✓	✓	✓
- Jukes-Cantor model	✓	X	✓	✓	✓	✓
- Kimura 2-parameter model	✓	X	✓	✓	✓	✓
- Tajima-Nei model	✓	X	✓	✓	✓	X
- Tamura 3-parameter model	✓	X	✓	✓	✓	X
- Tamura-Nei model	✓	X	✓	✓	✓	X
- Maximum Composite Likelihood	✓	X	✓	✓	✓	X
- LogDet (Tamura-Kumar)	✓	X	X	✓	✓	✓
3. Minimum-Evolution						
- No. of differences	✓	X	✓	✓	✓	✓
- p-distance	✓	X	✓	✓	✓	✓
- Jukes-Cantor model	✓	X	✓	✓	✓	✓
- Kimura 2-parameter model	✓	X	✓	✓	✓	✓
- Tajima-Nei model	✓	X	✓	✓	✓	X
- Tamura 3-parameter model	✓	X	✓	✓	✓	X
- Tamura-Nei model	✓	X	✓	✓	✓	X
- Maximum Composite Likelihood	✓	X	✓	✓	✓	X
- LogDet (Tamura-Kumar)	✓	X	X	✓	✓	X
4. UPGMA						
- No. of differences	✓	✓	X	✓	✓	X
- p-distance	✓	✓	X	✓	✓	X
- Jukes-Cantor model	✓	✓	X	✓	✓	X
- Kimura 2-parameter model	✓	✓	X	✓	✓	X
- Tajima-Nei model	✓	✓	X	✓	✓	X
- Tamura 3-parameter model	✓	✓	X	✓	✓	X
- Tamura-Nei model	✓	✓	X	✓	✓	X
- Maximum Composite Likelihood	✓	✓	X	✓	✓	✓
- LogDet (Tamura-Kumar)	✓	✓	X	✓	✓	X
5. Maximum Parsimony	✓	X	✓	✓	X	X
MrBayes	✓	✓	✓	✓	✓	✓

The most appropriate phylogenetic methods for parrot and cockatoo families in this study can be found in Figure 3.26 (K2P/NJ) and Figure 3.27 (Bayesian analysis) and the other trees can be found in the Appendix D.

The combining NJ and bootstrap analysis [41] is the optimum way to evaluate trees using distance methods [35]. This method is related to clustering without assuming a clock-like behaviour. Nevertheless, most clusters in the constructed using K2P model with NJ method are composed of the same species.

MrBayes program uses Bayesian analysis to reconstruct the phylogenetic tree by using a variant of Markov chain Monte Carlo (MCMC) algorithm for approximating the posterior probabilities [58, 59]. A 50% majority rule consensus tree is constructed and found that this method is better than the other in distinguishing and placing the parrots and cockatoos in this study to an appropriate taxonomic group.

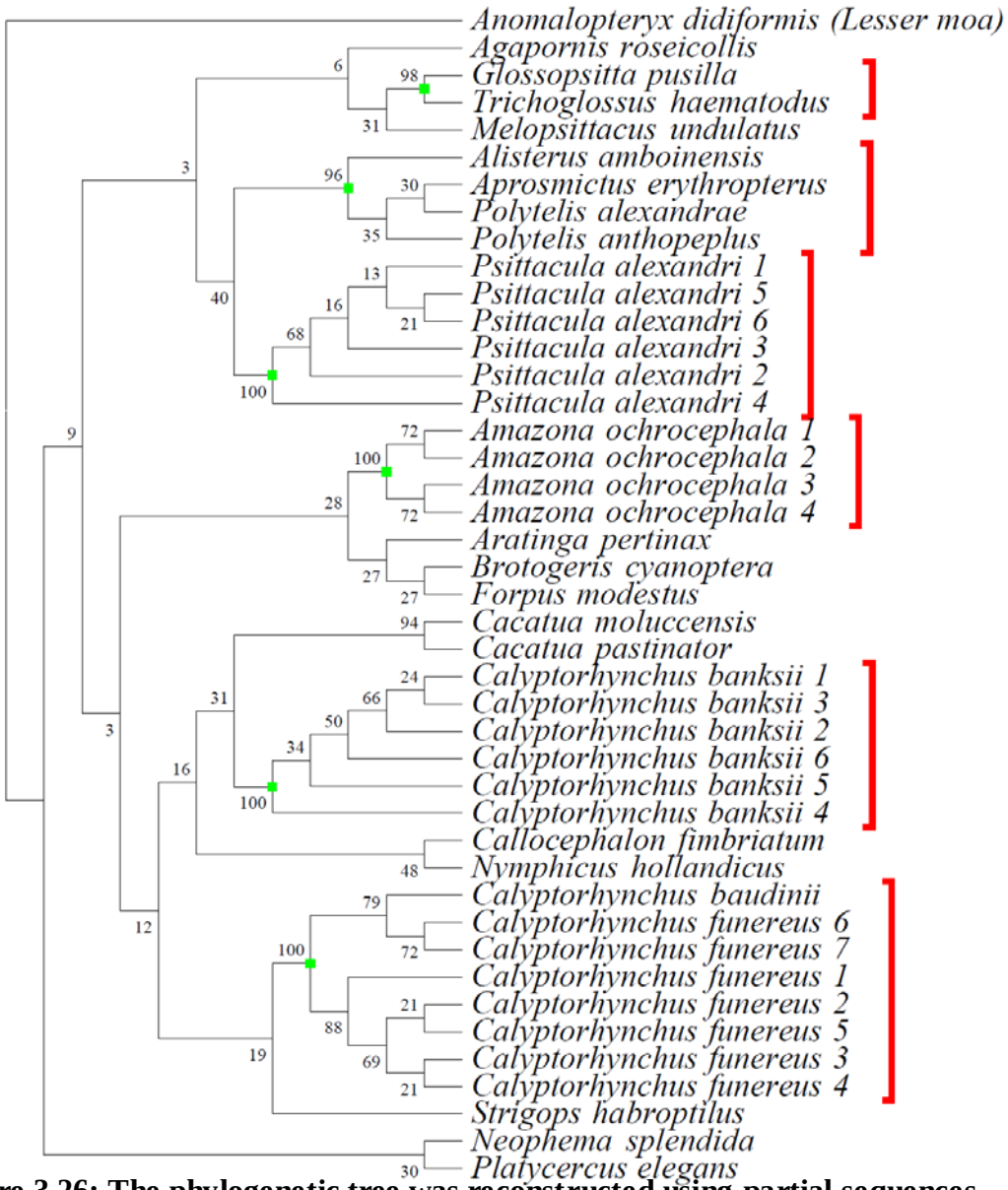


Figure 3.26: The phylogenetic tree was reconstructed using partial sequences at 5' terminus (57-508, 452 bp) of the ND5 gene of 15 species of parrots and 8 species of cockatoos. The out group was a moa. Evolutionary analyses were conducted in MEGA5 [43]. The evolutionary history was inferred using the Neighbor-Joining method [44]. The bootstrap consensus tree inferred from 1000 replicates [45] is taken to represent the evolutionary history of the taxa analyzed [45]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [45]. The evolutionary distances were computed using the Kimura 2-parameter method [46] and are in the units of the number of base substitutions per site. The analysis involved 43 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 452 positions in the final dataset. The green dots indicate the bootstrap value for a given interior branch is 95% or higher, then the topology at that branch is considered correct [47].

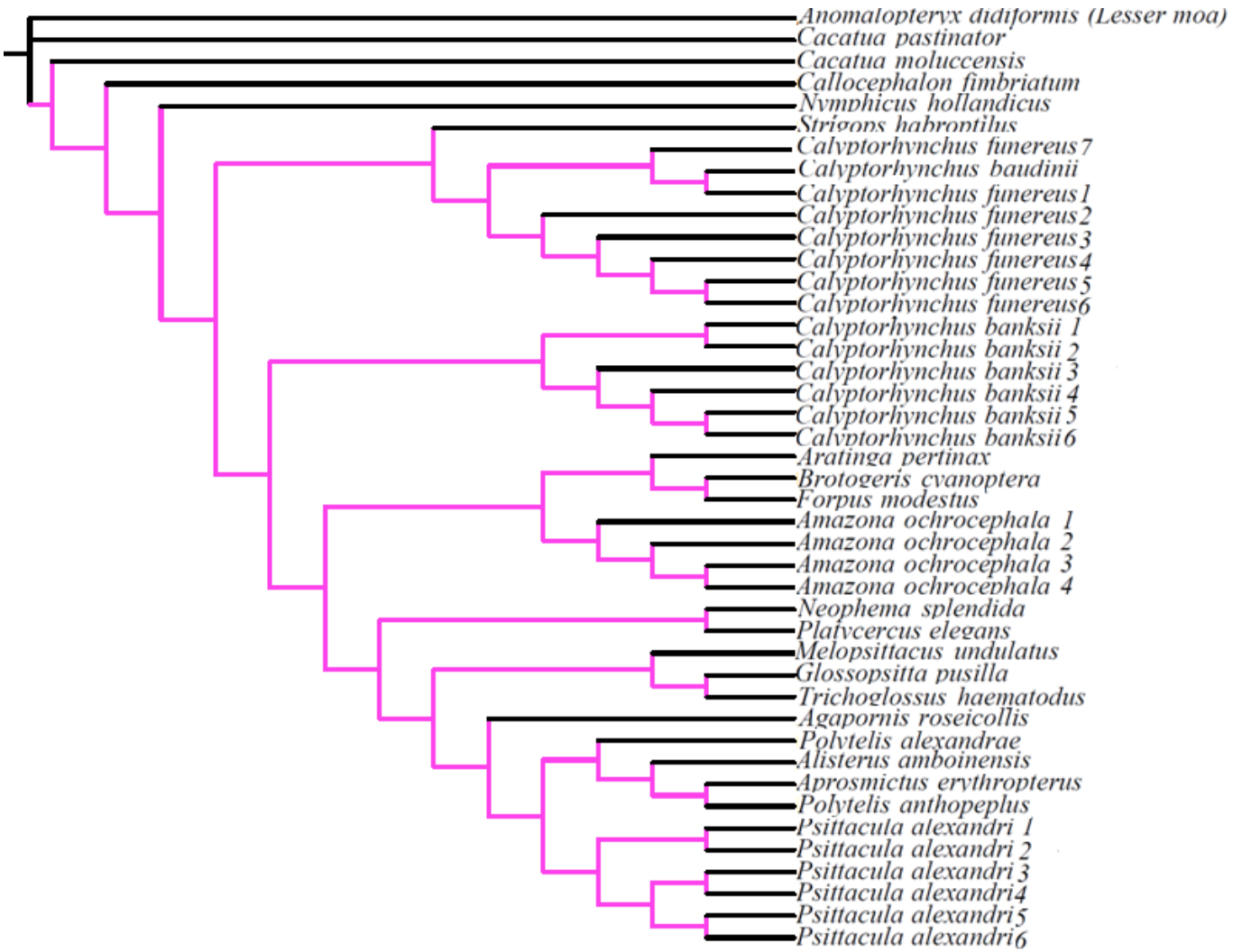


Figure 3.27: The phylogenetic trees generated using a range of parrot and cockatoo species using the partial sequences of the ND5 gene at 5' terminus created from the MrBayes program. The out-group was a moa. The pink branches indicate the bootstrap value for a given interior branch is 95% or higher, then the topology at that branch is considered correct [47].

3.5 Inter- and intra-species variations in closely related species

As highly degraded DNA is typical of the type of samples encountered in forensic science, an aim is to identify the smallest section of a locus that can identify unambiguously any member of an avian species. In line with previous work in this thesis, the ideal section for this purpose would also show little intra-species variation and high inter-species variation to allow species that are close genetic relatives to be distinguished.

3.5.1 Inter- and intra-species of the Fringillidae family (finches) at genus and species taxonomic level using partial sequences of the cyt *b*, COI and ND2 loci

The partial sequences of the ND2 gene obtained from this study are five sequences of *F. coelebs*, six sequences of *F. montifringilla*, five sequences of *C. chloris*, six sequences of *C. carduelis* and six sequences of *C. spinus*. The other sequences were obtained from the GenBank DNA database. The pairwise distance between finch species was calculated by the MEGA 5 program using Kimura 2-parameter model. The results are shown in Figures 3.28-3.30. The ND5 locus was not used in this part of study according to insufficient sequence data of this locus from *Fringilla* and *Carduelis* genera on the database.

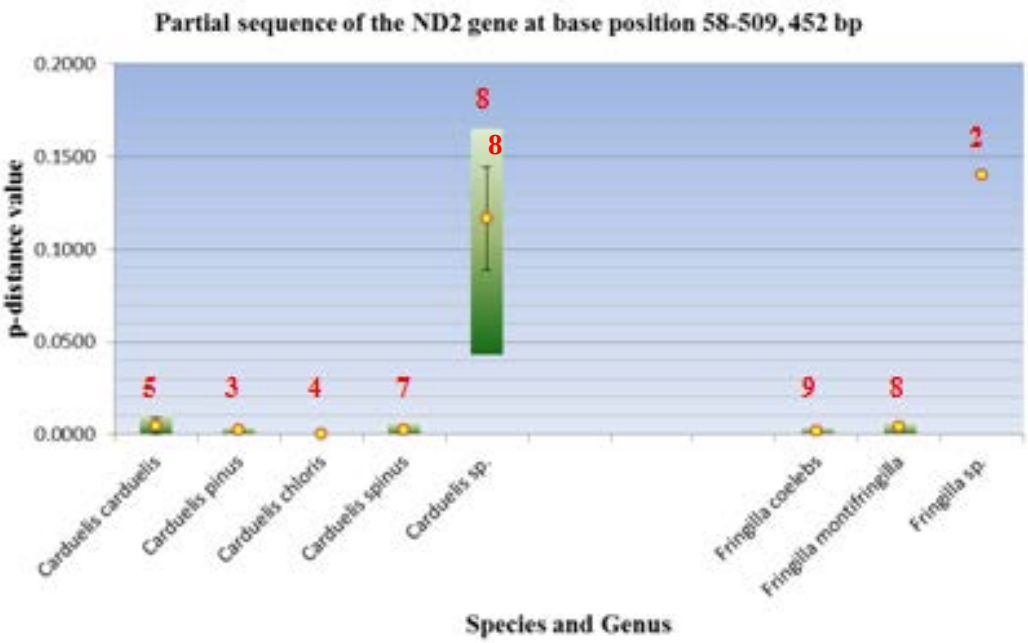


Figure 3.28: Inter- and intra-species of the Fringillidae family (finches) at genus and species taxonomic level using partial sequences of the ND2 gene at base positions 58-509, 452 bp. The numbers in red indicates the numbers of sequences used in this study.

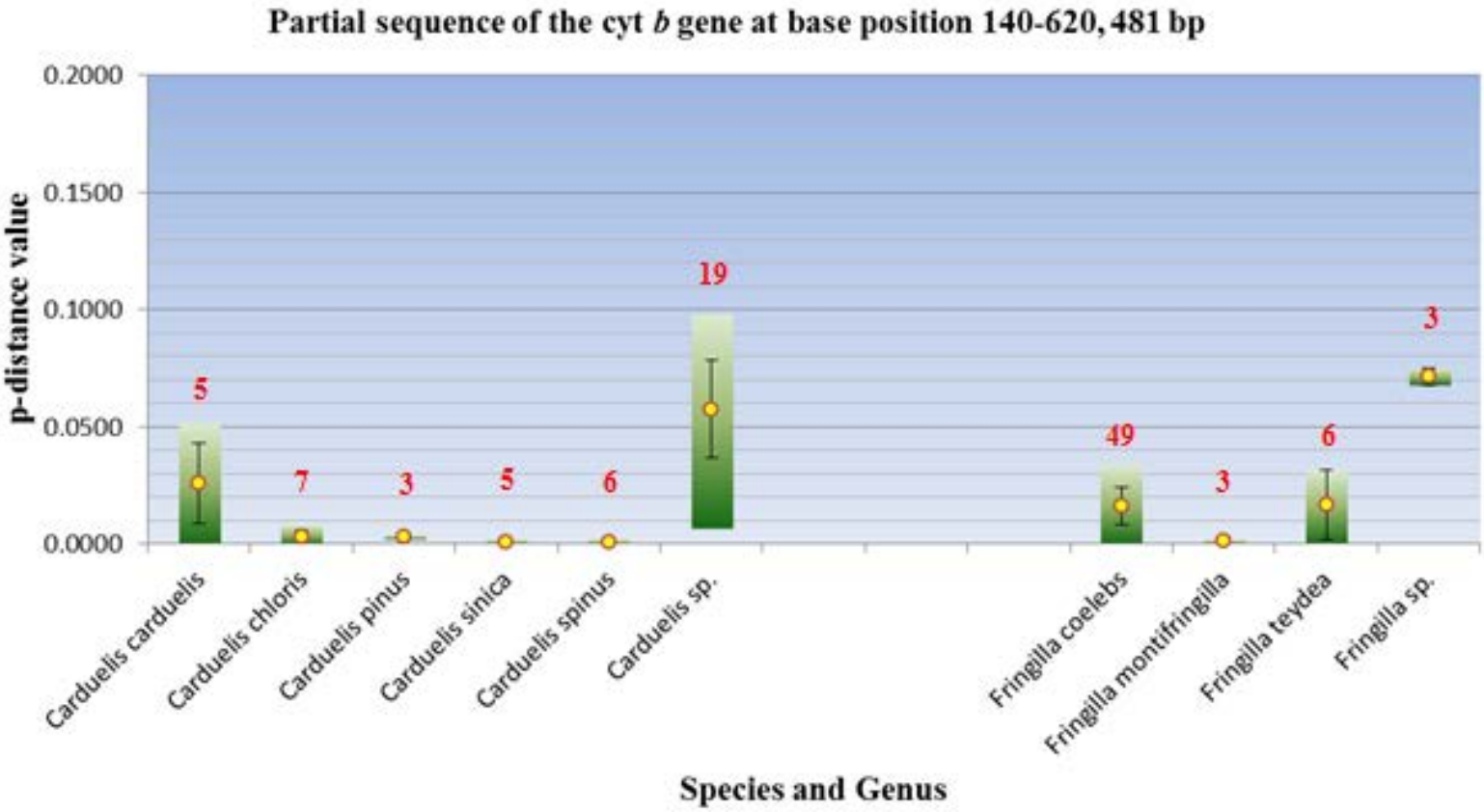


Figure 3.29: Inter- and intra-species of the Fringillidae family (finches) at genus and species taxonomic level using partial sequences of the cyt *b* gene at base positions 140-620, 481 bp. The numbers in red indicates the numbers of sequences used in this study.

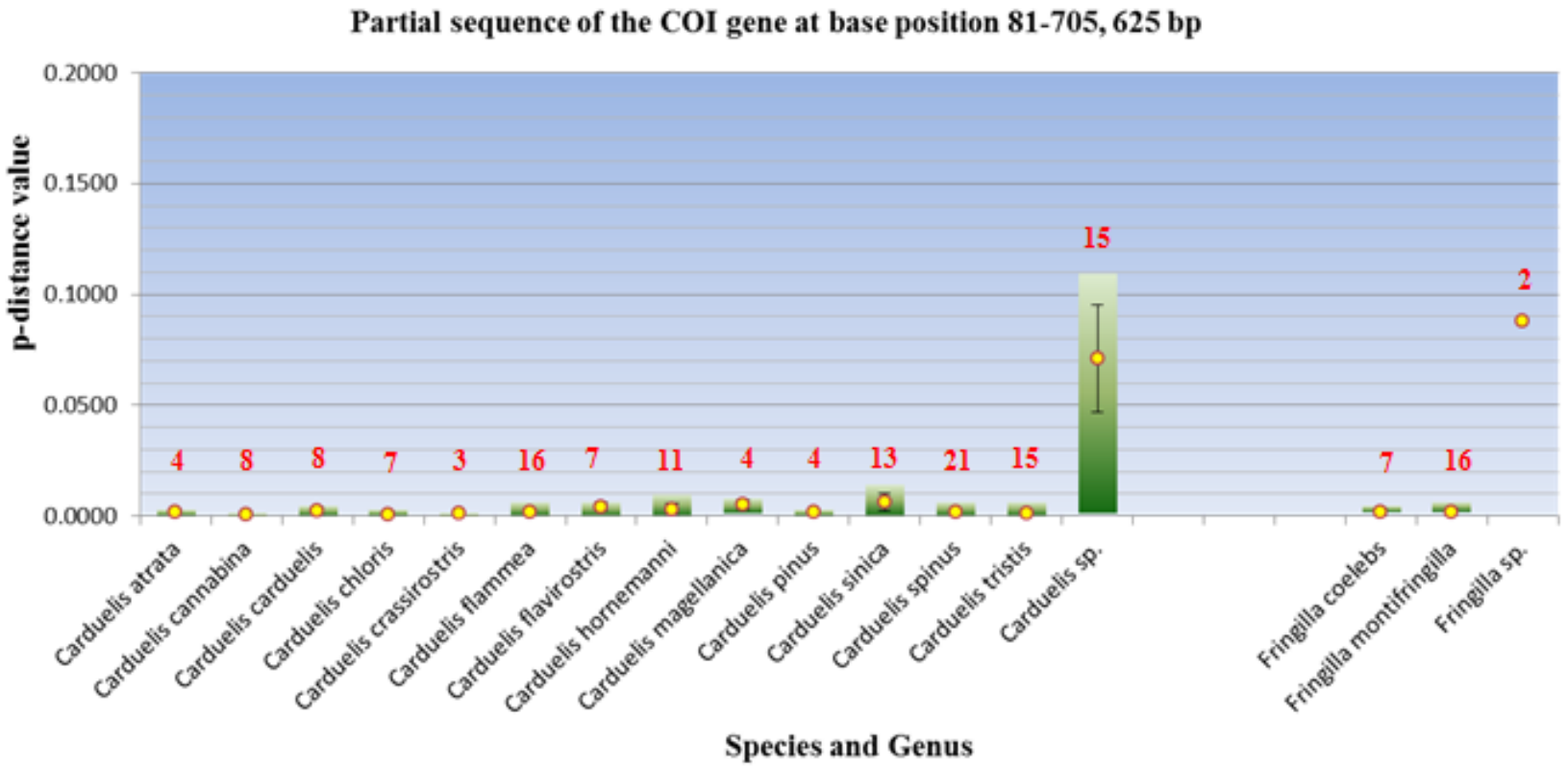


Figure 3.30: Inter- and intra-species of the Fringillidae family (finches) at genus and species taxonomic level using partial sequences of the COI gene at base positions 81-705, 625 bp. The numbers in red indicates the numbers of sequences were used in this study.

3.5.2 Inter- and intra-species Psittacidae family (parrots) and Cacatuidae family (cockatoos) at genus and species taxonomic level using partial sequences of the cyt *b*, COI, ND2 and ND5 loci

The pairwise distance between parrot and cockatoo species were calculated by MEGA 5 program using Kimura 2-parameter model. The results are shown in Figure 3.31- 3.35.

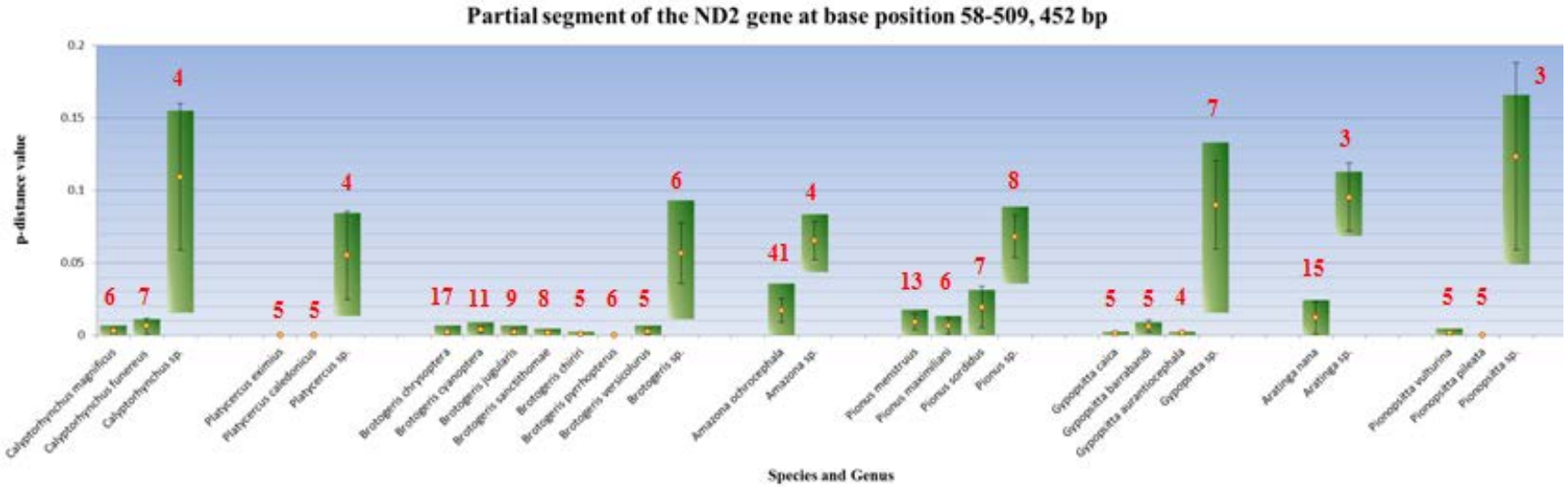


Figure 3.31: Inter- and intra-species of the Psittacidae family (parrots) and Cacatuidae family (cockatoos) at genus and species taxonomic level using partial sequences of the ND2 gene at base positions 58-509, 452 bp. The numbers in red indicates the numbers of sequences used in this study.

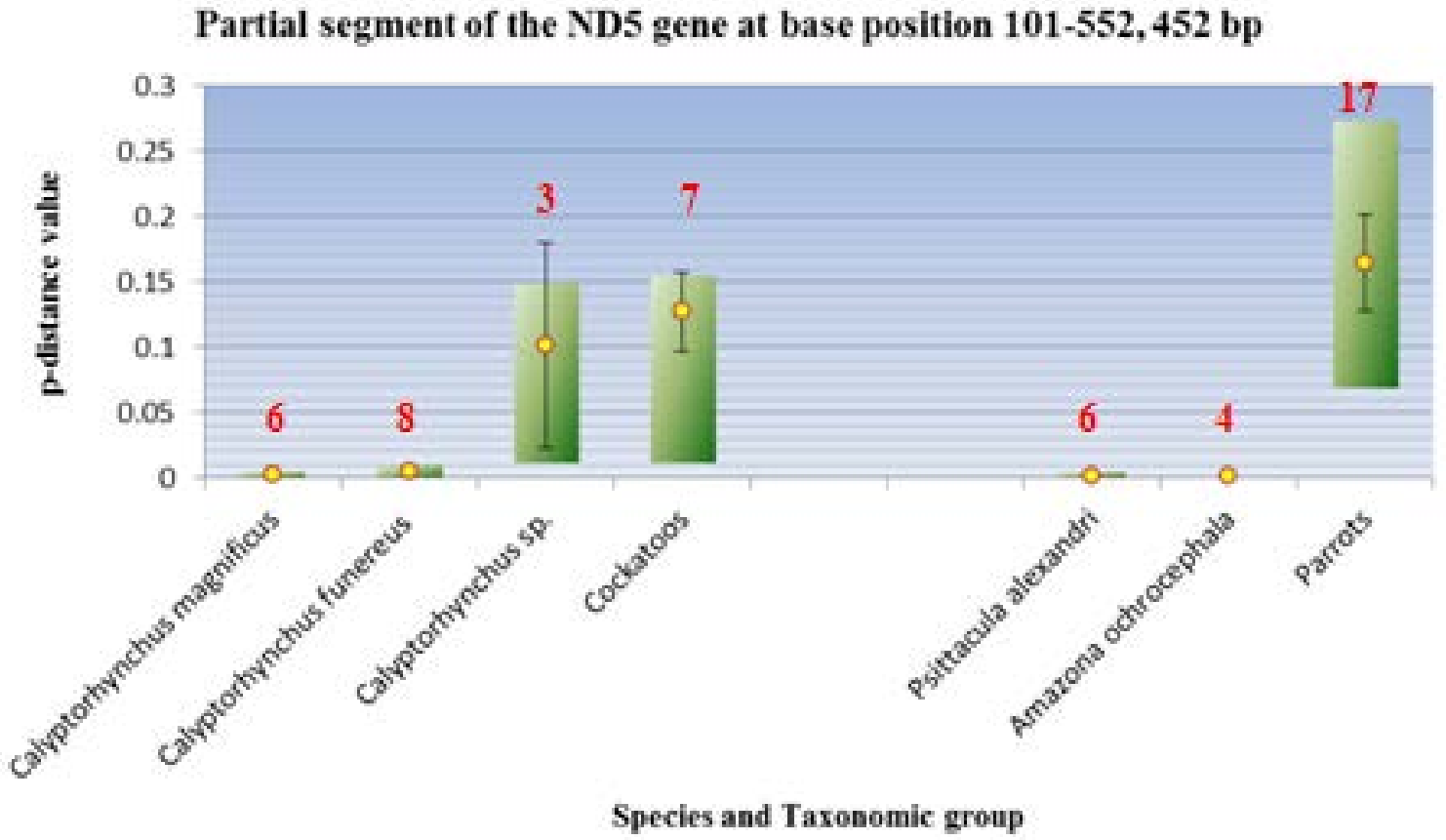


Figure 3.32: Inter- and intra-species of the Psittacidae family (parrots) and Cacatuidae family (cockatoos) at genus and species taxonomic level using partial sequences of the ND5 gene at base positions 101-552, 452 bp. The numbers in red indicates the numbers of sequences used in this study.

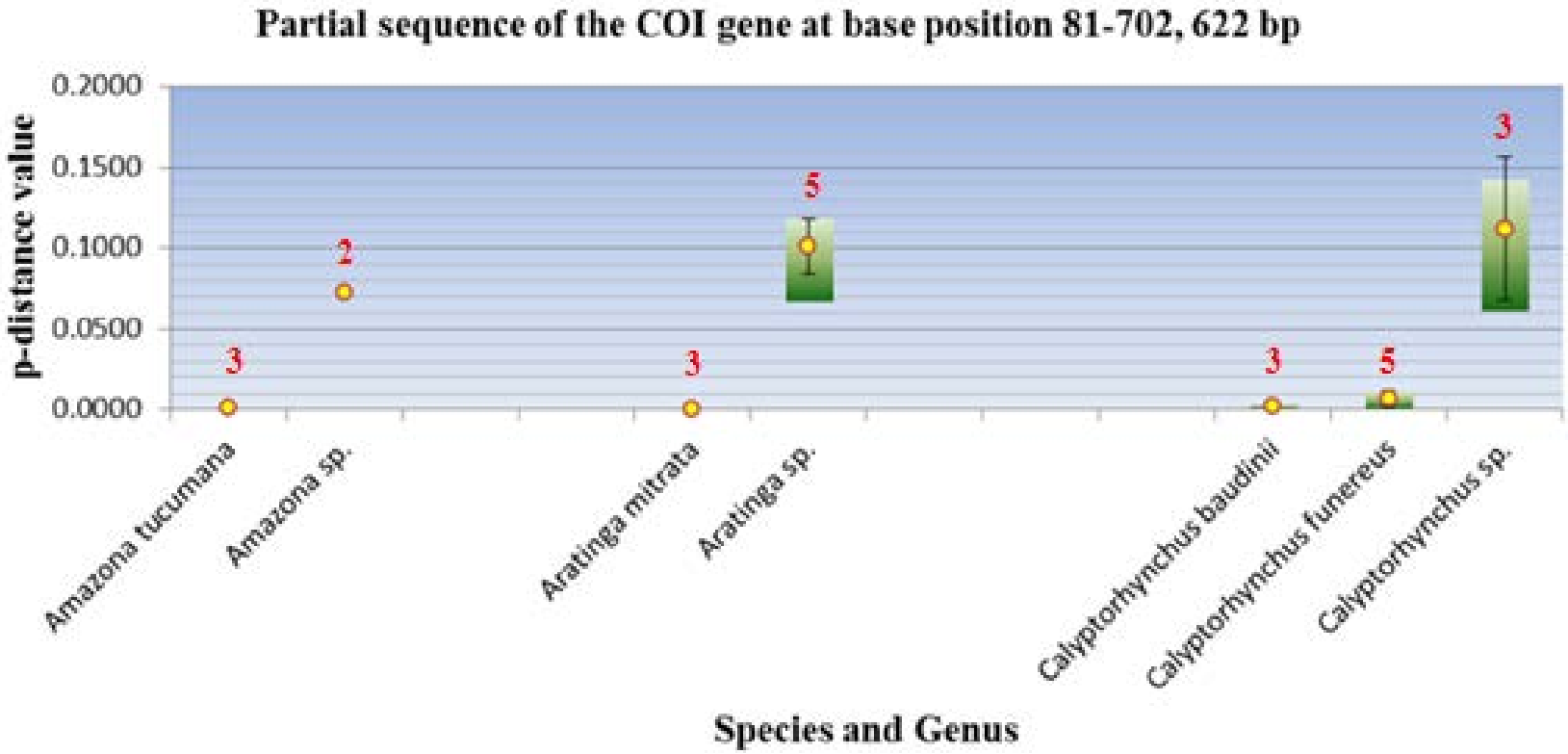


Figure 3.33: Inter- and intra-species of the Psittacidae family (parrots) and Cacatuidae family (cockatoos) at genus and species taxonomic level using partial sequences of the COI gene at base positions 81-702, 622 bp. The numbers in red indicates the numbers of sequences used in this study.

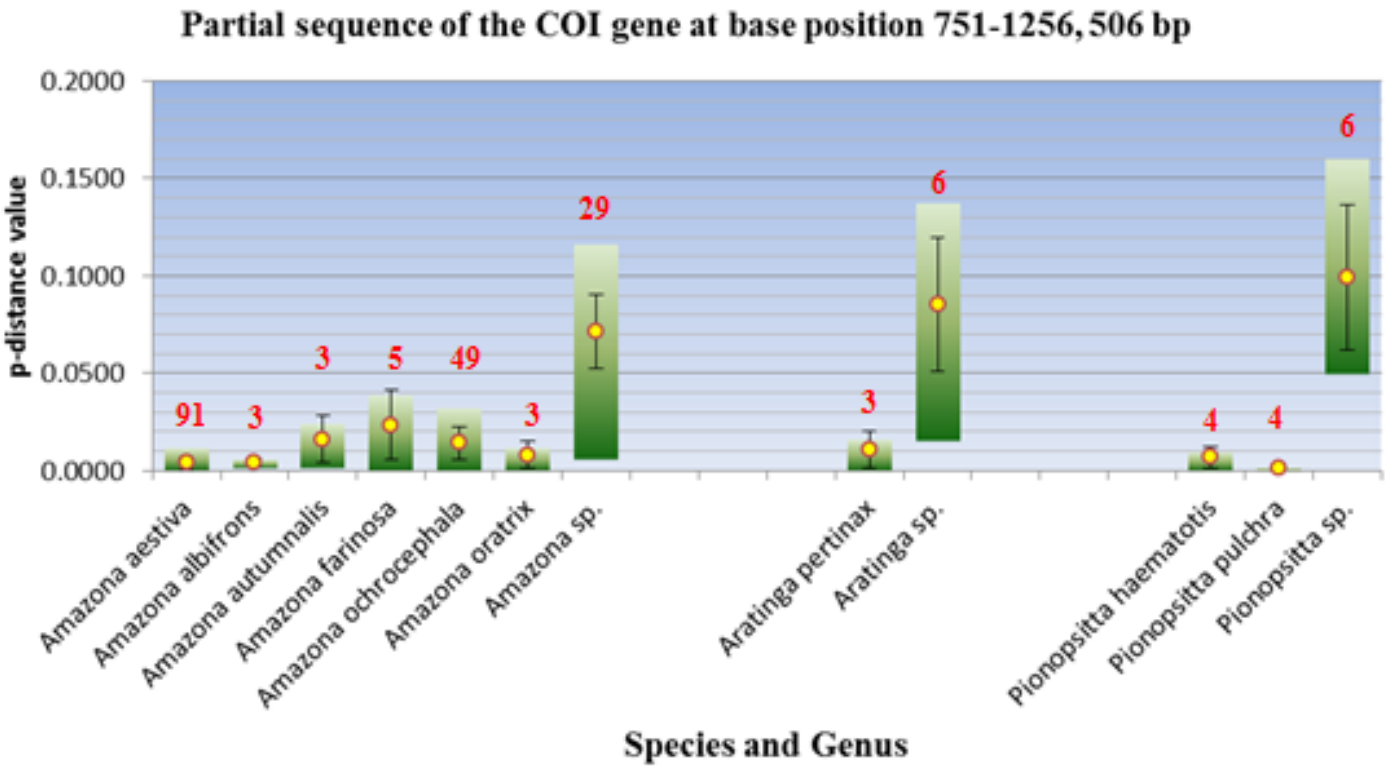


Figure 3.34: Inter- and intra-species of the Psittacidae family (parrots) and Cacatuidae family (cockatoos) at genus and species taxonomic level using partial sequences of the COI gene at base positions 751-1256, 506 bp. The numbers in red indicates the numbers of sequences used in this study.

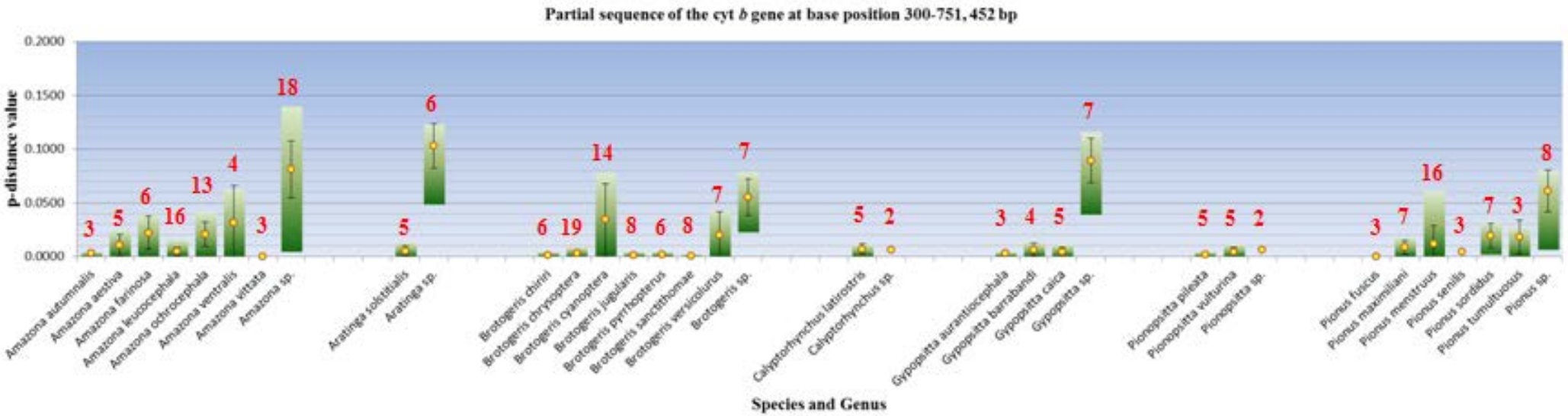


Figure 3.35: Inter- and intra-species of the Psittacidae family (parrots) and Cacatuidae family (cockatoos) at genus and species taxonomic level using partial sequences of the cyt *b* gene at base positions 300-751, 452 bp. The numbers in red indicates the numbers of sequences used in this study.

The partial sequences of the ND2 and the ND5 loci obtained from this study included a total of 47 sequences; six samples of ND2 and ND5 from *C. banksii*, seven samples of ND2 and ND5 from *C. funereus*, five sequences of ND2 and six partial sequences of ND5 from *P. alexandri* and six sequences of ND2 and four sequences of ND5 from *A. ochrocephala*. The sequences from the other species were obtained from the GenBank DNA database.

The use of partial sequences of the ND2 gene at base positions 58-509 and the ND5 gene at base positions 101- 552 were able to identify closely related species of both the finch, parrot and cockatoo families, as shown in Figures 3.28, 3.31 and 3.32. The COI gene at 5' terminus at base positions 81-705 was found to be highly conserved; therefore this region cannot identify *Carduelis* sp. as shown in Figure 3.30. However based on sequence data from GenBank, this region can identify species of parrot and cockatoo as shown in Figure 3.33.

The genus Fringilla is a very small taxonomic group as there are three species within this genus; being *F. coelebs*, *F. teydea* and *F. montifringilla*. The segment from the cyt *b* locus at base positions 140-620 (480 bp) can distinguish individual species within this genus, as shown in Figure 3.29. The 452 bp at base positions 300-751 of the cyt *b* gene can distinguish closely related species of *Aratinga* sp. and *Gypopsitta* sp., as shown in Figure 3.35.

The 3’ terminus of the COI gene was found to show higher variation than the 5’ terminus, therefore 506 bp of this region at base positions 751-1256 was tested in parrot and cockatoo species. The result shows that 3’ region of the COI gene can distinguish the *Aratinga* sp. and *Pionopsitta* sp. as shown in Figure 3.34.

In conclusion, there are only the 452 segments at the 5’ terminus of the ND2 and ND5 loci that show no overlap of the gap between intra-species variation and inter-species variation. This suggests that both segments can identify closely related species of the finch, parrot and cockatoo families. These loci are currently not used in avian species identification yet this work illustrates the benefits of using these loci, and potentially using the both loci. Especially, these small segments of ND2 and ND5 are ideal for forensic application according to the amplicon size (452 bp) is applicability to low quality and quantity or degraded DNA samples.

Chapter 3 References

1. An, J., et al., *A molecular genetic approach for species identification of mammals and sex determination of birds in a forensic case of poaching from South Korea.* Forensic Science International Affairs, 2007. 167(1): p. 59-61.

2. Lee, J.C.I., et al., *A novel strategy for avian species identification by cytochrome b gene.* Electrophoresis, 2008. 29(11): p. 2413-2418.

3. Linacre, A. and S. Tobe, *An overview to the investigative approach to species testing in wildlife forensic science.* Investigative Genetics, 2011. 2(1): p. 2.

4. Namikoshi, A., et al., *Species identification of Alaska pollock, Gadus spp., and Micromesistius spp. in cod roe products using a PCR-based method,* in *Fisheries Science*. 2011, Springer Japan. p. 671-678.

5. Smith, M.A., N.A. Poyarkov, and P.D.N. Hebert, *DNA Barcoding: CO*1 *DNA barcoding amphibians: take the chance, meet the challenge.* Molecular Ecology Resources, 2008. 8(2): p. 235-246.

6. Holmes, B.H., D. Steinke, and R.D. Ward, *Identification of shark and ray fins using DNA barcoding.* Fisheries Research, 2009. 95(2-3): p. 280-288.

7. Hebert, P.D.N., et al., *Identification of Birds through DNA Barcodes.* PLoS Biol, 2004. 2(10): p. e312.

8. Park, H.Y., et al., *New DNA barcodes for identification of Korean birds.* Genes & Genomics, 2011. 33(2): p. 91-95.

9. Tavares, E. and A. Baker, *Single mitochondrial gene barcodes reliably identify sister-species in diverse clades of birds.* Bmc Evolutionary Biology, 2008. 8(1): p. 81.

10. Lohman, D., D. Prawiradilaga, and R. Meier, *Improved COI barcoding primers for Southeast Asian perching birds (Aves: Passeriformes).* Molec Ecol Res, 2009. 9(1): p. 37-40.

11. Wong, E.H.K., M.S. Shivji, and R.H. Hanner, *Identifying sharks with DNA barcodes: assessing the utility of a nucleotide diagnostic approach.* Molecular Ecology Resources, 2009. 9: p. 243-256.

12. Chang, C.-H., R. Rougerie, and J.-H. Chen, *Identifying earthworms through DNA barcodes: Pitfalls and promise.* Pedobiologia, 2009. 52(3): p. 171-180.

13. Huang, J., et al., *Identifying earthworms through DNA barcodes.* Pedobiologia, 2007. 51(4): p. 301-309.

14. Hebert, P.D.N., et al., *Biological identifications through DNA barcodes.* Proceedings of the Royal Society of London. Series B: Biological Sciences, 2003. 270(1512): p. 313-321.

15. Hebert, P.D.N., et al., *Barcoding animal life: cytochrome c oxidase subunit* 1 *divergences among closely related species.* Proceedings of the Royal Society of London Series B-Biological Sciences, 2003. 270: p. S96-S99.

16. Wilson-Wilde, L., et al., *Current issues in species identification for forensic science and the validity of using the cytochrome oxidase I (COI) gene.* Forensic Science Medicine and Pathology, 2010. 6(3): p. 233-241.

17. Zhou, H., et al., *Molecular phylogeny of Nearctic species of Rhynchelmis (Annelida).* Zoologica Scripta, 2010. 39(4): p. 378-393.

18. Meusnier, I., et al., *A universal DNA mini-barcode for biodiversity analysis.* BMC Genomics, 2008. 9(1): p. 214.

19. Hajibabaei, M., et al., *A minimalist barcode can identify a specimen whose DNA is degraded.* Molecular Ecology Notes, 2006. 6: p. 959 - 964.

20. Rasmussen, R.S., M.T. Morrissey, and P.D.N. Hebert, *DNA Barcoding of Commercially Important Salmon and Trout Species (Oncorhynchus and Salmo) from North America.* Journal of Agricultural and Food Chemistry, 2009. 57(18): p. 8379-8385.

21. Kocher, T.D., et al., *Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers* Proceedings of the National Academy of Sciences 1989 86 (16): p. 6196-6200

22. Johnson, K. and M. Sorenson, *Comparing Molecular Evolution in Two Mitochondrial Protein Coding Genes (Cytochrome b and ND*2) *in the Dabbling Ducks (Tribe : Anatini).* Blood, 1998. 10(1): p. 82-94.

23. Parson, W., et al., *Species identification by means of the cytochrome b gene.* International Journal of Legal Medicine, 2000. 114(1): p. 23-28.

24. López-Andreo, M., et al., *Identification and quantitation of species in complex DNA mixtures by real-time polymerase chain reaction.* Analytical Biochemistry, 2005. 339(1): p. 73-82.

25. Coghlan, M.L., et al., *Egg forensics: An appraisal of DNA sequencing to assist in species identification of illegally smuggled eggs.* Forensic science international. Genetics, 2012. 6(2): p. 268-273.

26. Whiteman, N.K., et al., *Molecular and Morphological Divergence in a Pair of Bird Species and Their Ectoparasites.* Journal of Parasitology, 2009. 95(6): p. 1372-1382.

27. Dawnay, N., et al., *Validation of the barcoding gene COI for use in forensic genetic species identification.* Forensic Science International, Genetics Supplement Series, 2007. 173(1): p. 1-6.

28. Folmer, O., et al., *DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates.* 1994. 3: p. 294–299.

29. Kerr K.C.R., et al., *Comprehensive DNA barcode coverage of North American birds.* Mol Ecol Notes, 2007. 7: p. 535-543.

30. Campagna, L., et al., *DNA barcodes provide new evidence of a recent radiation in the genus Sporophila (Aves: Passeriformes).* Molecular Ecology Resources, 2010. 10(3): p. 449-458.

31. Yoo, H., et al., *DNA barcoding Korean birds.* Molec Cells, 2006. 22(3): p. 323 - 327.

32. Johnsen, A., et al., *DNA barcoding of Scandinavian birds reveals divergent lineages in trans-Atlantic species.* Journal of Ornithology, 2010. 151(3): p. 565-578.

33. Khan, H.A., I.A. Arif, and M. Shobrak, *DNA Barcodes of Arabian Partridge and Philby's Rock Partridge: Implications for Phylogeny and Species Identification.* Evolutionary Bioinformatics, 2010. 6: p. 151-158.

34. Kocher, T., et al., *Dynamics of mitochondrial-DNA evolution in animals - amplification and sequencing with conserved primers.* Proc Natl Acad Sci USA, 1989. 86(16): p. 6196 - 6200.

35. Rudnick, J.A., et al., *Species identification of birds through genetic analysis of naturally shed feathers.* Molecular Ecology Notes, 2007. 7(5): p. 757-762.

36. Tobe, S., A. Kitchener, and A. Linacre, *Reconstructing mammalian phylogenies: a detailed comparison of the cytochrome b and cytochrome oxidase subunit I mitochondrial genes.* PLoS ONE, 2010. 1(1): p. 71-73.

37. Irwin, D., T. Kocher, and A. Wilson, *Evolution of the cytochrome b gene of mammals.* Journal of Molecular Evolution, 1991. 32(2): p. 128-144.

38. Castresana, J., *Cytochrome b Phylogeny and the Taxonomy of Great Apes and Mammals.* Molecular Biology and Evolution, 2001. 18(4): p. 465-471.

39. Bellinger, M.R., et al., *Taxonomic relationships among Phenacomys voles as inferred by cytochrome b.* Journal of Mammalogy, 2005. 86(1): p. 201-210.

40. Unajak, S., et al., *Identification of species (meat and blood samples) using nested-PCR analysis of mitochondrial DNA.* African Journal of Biotechnology, 2011. 10(29): p. 5670-5676.

41. Hackett, S.J., *Molecular phylogenetics and biogeography of tanagers in the genus Ramphocelus (Aves).* Molecular Phylogenetics and Evolution, 1996. 5(2): p. 368-382.

42. Ian, L., *Examples of Comparative Biology. Diagrams for the peptide CYTB from several species.* [*cited* 16 *August* 2011.]*; Available from: http://www.ianlogan.co.uk/TMRjpgs/comp_biol/cytb_comp.htm.* 2011.

43. Tamura K, et al., *MEGA*5: *Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods.* Molecular Biology and Evolution, 2011. 28: p. 2731-2739.

44. Saitou N and N. M, *The neighbor-joining method: A new method for reconstructing phylogenetic trees.* Molecular Biology and Evolution, 1987. 4: p. 406-425.

45. Felsenstein, J., *Confidence Limits on Phylogenies: An Approach Using the Bootstrap.* Evolution, 1985. 39(4): p. 783-791.

46. Kimura, M., *A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences.* Journal of Molecular Evolution, 1980. 16(2): p. 111 - 120.

47. Nei, M. and S. Kumar, *Molecular Evolution and Phylogenetics.* Oxford University Press, New York, 2000.

48. Gabriel, M.N., et al., *Improved MtDNA sequence analysis of forensic remains using a "mini-primer set" amplification strategy.* Journal of Forensic Sciences, 2001. 46(2): p. 247-253.

49. Divine, A.-M., *Evaluation of New technologies for Forensic DNA analysis.* Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. Acta Universitatis Upsaliensis, 2005. 27: p. 59 pp.

50. Edlund, H., *Sensitive Identification Tool in Forensic DNA analysis.* Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. Acta Universitatis Upsaliensis, 2010. 606: p. 49 pp.

51. van Berkum, P., et al., *Discordant phylogenies within the rrn loci of rhizobia.* Journal of Bacteriology, 2003. 185(10): p. 2988-2998.

52. van der Kuyl, A.C., et al., *Routine DNA analysis based on* 12*S rRNA gene sequencing as a tool in the management of captive primates.* Journal of Medical Primatology, 2000. 29(5): p. 309-317.

53. Miller S.R., M.D. Purugganan, and S.E. Curtis, *Molecular population genetics and phenotypic diversification of two populations of the thermophilic cyanobacterium Mastigocladus laminosus*. Applied and Environmental Microbiology, 2006. 72.(4): p. 2793-2800.

54. Hafner, J.C., et al., *Basal clades and molecular systematics of heteromyid rodents*. Journal of Mammalogy, 2007. 88(5): p. 1129-1145.

55. Parker, M.A., et al., *Conflicting phylogeographic patterns in rRNA and nifD indicate regionally restricted gene transfer in Bradyrhizobium.* Microbiology, 2002. 148(8): p. 2557-2565.

56. Wertz, J.E., et al., *A molecular phylogeny of enteric bacteria and implications for a bacterial species concept*. Journal of Evolutionary Biology, 2003. 16(6): p. 1236-1248.

57. Hackett, S.J., et al., *A phylogenetic study of birds reveals their evolutionary history*. Science, 2008. 320: p. 1763-1767.

58. Ronquist, F. and J.P. Huelsenbeck, *MRBAYES* 3: *Bayesian phylogenetic inference under mixed models*. Bioinformatics, 2003. 19: p. 1572-1574.

59. Huelsenbeck, J.P. and F. Ronquist, *MRBAYES: Bayesian inference of phylogeny*. Bioinformatics, 2001. 17: p. 754-755.