

Student: Tom Garden

Supervisors: Howard Ross and Craig Millar

## **Summer Studentship Report: DNA Barcoding the Birds of New Zealand** **– How reliable will be this identification system?**

### **Executive Summary**

The genetic distinctiveness of 210 New Zealand bird species was analysed by assessing the ratio of intraspecific distance over interspecific distance to the species' closest relative, using 7653 DNA barcode sequences downloaded from online databases. It was found that approximately 82% of species are genetically distinctive enough to enable reliable species identification by DNA barcoding. A significant number of sequences present in the databases were found to be mislabeled and/or misidentified; these must be identified and corrected to improve the reliability further. There are also some bird genera which are severely under-sampled in the databases, which can lead to species appearing more distinctive than they actually are.

### **Background/ Introduction**

DNA barcoding in animals involves sequencing of a standardized portion of the mitochondrial cytochrome c oxidase I (COI) gene in many species. This sequence data is then deposited in a public reference database to aid in the quick identification of future unknown specimens (Frezal & Leblois, 2008). The reliability of these techniques has been criticized, for reasons including: assuming that within-species variation is well sampled in the data set; not being able to recognize when no members of the test species are in the data set; assuming that species are reciprocally monophyletic, and being unable to deal with hybridization (Ross et al., 2008).

In the last 6 years some large scale projects have been undertaken in barcoding birds, such as a survey of 260 species of North American birds in 2004 (Herbert et al., 2004), and 643 North American birds in 2007 (Kerr et al., 2007), which have found species identification by DNA barcoding to be generally straightforward (Herbert et al., 2004). The reliability of species identification by DNA barcoding was tested by simulation by Ross et al. (2008). It was found that the level of genetic differentiation between species was critical in determining how successful an identification will be, with the probability of a correct identification falling greatly when the ratio of intraspecific/interspecific distance increases over ~0.6 – 0.7 (Ross et al., 2008).

Currently an effort is being undertaken to barcode all the birds species of New Zealand, and the majority have had their barcode sequences determined. This aim of this summer studentship is to assess how reliably the birds of New Zealand can be identified to species level using these DNA barcodes. To do this, barcode sequences for New Zealand (native and introduced) species and their overseas relatives were downloaded from internet databases such as the NZ Birds barcoding project, Genbank, and the Barcode of Life Data (BOLD). Various software programs were used to align the sequences and calculate the

average genetic distance within the test species, determine which other species is its closest relative, and to calculate the average genetic distance to this relative. A ratio of the intra-specific distance divided by the inter-specific distance was then calculated.

The goal of this study is to assess the genetic distinctiveness of New Zealand birds, and therefore how reliable species identification using DNA barcoding will be in these birds.

## Results

In this project, 210 species New Zealand species were analysed; 65% of New Zealand's approximately 321 total bird species, and 7653 sequences used altogether (Table 1). A minimum of two barcode sequences was required to test the reliability of species identification. Consequentially 111 species could not be assessed for this reason, and also if no sequences for any of their close relatives could be found. An example is the Rifleman (*Acanthisitta chloris*), where no sequences for its single close relative could be found. Groups where the taxonomy is significantly unresolved were also left out because of the difficulty in assigning sequences to species. This is why the Kiwi group (genus *Apteryx*) was left out.

Order	Total # NZ species in order	Well identifiable (R < 0.6)	Poorly identifiable (R > 0.6)	Due to suspected mis-labeling	# in v. poorly sampled genus	Not assessed, insufficient # sequences
Anseriformes	21	6	3	1	0	12
Apodiformes	2	0	0	0	0	2
Charadriiformes	88	69	9	0	0	10
Ciconiiformes	16	3	2	2	0	11
Columbiformes	4	2	2	1	1	0
Coraciiformes	3	2	0	0	2	1
Cuculiformes	6	1	1	0	1	4
Falconiformes	5	2	1	0	0	2
Galliformes	10	6	1	0	2	3
Gruiformes	10	6	0	0	0	4
Passeriformes	43	26	3	0	5	14
Pelecaniformes	18	4	1	0	2	13
Phaethontiformes	2	0	0	0	0	2
Podicipediformes	4	1	0	0	0	3
Procellariiformes	60	32	5	3	4	23
Psittaciformes	10	6	1	0	0	3
Sphenisciformes	13	11	2	0	0	0
Strigiformes	3	2	0	0	1	1
Struthioniformes	3	0	0	0	0	3
<b>Total #</b>	<b>321</b>	<b>179</b>	<b>31</b>	<b>7</b>	<b>18</b>	<b>111</b>
<b>% of 210 species analysed</b>		<b>82.68</b>	<b>17.32</b>	<b>3.33</b>	<b>8.57</b>	

Table 1: A summary table showing number of species that are relatively well identifiable and poorly identifiable for each bird order in New Zealand. Of those that were found to be poorly identifiable, the number that are likely to be caused by mislabeling is shown. Also shown is the number of species who are a member of a very poorly sampled genus, and therefore whose closest relative may not be present in the database. The number of species which could not be assessed due to insufficient sequences in the database is also shown.

Most of the New Zealand bird species appear to be readily distinguishable from their close relatives by DNA barcoding, with low inter-specific distance to intra-specific distance ratios. An accurate formula used to determine the probability of attaining a correct identification from this ratio is not yet available.

There are some species which have very high ratios, and which therefore could not be distinguished. These include some duck species of the genus *Anas*, such as the Mallard duck (*Anas platyrhynchos*), which are known to frequently hybridise with other duck species (Tubaro & Lijtmaer, 2002). The gull *Larus dominicanus* (Black-Backed Gull) was also found to be indistinguishable from the Northern-Hemisphere Lesser Black-Backed Gull (*Larus fuscus*). The taxonomy of the *Larus* genus is known to be very complicated, and this can be clearly seen when a phylogenetic tree is created for the group. Very high ratios (high intra-specific distance compared to inter-specific distance) were also found for groups where it was debated whether a population was a sub-species or a separate species. For example the Macaroni Penguin (*Eudyptes chrysolophus*) was found to be indistinguishable from the Royal Penguin (*Eudyptes schlegeli*). The Royal Penguin is considered by some to be a sub-species of the Macaroni Penguin (Christidis and Boles, 2008), while others consider it a separate species (Baker et al., 2006). This is important because it means Macaroni and Royal penguins are likely to be very closely related and diverged relatively recently. It also means that some Royal Penguin barcode sequences could be labeled as belonging to the species *Eudyptes chrysolophus* if the worker considered them the same species. The same issue was found with the Greater Painted Snipe (*Rostratula benghalensis*) and the Australian Painted Snipe (*Rostratula australis*); the Australian Painted Snipe is sometimes considered a sub-species of the Greater Painted Snipe (Baker et al., 2007).

It was also found that there were many specimens belonging to the same species were labeled under different, synonymous names, often due to changing taxonomy. These were identified and corrected before analysis took place. This project also enabled some mislabeled and/or misidentified specimens to be identified. When a species' closest relative was found to be an apparently distantly related species, then the phylogenetic tree created from the sequences for this group could be examined. Often it was found that one sequence would be nestled amongst a clade of specimens belonging to a completely different species or genus, away from the rest of the specimens from that species. Approximately 7 suspicious results thought to be due to mislabeling/misidentification were found. A useful outcome of this research is that these instances can be reported to be investigated and corrected by the group who uploaded them into the database.

For the analysis undertaken, it is important that there is good sampling, so that the closest relative of the species of interest is present in the alignment used. To get a sense of how well sampled different groups are, for each species of interest the total number of extant species in the genus was found, as well as how many members of that genus are represented in the alignment. As expected, some genera were much better-represented than others. For example some genera within the passerines (Order Passeriformes) and the kingfishers (Order Coraciiformes) were very poorly sampled, while the penguins (Order Sphenisciformes) and shore-birds (Order Charadriiformes) were relatively well sampled.

## Conclusions

The majority of New Zealand birds surveyed (~82%) are genetically distinctive enough for identification of them to be reliable. A significant number were found genetically indistinguishable from their close relatives, however approximately a quarter of these are likely to be due to mislabeled and/or misidentified specimens in the reference database. There are also a relatively large number of species which are in a very poorly sampled group, and so their genetic distinctiveness is likely to be over-estimated due to their closest relative not being in the reference database. Also, 35% of New Zealand's bird species are represented by fewer than two sequences, and so cannot be used for this analysis. This number will reduce as the NZCOI Barcoding New Zealand Birds project continues to add more sequences. Further research in this area could include locating and correcting more sequences in the Barcode of Life database that are mislabeled/misidentified. A specific group that was found to include some genetically indistinguishable members (such as gulls or ducks) could also be investigated in more detail to determine the reasons why they are so genetically close.

## References

- Baker AJ, Pereira SL, Haddrath OP, Edge KA (2006). "Multiple gene evidence for expansion of extant penguins out of Antarctica due to global cooling". *Proc Biol Sci.* **273** (1582): 11–7.
- Baker, A.J., Pereira, S.L., Rogers, D.I., Elbourne, R. and Hassell, C.J. (2007). Mitochondrial-DNA evidence shows the Australian Painted Snipe is a full species, *Rostratula australis*. *Emu* **107**: 185-189.
- Christidis L, Boles WE (2008). *Systematics and Taxonomy of Australian Birds*. Canberra: CSIRO Publishing. p. 98.
- Ross, H.A., Murugan, S. and Li, W.L.S. (2008). Testing the reliability of genetic methods of species identification via simulation. *Systematic Biology* **57** (2): 213-230.
- Tubaro, P.L. and Lijtmaer, D.A. (2002). Hybridization patterns and the evolution of reproductive isolation in ducks. *Biological Journal of the Linnean Society* **77**: 193-200.