

Is a Large-Scale DNA-Based Inventory of Ancient Life Possible?

D. M. LAMBERT, A. BAKER, L. HUYNEN, O. HADDRATH, P. D. N. HEBERT, AND C. D. MILLAR

From the Allan Wilson Centre for Molecular Ecology and Evolution, Institute of Molecular BioSciences, Massey University, Private Bag 102 904, Auckland, New Zealand (Lambert and Huynen); Centre for Biodiversity and Conservation Biology, Royal Ontario Museum, 100 Queen's Park, Toronto ON, Canada M5S 2C6 and Department of Zoology, University of Toronto, Toronto ON, Canada M5S 1A1 (Baker and Haddrath); Department of Zoology, University of Guelph, Guelph ON, N1G 2W1, Canada (Hebert); and Allan Wilson Centre for Molecular Ecology and Evolution, School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand (Millar).

Address correspondence to David Lambert at the address above, or e-mail: d.m.lambert@massey.ac.nz

Abstract

A complete DNA-based inventory of the Earth's present biota using large-scale high-throughput DNA sequencing of signature region(s) (DNA barcoding) is an ambitious proposal rivaling the Human Genome Project. We examine whether this approach will also enable us to assess the past diversity of the earth's biota. To test this, we sequenced the 5' terminus of the mitochondrial cytochrome *c* oxidase I (COI) gene of individuals belonging to a group of extinct ratite birds, the moa of New Zealand. Moa comprised a large number of taxa that radiated in isolation on this oceanic landmass. Using a phylogenetic approach based on a large data set including protein coding and 12S DNA sequences as well as morphology, we now have precise information about the number of moa species that once existed. We show that each of the moa species detected using this extensive data set has a unique COI barcode(s) and that they all show low levels of within-species COI variation. Consequently, we conclude that COI sequences accurately identify the species discovered using the larger data set. Hence, more generally, this study suggests that DNA barcoding might also help us detect other extinct animal species and that a large-scale inventory of ancient life is possible.

The proposal to produce a DNA-based inventory of all extant species (www.barcodinglife.org) results from the pressing need to identify and catalog the current biota of the Earth (Blaxter 2003; Gould 2002; Wilson 2000). There is a broad acceptance that a new approach is required for estimating the biodiversity of species because of the current high rates of extinction and the slow rate at which we are presently able to describe them. It is clear that this task will require new DNA methodology, although the precise manner in which the technology will be employed has been widely debated.

In recent years Hebert and co-workers have argued strongly for a "DNA barcoding" approach to the construction of inventories of biodiversity. DNA barcoding involves sequencing of signature region(s) of the mitochondrial genome in order to catalog and identify species. In a recent study, Hebert et al. (2004a) used a single mtDNA gene, COI, to distinguish a large number of North American avian species. The COI gene is one of only two mitochondrial protein-coding genes found in all eukaryotes. COI possesses a very slow rate of amino acid change (Lynch and Jarrell 1993) and is highly conserved at the DNA sequence level, particularly within species. Using ~ 600 bp of sequence from

the 5' terminus of COI, Hebert et al. (2004a) were able to identify almost all of the 260 avian species investigated. They showed that the species studied were characterized by low levels of within-species COI variation (average 0.27%), with a maximum average within-species difference of 1.24%, together with generally high levels of between-species differences (congeneric average 7.93%). It remains a central, but empirical question to what extent DNA barcodes "calibrated" for any one group are applicable to others. Hebert et al. (2003a, b; 2004b) demonstrated that these levels of COI variation are also similar to those recorded for some insect species that in turn suggests that the approach is widely applicable across phylogenetically distant animal groups.

Moritz and Cicero (2004) expressed reservations about the DNA barcoding approach employed by Hebert and colleagues. They emphasized the need to distinguish between DNA barcoding as a molecular tool to identify individuals of already described taxa and DNA barcoding as a method to enhance the discovery of new species. They also raised a number of issues, including the need to compare sister taxa and not distantly related combinations of species, the need to

examine more than a single DNA sequence, and the need for precise data relating to geographic variation in COI sequences. Moritz and Cicero (2004) generally argued against the novelty and efficacy of DNA barcoding while accepting that in many cases “There is little doubt that large-scale and standardized sequencing, when integrated with existing taxonomic practice, can contribute significantly to the challenges of identifying individuals and increasing the rate of discovering biological diversity.” Certainly the general demonstration by Hebert et al. (2004a) that all the North American avian species examined had a different COI barcode(s) clearly shows the power of DNA barcoding to assigning individuals to previously described taxa. Hence, DNA barcoding as a method for the discovery of new species is at least possible.

The central issue is going to be how feasible this approach will be in broad terms. The use of multiple DNA barcodes will of course be required in some cases, but this will not detract from the general approach. An additional empirical question is whether this approach is possible in the case of ancient life.

The extinct moa of New Zealand represent an ideal test of the feasibility of DNA barcoding ancient faunas. Until recently, an accurate taxonomy of moa remained elusive with uncertainty about the number of species. For example, although 64 species names have been used in the past

(Worthy and Holdaway 2002), we now know that much of this overestimated was the result of high levels of sexual dimorphism (Cracraft 1976; Huynen et al. 2003). Not surprisingly, then, a large number of misidentified moa specimens still remain (Baker et al. 2004). This could potentially represent a significant problem, since it has been recently calculated that there were as many as 12 million individual moa in New Zealand (Gemmell et al. 2004)! Although this estimate of population sizes is likely to be inaccurate and is certainly based on a simplistic genetic analysis, it does suggest that large numbers of moa specimens might remain to be discovered. We have DNA barcoded the moa of New Zealand to test the feasibility of a DNA-based inventory of these ancient species. Using 26 subfossil moa bones, we examined COI variation from a variety of taxa and compared the number of species detected with the number of species identified using a more comprehensive data set (Baker et al. 2004).

Materials and Methods

The samples used in this work and their locations are detailed in Table 1 and Figure 1, respectively. Samples were obtained from the following museums: Auckland, Whanganui

Table 1. Moa samples used in this study.

Sample No.	Catalog No.	Species	Location	GenBank Accession No.
1	W 1152	<i>Pachyornis mappini</i>	Makirikiri	AY833111
2	MNZ 37844	<i>Pachyornis mappini</i>	Puketitiri	AY833109
3	MNZ 37845	<i>Pachyornis mappini</i>	Puketitiri	AY833110
4	CM SB301	<i>Pachyornis elephantopus</i>	Cheviot	AY833108
5	CM Av9209	<i>Pachyornis elephantopus</i>	Enfield	AY833107
6	CM Av8927	<i>Pachyornis elephantopus</i>	Hamiltons	AY833106
7	OM Av4139	<i>Pachyornis elephantopus</i>	Enfield	AY833105
8	AIM B6221	<i>Pachyornis australis</i>	Mt Arthur	AY833112
9	CM Av21331	<i>Pachyornis australis</i>	Takaka	AY833113
10	CM Av9243	<i>Euryapteryx geranoides</i>	Tom Bowling Bay	AY833114
11	AIM B6595ii	<i>Euryapteryx curtus</i>	Tokerau Beach	AY833118
12	CM Av8378	<i>Euryapteryx geranoides</i>	Pyramid Valley	AY833115
13	CM Av21330	<i>Euryapteryx geranoides</i>	Takaka	AY833119
14	CM Av9188	<i>Euryapteryx geranoides</i>	Kapua	AY833116
15	OM Av9821	<i>Euryapteryx geranoides</i>	Paerau	AY833117
16	CM Av8320	<i>Emeus crassus</i>	Pyramid Valley	AY833120
17	CM Av9132	<i>Emeus crassus</i>	Kapua	AY833121
18	Adid OH	<i>Anomalopteryx didiformis</i>	unknown	NC 002779
19	CM Av21547	<i>Anomalopteryx didiformis</i>	Mahoenui	AY833122
20	MNZ S34094	<i>Dinornis robustus</i>	Takaka	NC 002672
21	CM Av30497	<i>Dinornis robustus</i>	Culverden/Waikari	AY833123
22	CM Av9037	<i>Dinornis robustus</i>	Kapua	AY833124
23	CM Av20591	<i>Dinornis robustus</i>	Glendhu Bay	AY833125
24	CM Av30495	<i>Dinornis robustus</i>	Culverden/Waikari	AY833127
25	CM Av8579	<i>Dinornis robustus</i>	Roxburgh Gorge	AY833128
26	CM Av36435	<i>Dinornis robustus</i>	Karamea	AY833126
27	AIM B6310	<i>Dinornis novaezealandiae</i>	Waikaremoana	AY833129
28	OM Av10049	<i>Megalapteryx didinus</i>	Serpentine Range	AY833130

Samples are shown according to their taxonomic status (based on bone morphology and / or DNA analysis), museum catalogue number and the location where it was found. Abbreviations: AIM, Auckland Institute and Museum; CM, Canterbury Museum; MNZ, Museum of New Zealand Te Papa Tongarewa, Wellington; OM, Otago Museum; W, Whanganui Regional Museum; Sample numbers are as shown in Figure 2.



Figure 1. Details of the geographic distribution of moa samples used in this work.

Regional, Museum of New Zealand Te Papa Tongarewa (Wellington), Canterbury (Christchurch), and Otago (Dunedin). Femora were drilled using an 8-mm drill bit. The drill bit was sterilized in 5% (v/v) NaClO between uses, and initial shavings obtained from the bone surface were discarded. DNA was extracted from ~ 0.05 g of bone shavings by rotation overnight at 55°C in 0.8 ml of 0.5 M EDTA (pH 8.0) with ~ 2 mg proteinase K. The DNA was purified by phenol/phenol:chloroform:isoamyl alcohol/butan-2-ol extraction then concentrated and washed using Vivaspin concentrators (30,000 MWCO PES; Vivascience). All DNA amplifications were carried out in 10- μ l reactions containing 50 mM Tris-Cl, pH 9.0, 20 mM NH₄SO₄, 2.5 mM MgCl₂, 100 μ M of each dNTP, 0.3 U AmpliTaq DNA Polymerase (Perkin Elmer), 40 ng of each primer, and ~ 1–20 ng DNA. DNA extraction, primers, amplification conditions, and tree building methods used for the 2859 bp of moa mitochondrial sequence are described in Baker et al. (2004). Oligonucleotide primers

used for amplification of the mitochondrial COI locus are: mCOIF1b (5'-ACAGCCCTCAGCCTACTCAT-3')/mCOIF3 (5'-CCGATATAGCATTTCCAC-3'), mCOIF4b (5'-TCCATCCTAGGAGCTATCAA-3')/mCOIR2 (5'-CAGGGTGTCCGAAGAATC-3'), and mCOIR3b (5'-GGTAGGAGTCAGAAGCTTAT-3')/mCCIR4 (5'-ATGTTAATTGCTGTGGT-3'). These primers amplified three overlapping fragments to give a total of 596 bp of sequence from the 5' terminus of COI. All sequences were edited and aligned in Sequencher™ and sequence divergences were calculated in PAUP*4.0b10 using the Kimura 2-parameter algorithm.

Strict ancient DNA procedures were followed throughout this study involving independent replication in two physically separate laboratories: a dedicated ancient DNA laboratory at Massey University, Albany, and a similar facility at the University of Auckland. Sequences were obtained from multiple amplifications from individual samples.

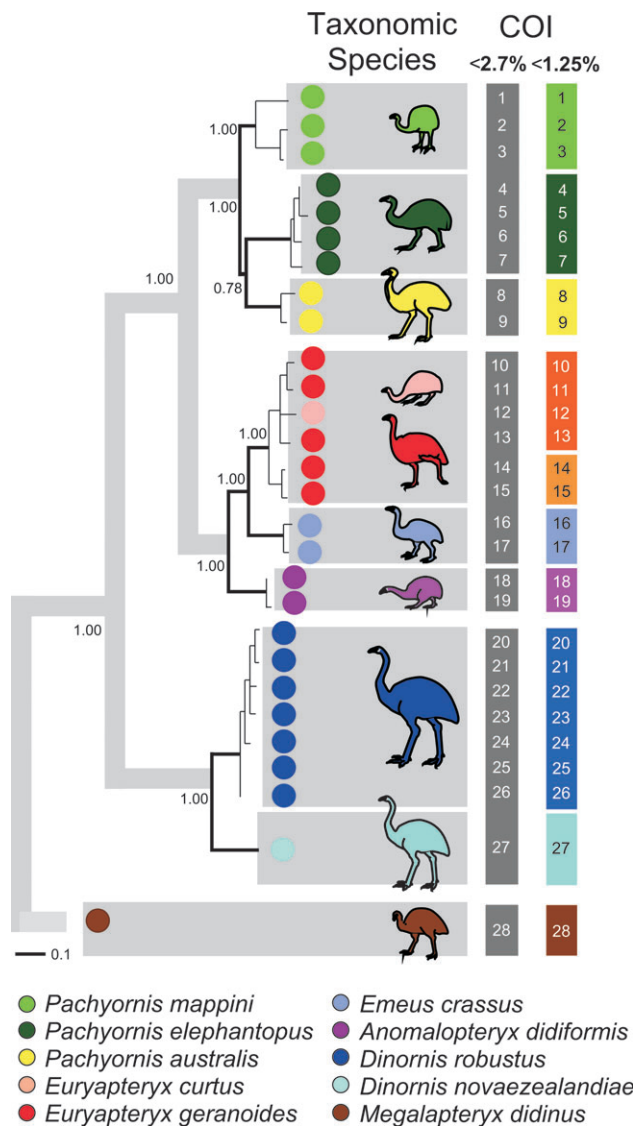


Figure 2. Testing the efficacy of DNA barcoding of ancient moa against a comprehensive phylogenetic tree. At left is a Bayesian partitioned likelihood tree constructed from control region data, together with partial sequences from seven mitochondrial protein coding regions and the ribosomal 12S gene (in total, 2859 bp). Bayesian probability values are shown. The suggested species limits are indicated by the light gray boxes. At right are the species limits detected by DNA barcoding using mitochondrial COI sequences (596 bp), and two intraspecific thresholds of genetic distance (<2.7% and <1.25%). Sample numbers refer to those shown in Table 1.

Results

Our aim was to compare the number of moa species detected using a large data set comprising DNA sequences from the mitochondrial control region and including seven partial protein-coding sequences and regions of the ribosomal 12S gene, with those detected using COI variation and divergence levels for the same moa specimens. We report that

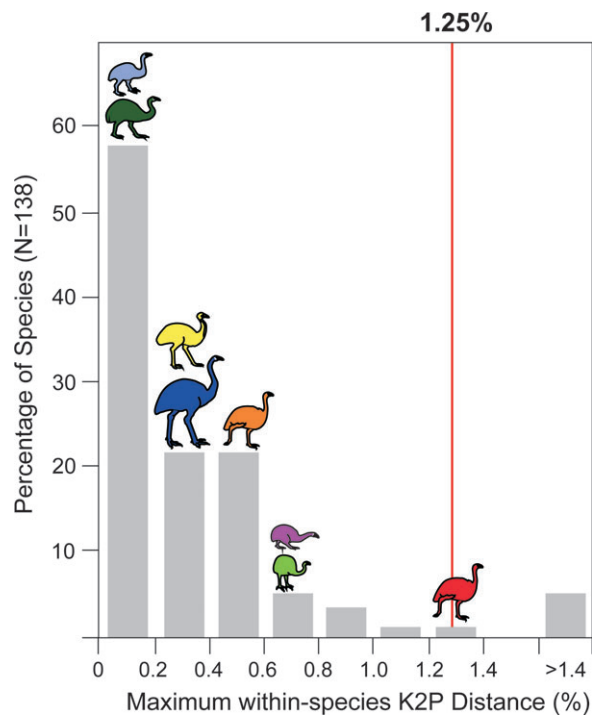


Figure 3. Maximum levels of intraspecific genetic distances between individuals belonging to 130 species of North American birds. Within-species values for moa are indicated. Colors refer to species moa as shown in Figure 2.

each species recognized by the comprehensive data set (Baker et al. 2004) had a unique COI barcode or barcodes. In addition, within-species COI sequence differences ranged from 0 to 1.24%. These data suggest that for moa, as a method of identification against known species, short COI sequences enable identification of unknown specimens, to the species level.

In relation to the discovery of new species using DNA barcoding based on calibrations from other avian groups, we performed a number of analyses. First, we used a standard screening threshold (Hebert et al. 2004a) of COI sequence difference of 2.7%. This value is 10 times the average within-species difference calibrated for North American birds and is therefore a very conservative estimate of species limits. Using this estimate as the upper boundary of within-species variation, six moa species were detected. These represent either single species or groups of related species that were distinguished using the larger data set (Figure 2). Relatively low between-species levels of divergence (range 1.3–3.8%, excluding the complex arrangement found in *Euryapteryx*) accompanied these comparisons.

A value of 1.24% was also used as a limit of allowable within-species variation in moa. This figure is less conservative and also derives from Hebert et al. (2004a). It represents the maximum average level of within-species difference detected for nearly all of the 130 North American avian species sampled (Figure 3). Not surprisingly, this value

resulted in an increased number of moa species detected, compared to the result when 2.7% was used. However, importantly, the 10 moa species identified in this way corresponded to those known from the larger study, with one exception (Figure 2). The latter analysis suggested a possible complex of species in the genus *Euryapteryx*. This genus is taxonomically complex in that the species *Euryapteryx curtus*, currently recognized, is now embedded within *Euryapteryx geranoides*.

Discussion

Our results show that for a moa extinct group of very closely related species, distinct COI barcodes exist for the taxa recognized from a larger, more comprehensive analysis. This gives confidence that a COI-based species identification system would be useful in identifying the large number of unknown or misidentified moa specimens in museums and other collections. This approach, properly applied, will therefore enhance the value of these collections and allow more detailed geographical and morphological analyses. For example, our knowledge of the geographic distributions of moa species is tentative, and more precise estimates are required. COI barcodes of a large number of moa specimens will inevitably improve these estimates.

More generally, the success in moa species identification using COI DNA barcodes calibrated on North American avian species is encouraging. From a practical perspective this is significant because it demonstrates that such studies of ancient life are possible. Furthermore, moa are a highly speciose and closely related group that is phylogenetically distant from the North American species on which the species-level calibrations were performed. Hence, this suggests that most avian groups will be amenable to COI DNA barcoding using the same calibrations. Moritz and Cicero (2004) have criticized these calibrations. They argued that a true test of the precision of mtDNA barcodes to assign individuals to species would include comparisons with sister species—"the most closely related extant relatives." They have suggested that such comparisons were not included in those made by Hebert et al. (2004a). However, the calibrated DNA barcodes for North American birds accurately detect ancient species of moa. This group is restricted to a defined geographic region and member species are closely related. Hence, for moa at least, the Moritz and Cicero (2004) criticism does not appear to be valid. Moritz and Cicero (2004) also suggested that other avian data are inconsistent with Hebert et al.'s (2004a) general interpretation. For example, they refer to Johnson and Cicero's (2004) finding that nearly 74% of sister species comparisons differed by less than the 2.7% difference threshold suggested by Hebert et al.'s (2004a). On the surface, this appears to be a serious criticism of the DNA barcoding proposition. However, these comparisons are invalid because Johnson and Cicero's (2004) results—on which Moritz and Cicero's (2004) claims are based—mainly draw on data derived from a large number of studies using different methods (including RFLP) and

regions other than COI. In fact, of the 39 species comparisons reported by Johnson and Cicero (2004), only 8 actually use COI data to arrive at their conclusions. Of particular concern is the suggestion that Johnson and Cicero (2004) have detected species with identical DNA barcodes, because they report an example of distinct species with no sequence divergence. In reference to that sample, Table 1 from Johnson and Cicero (2004) refers to an unpublished 723-bp sequence of ND6. However, to date, ND6 has not been suggested as a likely candidate for DNA barcoding, and no species-level calibrations are available.

Although the DNA barcoding approach is focused on extant biota, the number and characteristics of the species that once inhabited the Earth is less well known. Recent advances in DNA biotechnology offer a platform on which a DNA-based taxonomy of ancient forms might be possible. These advances include the successful recovery of mitochondrial DNA sequences from a large number of species of extinct animals (Haddrath and Baker 2001; Huynen et al. 2003; Lambert et al. 2002), as well as a variety of tissues (e.g., hairs). Even the routine amplification of short single locus nuclear DNA sequences from similar ancient material is now possible (Huynen et al. 2003). The latter is remarkable given that only recently, this was argued to be impossible owing to the low copy number of target nuclear DNA sequences.

Our success in amplifying this COI region from a large number of sub-fossil moa, in combination with our finding of unique COI barcodes offers the prospect of being able to identify, to the species level, not only the many moa specimens currently stored in museums, but also many of the other sub-fossil remains of extinct animals. Recent developments in high throughput sequencing, combined with new ancient DNA technologies, suggests that a DNA-based inventory, in concert with the "Tree of Life" approach, seem destined to result in a new era upon us in the study of biodiversity, past and present.

Acknowledgments

We thank the Marsden (97-MAU-LFS-0025) and the Centres of Research Excellence Funds for financial support. We are grateful to the following institutions and people: Museum of New Zealand Te Papa Tongarewa, Canterbury Museum, Auckland Institute and Museum, Otago Museum, Whanganui Regional Museum, Institute of Geological and Nuclear Sciences, J. Anderson, J. A. Bartle, B. Gill, A. Tennyson, T. Worthy, and V. Ward. We appreciate the support of a number of iwi, especially Ngai Tahu and Ngati Kahungunu. This paper is based on a presentation given at the symposium entitled "Genomes and Evolution 2004," cosponsored by the American Genetic Association and International Society of Molecular Biology and Evolution, at the Pennsylvania State University, State College, PA, USA, June 17–20, 2004.

References

- Baker AJ, Huynen L, Haddrath OP, Millar CD, and Lambert DM, 2004. Ancient DNA clarifies taxonomy and evolutionary history of the giant extinct moa of New Zealand. Submitted for publication.
- Blaxter M, 2003. Counting angels with DNA. *Nature* 421:122–124.

- Cracraft J, 1976. The species of moas. *Smithsonian Contrib Paleobiol* 27:189–205.
- Gemmell NJ, Schwartz MK, and Robertson BC, 2004. Moa were many. *Biol Lett* 271:S430–S432.
- Gould SJ, 2002. *The structure of evolutionary theory*. Cambridge, MA: Harvard University Press.
- Haddrath O and Baker AJ, 2001. Complete mitochondrial DNA genome sequences of extinct birds; ratite phylogenetics and the vicariance biogeography hypothesis. *Proc R Soc Lond B* 268:939–945.
- Hebert PDN, Cywinska A, Ball SL, and deWaard JR, 2003a. Biological identifications through DNA barcodes. *Proc R Soc Lond B* 270:313–321.
- Hebert PDN, Ratnasingham S, and deWaard JR, 2003b. Barcoding animal life: cytochrome c oxidase subunit 1 divergences amongst closely related species. *Proc R Soc Lond B (Suppl.)* 270:S96–S99.
- Hebert PDN, Stoeckle MY, Zemplak TS, and Francis C, 2004a. Identification of birds through DNA barcodes. *PLoS Biology* 10:1657–1663.
- Hebert PDN, Penton EH, Burns JM, Janzen DH, and Hallwachs W, 2004b. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc Natl Acad Sci* 101:14812–14817.
- Huynen L, Millar CD, Scofield RP, and Lambert DM, 2003. Nuclear DNA sequences detect species limits in ancient moa. *Nature* 425:175–178.
- Johnson NK and Cicero C, 2004. New mitochondrial DNA data affirm the importance of Pleistocene speciation in North American birds. *Evolution* 58:1122–1130.
- Lambert DM, Ritchie PA, Millar CD, Holland B, Drummond AJ, and Baroni C, 2002. Rates of evolution in ancient DNA from Adelie penguins. *Science* 295:2270–2273.
- Lynch M and Jarrell PE, 1993. A method for calibrating molecular clocks and its application to animal mitochondrial DNA. *Genetics* 135:1197–1208.
- Moritz C and Cicero C, 2004. DNA barcoding: promise and pitfalls. *PLoS Biology* 2:1529–1531.
- Wilson EO, 2000. On the future of conservation biology. *Conservation Biology* 14:1–3.
- Worthy TH and Holdaway RN, 2002. *The lost world of the moa*. Christchurch, New Zealand: Canterbury University Press.

Corresponding Editor: Shozo Yokoyama