Genetic variation within and among populations of North Island kokako

SCIENCE & RESEARCH INTERNAL REPORT 176

Michael Double and Steve Murphy

Published by
Department of Conservation
P.O. Box 10-420
Wellington, New Zealand

Science & Research Internal Reports are written by DOC staff or contract science providers on matters which are on-going within the Department. They include reports on conferences, workshops, and study tours, and also work in progress. *Internal Reports* are not normally subject to peer review. Publication was approved by the Manager, Science & Research Unit, Science Technology and Information Services, Department of Conservation, Wellington.
This report was prepared by Michael Double and Steve Murphy, Division of Botany and Zoology, Australian National University, Canberra ACT 0200. E-mail Mike.Double@anu.edu.au.

ISSN 0114–2798
ISBN 0–478–21954–7

Cataloguing-in-Publication data

Double, Michael
Genetic variation within and among populations of North Island kokako / Michael Double and Steve Murphy. Wellington, N.Z. : Dept. of Conservation, 2000.
1 v. ; 30 cm. (Science & Research internal report, 0114-2798 ; 176)
Includes bibliographical references.
ISBN 0478219547
1. Callaeas cinerea wilsoni—Research—New Zealand.
I. Murphy, Steve. II. Title. Series: Science and Research internal report ; 176.

CONTENTS

Abstract

Genetic relationships among extant populations of North Island kokako were investigated by sequencing the 806 bases of the mitochondrial control region for 23 kokako from ten of the remaining 16 populations. There was very little sequence variation between individuals; only 1.9% of the 806 bases were variable. The sequence variants (haplotypes) were often found in more than one population and there was no apparent geographic clustering of haplotypes. This result indicated that the contemporary kokako populations are not genetically distinct. We therefore suggest that either (1) until relatively recently gene flow occurred between the kokako populations or (2) all current populations are derived from a population that dispersed from a single refuge following the last Pleistocene glaciation. Multilocus DNA fingerprinting was used to address whether the genetic variability within the recently bottlenecked Mapara population of North Island kokako was lower than that found in the larger Te Urewera kokako population. This technique allows the variation across multiple 'minisatellite' loci to be quantified in a single step. By contrasting the DNA fingerprints from multiple individuals it is possible to quantify the genetic variation found within and between populations. We DNA fingerprinted twelve and eight individuals from Mapara and Te Urewera respectively. These fingerprints indicate that the Mapara kokako population has not lost a significant amount of genetic variation when compared with Te Urewera. This result concurs with our theoretical analysis that predicted the bottleneck at Mapara was neither long nor severe enough to significantly reduce the genetic variability within this population.

1. Determining genetic relationships among extant populations of North Island kokako

1.1 INTRODUCTION

Ancestors of kokako *Callaeas cinerea* are thought to have arrived in New Zealand from Australia about 65 million years ago (Stevens et al. 1988; Williams 1976). If kokako have always depended on forest habitat, then their populations will have expanded and contracted in concert with the distribution of forests. Three main factors have greatly influenced the distribution of forests in New Zealand: tectonic activity, glaciation and human activities.

Approximately 2.4 million years ago an extensive period of global cooling began. Twenty glacial phases occurred in New Zealand between this time and 14 000 years ago (the Pleistocene epoch; Fleischer et al. 1994; Fleming 1979).

During each glacial phase, water was locked up in ice sheets and thickening polar ice caps, which caused sea levels to fall by up to 165 m (Stevens et al. 1988). A land bridge formed between the North and South Islands, and offshore islands were joined to the mainland. Falling sea levels may have allowed many terrestrial taxa, including kokako to move between New Zealand's islands (Stevens et al. 1988). Gymnosperm-hardwood forests were probably confined to moist frost-free sites, possibly in parts of the continental shelf exposed by the falling sea levels. Only the northern one-third of the North Island was likely to have had extensive forested habitat (Fleming 1979; McGlone 1985; Fig. 1).

About 10 000 years ago New Zealand's climate began to stabilise and became warmer. This allowed gymnosperm-hardwood forests to expand. Around 1500 years ago such forest is thought to have covered between 75–80% of New Zealand (Stevens et al. 1988). The distribution of these forests was again radically altered after the arrival of humans in New Zealand approximately 1000 years ago. Within the last few centuries much of the lowland forests have been cleared or burnt. Between 1847 and 1923, 6.5 million ha (57%) of the North Island's lowland forests were cleared (Lavers 1978). An assemblage of mammals were also introduced by human colonisers including rats, mustelids, and possums; species which have since been implicated in the decline of many terrestrial species including kokako (King 1984; Leathwick et al. 1983).

1.1.1 Phylogeography of a species

A phylogeographic study draws on geological and geographical information such as that described above and then adds genetic data to build a picture of a population's history and its evolutionary relationship with other populations of the same species. For example, Baker et al. (1995) uncovered extreme genetic

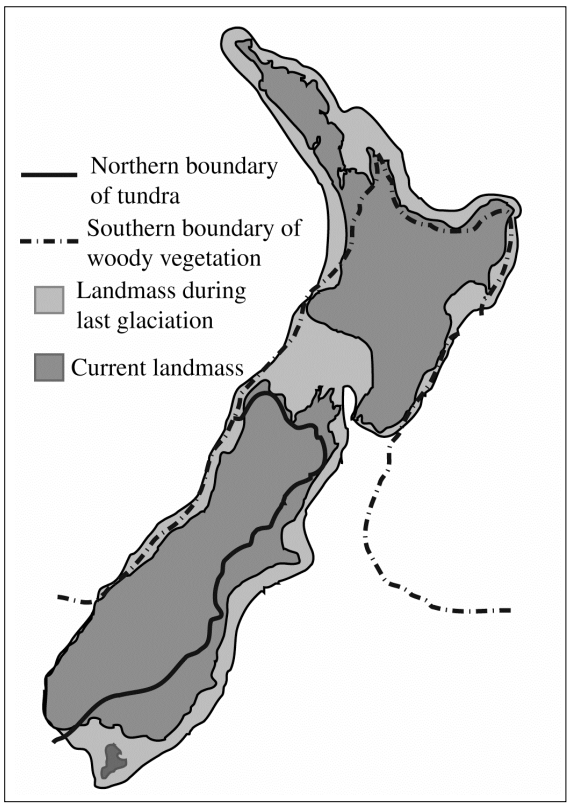


Figure 1. New Zealand during the last glaciation (about 14 000 years ago). Based on a map by M. Gage and R.P. Suggate, subsequently reproduced by Fleming (1972). Extensive habitat suitable for kokako was only likely to have occurred above the black dotted line. From evidence for scrub refugia, Fleming (1962) suggests the boundary for woody vegetation is too far north.

divergence among contemporary populations of the brown kiwi *Apteryx australis* after sequencing regions of mitochondrial DNA. Baker et al. (1995) suggested that the populations were isolated during the glaciations of the Pleistocene epoch and consequently diverged genetically. Following the glacial period there was little gene flow between the remnant populations; perhaps because of the poor dispersal capabilities of the kiwi or due to a reproductive barrier established through genetic or behavioural divergence. The factors that created such divergence in kiwi populations could also have had a similar impact on other species such as kokako.

In this study we investigated the genetic relationship among extant populations of kokako and attempted to explain the results given the geological and geographic factors which will have influenced these populations in the past.

1.2 METHODS

1.2.1 Kokako

Kokako belong to the New Zealand wattlebird family Callaeidae. This family also includes the tieke (saddleback; *Philesturnus carunculatus*) and huia (*Heteralocha acutirostris*). There are two subspecies of kokako, the South Island kokako (*Callaeas cinerea cinerea*) with orange wattles and the blue-wattled North Island kokako (*C. c. wilsoni*).

Populations of North Island kokako are currently confined to mixed gymnosperm-hardwood forest between 400–600 m above sea level. Generally they exist in small, declining, fragmented populations totalling approximately 1000 individuals (Fig. 2). The International Union for the Conservation of Nature has classified the North Island kokako as endangered (Baillie et al. 1996). The South Island kokako is now thought to be extinct (Falla et al. 1981).

1.2.2 Sampling

The origins of samples used in this study are shown in Fig. 2. The South Island kokako tissue sample was taken from a skin at the National Museum of New Zealand, Wellington. The kokako skin was originally collected from Milford Sound by Reischek in 1888. Tieke DNA was provided by D. Lambert and T. King, Massey University, Palmerston North.

1.2.3 DNA extraction, amplification and sequencing

See Murphy (1998) for detailed account of methodologies.

DNA was extracted using either a standard phenol:chloroform extraction protocol (Sambrook et al. 1990) or a QIAamp Blood and Tissue Kit (Qiagen). DNA samples were then used in Polymerase Chain Reactions (PCR) to specifically amplify two regions of kokako mitochondrial DNA; cytochrome *b* and the control region.

The cytochrome *b* region was amplified using primers developed by Crozier et al. (1991) and Palumbi et al. (1991; CB1 & CB3-H). Primers used to amplify the control region (H1248 and L437) were developed by Tarr (1995). A third PCR primer (L-DomIII; 5' – CTC ACA CTT TGC CCT GAT GC – 3') was designed *de*

novo and in conjunction with L437 specifically amplifies a small, but variable part of the control region (Domain III (see Mindell 1997)). This primer was also used to sequence the larger fragment generated by the H1248 and L437 primer combination.

PCR products were run in, and subsequently excised from, a 2% agarose gel. The products were then extracted and cleaned using the Bresaclean DNA Purification Kit (Bresatec). The sequencing reaction used Big-Dye Terminator Sequencing (Perkin-Elmer) and the products run on a ABI Prism 377 automated sequencer (Perkin-Elmer).

1.2.4 Data analysis

DNA sequences generated using the protocols outlined above were checked for possible base calling errors in the DNA analysis software Sequencher (Gene Codes Corporation). Sequences were then aligned using Clustal W version 1.6 (Thompson et al. 1994).

Principal Coordinates Analysis

A Principal Coordinates Analysis (PCA) was used to test if there was any underlying pattern to the observed genetic variation by reducing the hyperdimensional nature of the pairwise genetic distance matrix into fewer dimensions (Gnanadesikan 1997). The PCA was performed using the computer software package GenAlEx (Peakall & Smouse 1998). The PCA used pairwise genetic distances among haplotypes to track the distribution of individuals with the same or similar haplotypes.

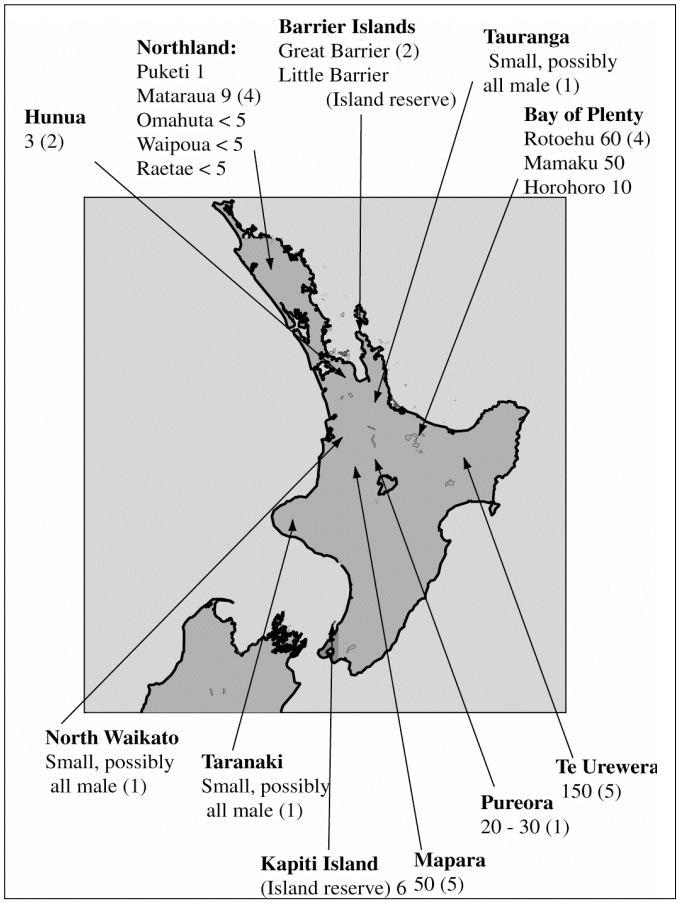


Figure 2. Contemporary North Island kokako populations. Numbers indicate the population size in breeding pairs. Numbers within parentheses indicate the sample size for this study.

Tree-based analyses

Using the software package PAUP* v4.0.0d64 (Swofford 2000), two tree-based analytical methods were used to graphically represent genetic distances among haplotypes and to track their evolutionary origin. In the first analysis, a heuristic search of the data with ten random starts using a tree-bisection-reconnection (TBR) algorithm allowed a parsimony tree to be constructed. Support for interior branches was non-parametrically tested using 100 bootstrap replicates (Felsenstein 1985). The bootstrapping technique involves randomly sampling the dataset with replacement to make new datasets the same size as the original. Trees are then formed from the new datasets and the number of times nodes or groups in the observed tree appear in the simulated trees are counted. The Felsenstein bootstrap proportion cannot be interpreted directly as a confidence limit so often a 'bootstrap majority rule' is adopted. Thus clades which appear in over 50% of the trees generated from bootstrapped data are considered well supported.

The second tree-based analysis involved the construction of neighbour-joining trees based on two alternative evolutionary models. The uncorrected P model simply considered the percentage of nucleotide site differences between individual haplotypes. In contrast, the HKY85 model allowed for unequal base frequencies among taxa, and weighted transversions (purine ↔ pyrimidine) over transitions (purine ↔ purine; pyrimidine ↔ pyrimidine) (Hasegawa et al. 1985). In a neighbour-joining tree the genetic distance among taxa is proportional to branch length (Swofford et al. 1996).

1.3 RESULTS

1.3.1 Cytochrome b sequence variation

North Island kokako and tieke cytochrome *b* sequences differed at 11% of sites (36/323 bases; see Murphy 1998). This region of cytochrome *b* was sequenced for three North Island kokako from geographically distant populations: Northland, Great Barrier Island and Mapara. Only one of 323 bases was found to be variable (0.3% variation) among these individuals. DNA sequence with such low variation contains little or no phylogenetic information and thus data from cytochrome *b* was not used in later analyses.

1.3.2 Control region sequence variation

Tieke versus North Island kokako

Individual North Island kokako and tieke were found to differ at approximately 106 of 806 bases (13.2% variation) sequenced. The majority of variable sites were in domain III, a pattern consistent with that observed for this region in other avian taxa (Baker & Marshall 1997; Wenink et al. 1994).

North versus South Island kokako

The DNA extracted from the South Island kokako tissue was highly degraded and therefore only the short (406-bp) fragment of domain III could be sequenced. Approximately 5.5% of sites (23/406) were variable between South Island and North Island kokako (Table 1).

TABLE 1. SEQUENCE VARIATION WITHIN DOMAIN III OF THE CONTROL REGION FOR NORTH AND SOUTH ISLAND KOKAKO.

HAPLOTYPE	HAPLOTYPE SEQUENCE	POPULATION (NUMBER OF INDIVIDUALS)
A	CACACACAAACG----------CGTAACAAACAAATCAGTCG	Rotoehu (1)
B	..........----------.............A...	Te Urewera (4) Pureora (1)
C	..........A----------.............A...	Rotoehu (2)
D	..........----------..G.............A...	Tauranga (1)
E	..........----------.........G.....A...	Rotoehu (1)
F	....T.....----------.............A...	Great Barrier Island (2)
G	..........----------........--......A...	Taranaki (1)
H	..........----------..........--...A...	Mataraua (2)
I	..........TCATTTTATCA.............A..A	Te Urewera (1) Mapara (1) Hunua (1)
J	..........TCATTTTATCA......G.........A..A	Mapara (4) Hunua (1) North Waikato (1)
K	..........T.TCATTTTATCA.............A..A	Mataraua (2)
South Island kokako	TGTG.GTGGG..----------AAC.-T.G.....CTG.CCG	Milford (1)

Within North Island kokako

A total of 806 bases of the control region were sequenced for 23 North Island kokako. Only 1.9% of these 806 bases was found to be variable. The DNA from three other individuals was degraded so only domain III could be sequenced (Table 1).

Haplotypes (sequence variants) were defined on the basis of differences generated by insertion or deletion of bases (indels) and simple base substitution. Within North Island kokako a total of 12 haplotypes were defined by such variable sites within domains II and III. Eleven were defined by domain III alone (Table 1). As a larger dataset was available for domain III (N = 26), the phylogeographic analysis was based on this region only. The geographic locations of haplotypes are shown in Fig. 3. Five of the ten populations sampled were represented by unique haplotypes. The 11-bp insertion was found in five populations: Mataraua, Mapara, Hunua, Te Urewera and North Waikato (Fig. 3).

1.3.3 Geographic distribution of variation within the control region of North Island kokako

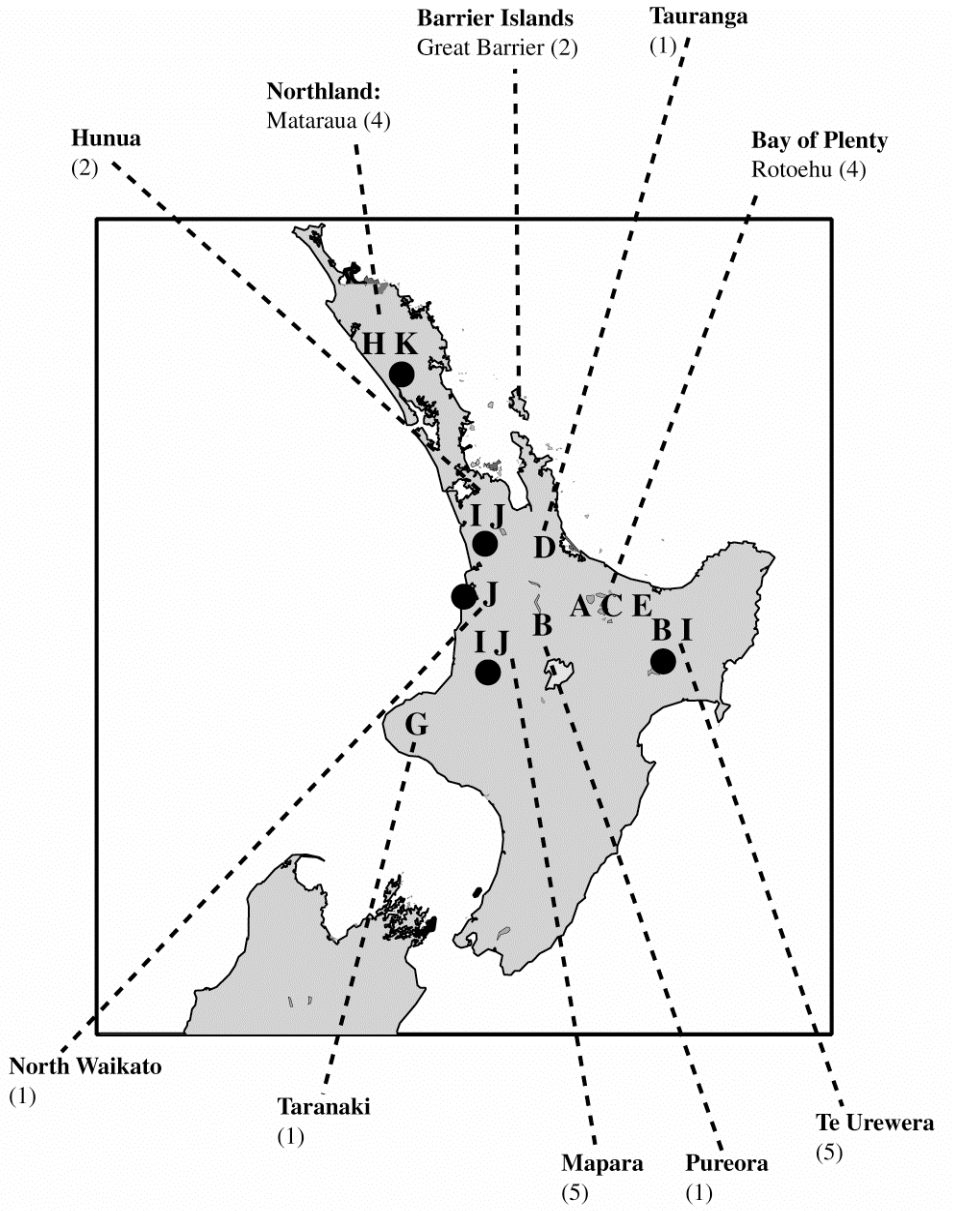
Principal Coordinates Analysis

The PCA of genetic distance among haplotypes showed some clustering of haplotypes, but clustered haplotypes were not necessarily from the same population or neighbouring populations (Fig. 4). For example, the large cluster in the bottom left of Fig. 4 includes haplotypes from populations up to 400 km apart (Taranaki and Mataraua).

Tree-based analyses

For a detailed description of tree analyses see Murphy (1998).

Figure 3. Geographical distribution of haplotypes. Numbers in parentheses are sample sizes. Letters refer to haplotypes described in Table 1. Black circles indicate populations that have at least one haplotype with the eleven base-pair insertion.



The neighbour-joining tree shows that the genetic distances among North Island kokako mtDNA control region lineages was low, with HKY85 distances ranging from 0.28% to 0.57%. By contrast, distances between North Island kokako and tieke ranged from 15.9% to 16.4%. and distances between North Island kokako and South Island kokako ranged from 4.6% to 4.9% (Fig. 5).

Most parsimonious trees were created using the uncorrected P and HKY85 evolutionary models. Parsimony in this context means that the 'simplest explanation consistent with the dataset will be chosen over more complex explanations' (Stewart 1993). These trees had identical topology to the neighbour-joining tree (Fig. 5). All observed ts/tv ratios exceeded the expected ts/tv ratio at saturation which indicated that homoplasy (convergent sequence evolution) was not likely to confuse the phylogenetic signal. Base frequencies among taxa were very similar which suggests that the relative probabilities of base substitutions among taxa were equal. Therefore, there was little bias in the estimated genetic distances among taxa.

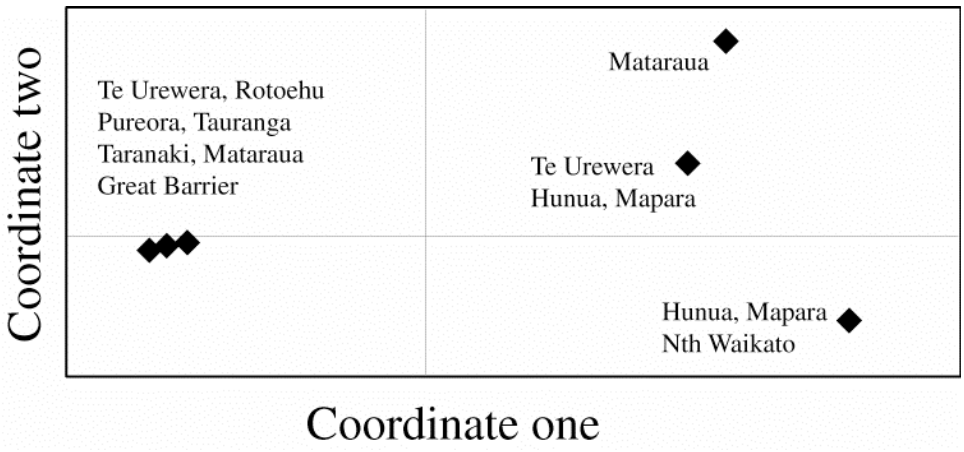


Figure 4. Principal Coordinates Analysis (PCA) based on pairwise genetic distances among haplotypes. Co-ordinates one and two were the dimensions which captured the most variation in the underlying distribution of genetic distances.

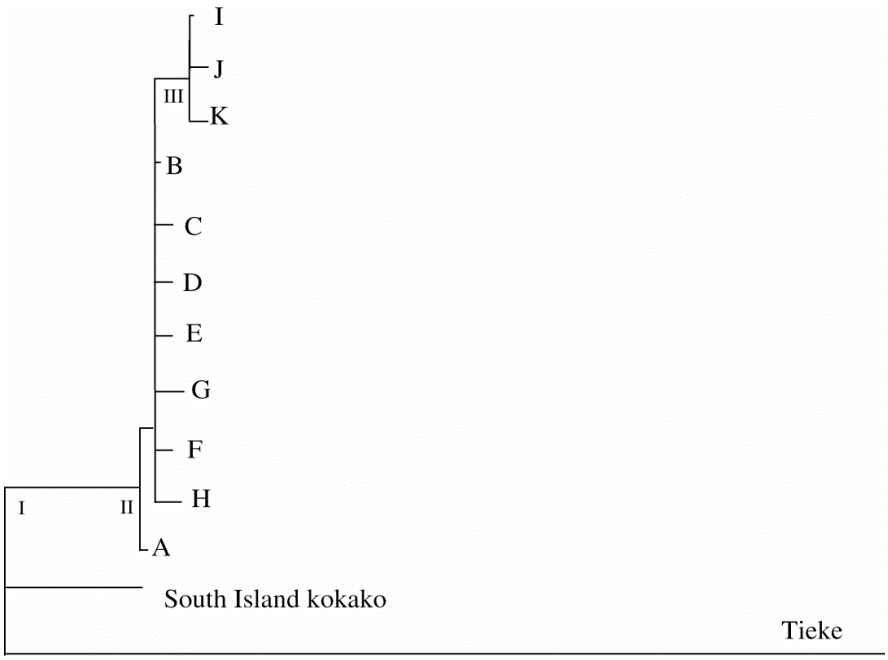


Figure 5. Neighbour-joining tree based on sequences differences among North Island kokako Domain III haplotypes. The tieke (saddleback) and South Island kokako Domain III sequences were defined as outgroups. Letters correspond to haplotypes described in Table 1. Genetic distances among haplotypes are proportional to the branch lengths. A parsimony analysis gave a tree with identical topology to the one presented in which the boostrap values for nodes I, II and III were 96, 64 and 72 respectively.

The lack of geographical structuring observed in the PCA was also reflected in the tree-based analyses. There was significant support for the clade which included haplotypes I, J and K (bootstrap = 72/100). These lineages were discovered in Te Urewera, Mapara, Hunua, North Pureora, North Waikato and Mataraua and all have the 11-bp insertion (Table 1; Fig. 3). When the insertion was not included in the parsimony-based heuristic search, the same topology was supported. This indicated that the geographically unstructured pattern of genetic variation is supported by two different types of mutations: base substitutions and insertions.

1.3.4 Summary of results

There are five main points that can be drawn from these results:

1. Control region sequence data suggest that North and South Island kokako are genetically distinct (5.5% divergence);
2. There was very little variation (0.3 %) among North Island kokako within cytochrome *b*;
3. There were low levels of variation (1.9 %) among North Island kokako within the control region;
4. The PCA analyses showed that haplotypes from the same or adjacent populations do not necessarily cluster together, suggesting no apparent geographical structure; and
5. The tree-based analyses also showed no geographical structure in the genetic variation, and is supported by two different types of mutation (base substitutions and insertions/deletions).

1.4 DISCUSSION

The aim of this study was to determine the genetic relationship among extant populations of North Island kokako. A Principal Coordinates Analysis showed that although haplotypes clustered into four groups, these groups were comprised of individuals from multiple, geographically distant populations. The tree-based analyses supported this result. Haplotypes that clustered within the tree (clades) were not necessarily from the same or geographically adjacent populations. Thus, there is no apparent geographic structure in the distribution of North Island kokako lineages.

The low sample sizes for some populations have little impact on this general conclusion. Although some populations show an apparently higher level of haplotypic diversity, all North Island kokako haplotypes are very similar and many populations share haplotypes. This result implies little genetic differentiation between populations and generally low variation within North Island kokako mitochondrial DNA. This marker system was used to reveal relationships between populations and the data generated cannot be used to assess genetic variation within and between populations. To do this a more variable marker system and larger sample sizes would be required (see Section 2).

The relatively low level of genetic variation and the lack of geographic structuring suggest that North Island kokako may have undergone a bottleneck and that contemporary populations of North Island kokako radiated from a single Pleistocene refuge. If other refugia contributed to the current North Island kokako gene pool then one would expect a greater amount of genetic divergence. Of course, distinct lineages may have been present in populations not sampled in this study. An alternative explanation for the low level of genetic diversity is that a selectively advantageous form of mitochondrial DNA arose which has subsequently replaced inferior forms. This process is known as a selective sweep. The mitochondrial genome is contained within a single linkage group, therefore if a single gene confers an advantage then the genes on the rest of the genome will also increase in frequency within a population. To exclude this possibility it would be necessary to assess the genetic diversity within

nuclear genes. If these too show low levels of variation then the bottleneck scenario would be more likely than a selective sweep. Whichever process generated the similarities between contemporary populations of kokako described here, both would result in populations of low genetic diversity and minimal genetic differentiation.

The suggestion that contemporary North Island kokako radiated from a single refuge does not exclude the possibility that kokako survived in refugia elsewhere. Indeed, the relatively large differences between South Island kokako sequence data and the North Island haplotypes are most likely due to isolation in separate Pleistocene refugia. In a study of lesser snow geese, Quinn (1992), found a similar level of control region variation *within* a population (6.7%) as described here between North and South Island kokako. However, each goose haplotype fell into one of two distinct clades and Quinn argues that the high level of variation was generated by a vicariant origin of the population i.e. populations that differentiated due to isolation during the Pleistocene subsequently coalesced forming a single population of high genetic variation. Only the analysis of more samples from the South Island will reveal if South Island kokako were a coalescence of lineages from multiple refugia (perhaps including North Island lineages) or a race of kokako genetically distinct from their North Island cousins.

2. Determining the remaining level of genetic variability within the population of kokako at Mapara

The kokako population at Mapara is an example of one that has been through a recent bottleneck. In this section genetic variation of Mapara kokako is contrasted with that of the larger Te Urewera population.

2.1 INTRODUCTION

Genetic variation will be lost from a small population if its numbers remain low over multiple generations (Hartl & Clark 1989). Such an event could occur when a few individuals colonise new territory (founder effect), or when a population decreases dramatically in size (bottleneck). Whatever the cause, two processes act in concert to erode genetic variation: genetic drift and inbreeding (Hartl & Clark 1989). Genetic drift is a random sampling effect whereby only a subset of the variation that exists in one generation is passed on to the next. Inbreeding will also occur if bird numbers remain low and they produce offspring that contain genetic material that is identical by descent.

A number of studies now report the reduction of genetic variation due to population bottlenecks (e.g. Gottelli et al. 1994; Taylor et al. 1994). For example, the population of black robins *Petroica traversi* on Chatham Island crashed to only five individuals during the late 1970s and early 1980s (Arden & Lambert 1997). The extant population now numbers 200 individuals, all thought to be derived from one female. By comparing the minisatellite DNA profiles (see Methods) of black robins with those of a sister species (*Petroica australis australis)*, it was shown that this severe bottleneck reduced average heterozygosity by approximately 35%, and the proportion of alleles represented in every individual assayed increased from an estimated 0.024 to 0.375 (Arden & Lambert 1997).

The specific aim of this study was to determine if the population of North Island kokako at Mapara had lost genetic variation due to a suspected bottleneck. To do this we firstly used demographic information from Mapara to predict the loss of genetic variation theoretically. Secondly, we directly quantified the genetic variation within kokako from both the Mapara and Te Urewera populations. These data were then used to assess the severity of any bottleneck and the consequential loss of genetic diversity.

2.1.1 The North Island kokako population at Mapara

The Mapara population of North Island kokako inhabits a 1500 ha area of contiguous mixed gymnosperm-hardwood forest, occurring 300–500 m above sea level. In 1978 a New Zealand Wildlife Service report found the highest densities of North Island kokako at Mapara (Coker 1978). However, by 1989 the population seemed to have crashed (Ian Flux, Department of Conservation, pers. comm.). Censuses conducted by Department of Conservation (DOC) field staff revealed the number of breeding pairs was four or five in 1989/90, between five and seven pairs in 1990/91, and five or six in 1991/92. Also during this period the population was thought to be extremely male biased, probably because of predation of females during incubation. In 1989/90 DOC started a program of mammalian predator control at Mapara, resulting in an initially slow but definite recovery. The census size in 1996/97 was 22 females from a total population size of 86 individuals (Table 2).

TABLE 2. DEMOGRAPHIC DATA OF NORTH ISLAND KOKAKO AT MAPARA (I. FLUX PERS. COMM.).

BREEDING SEASON	TOTAL CENSUS SIZE	NUMBER OF BREEDING FEMALES
pre-1989	unknown	unknown
1989/90	52	5
1990/91	48	5
1991/92	47	~4
1992/93	52	~10
1993/1994	59	~10
1994/1995	64	18
1995/1996	86	20
1996/1997	86	22

2.1.2 The comparative population: Te Urewera

Studies which attempt to determine the effect of population bottlenecks on genetic variation ideally quantify the genetic variation before and after the event. Unfortunately this is not usually possible so as an alternative the genetic variation in a separate population that has not gone through a bottleneck is quantified e.g. (Sherwin et al. 1991; Taylor et al. 1994).

In this study the relatively large population of North Island kokako at Te Urewera was used as a comparison to the potentially 'bottlenecked' Mapara population. Te Urewera forest has a total area of about 50 000 hectares. About 150 breeding pairs of North Island kokako are thought to inhabit about half of this area (Ian Flux, pers. comm.). For many years, Te Urewera was thought to be a stronghold for North Island kokako, so the effects of genetic drift and inbreeding are thought to be negligible. It is approximately 160 km west of Mapara and has a similar altitude (300–600 m above sea level). Furthermore, it is assumed that Mapara and Te Urewera once experienced gene flow because the intervening area was contiguous forest up until about 1880 (Ian Flux, pers. comm.). We therefore assume that the genetic variation within the Te Urewera population is representative of that at Mapara prior to the bottleneck.

2.2 METHODS

2.2.1 Theoretical prediction of the genetic variation remaining at Mapara

Random genetic drift

The proportion of genetic variation remaining (*P*) in a population after a single bottleneck event can be approximated by the equation:

$$P = \left(1 - \frac{1}{2N}\right)$$ Equation 1 (Frankel & Soulé 1981)

where *N* is the total population size. However, the total population size (*N*) may inappropriately measure the severity of the bottleneck because not all of the surviving individuals will contribute genes to the next generation. For this reason, the effective population size (*Ne*) is used and is usually lower than the actual census size. Some important factors that affect *Ne* are overlapping generations, fluctuating population size, and the breeding system (Hartl & Clark 1989; Primack 1993).

For fluctuating population size over *t* generations, the effective population size is the harmonic mean of the actual census sizes:

$$\frac{1}{N_e} = \left(\frac{1}{t}\right)\left(\frac{1}{N_1} + \frac{1}{N_2} + \ldots \frac{1}{N_t}\right)$$ Equation 2 (Hartl & Clark 1989)

Therefore, Equation 1 can be rewritten as:

$$P = \left(1 - \frac{1}{2N_e}\right)$$ Equation 3

Equation 3 describes the effect of a bottleneck over one generation. However, data suggests that the bottleneck at Mapara lasted longer than one kokako generation (see later). Each subsequent generation effectively endures a single bottleneck event and the erosion of genetic variation will become more pronounced as the duration of the bottleneck increases (Frankel & Soulé 1981). Thus, the proportion of genetic variation remaining after the suspected bottleneck at Mapara (which lasted *t* generations) was estimated using the equation:

$$P = \left(1 - \frac{1}{2Ne}\right)^t$$ Equation 4 (Frankel & Soulé 1981)

Figure 6 shows the relationship between the proportion of genetic variation remaining, and the effective population size and bottleneck duration, based on Equation 4. Not surprisingly the proportion of genetic variation remaining decreases with smaller effective population sizes and longer bottleneck duration.

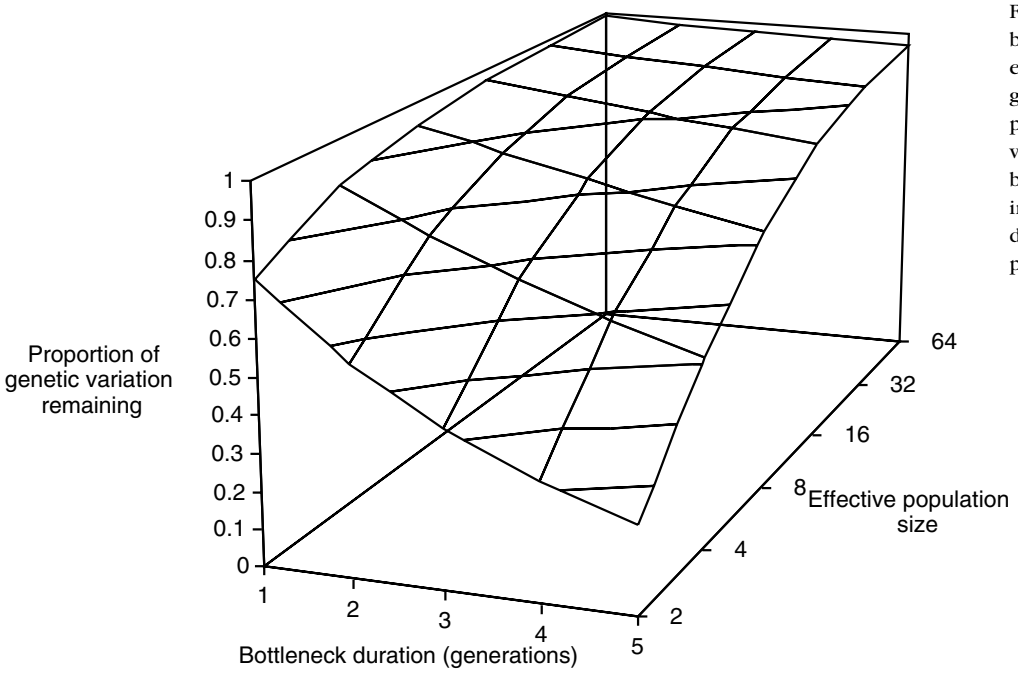


Figure 6. The impact of bottleneck duration and effective sample size on genetic variation. The proportion of genetic variation remaining after a bottleneck decreases with increasing bottleneck duration and decreasing population size.

2.2.2 Empirical estimation of the genetic variation remaining at Mapara

See Murphy (1998) for detailed account of methodologies.

Samples

Eight individuals from Te Urewera, and 12 from Mapara were used in the analysis of genetic variation. Sample sizes were limited by availability and the necessity for samples to be from unrelated individuals.

Minisatellite DNA markers

Minisatellites are non-coding, highly polymorphic sections of DNA that contain short repeated units, each about 16–64 bases long (Jeffreys et al. 1985b; Jeffreys et al. 1985c). Mutation events lead to differences in the number of times the unit is repeated to the extent that a simultaneous analysis of multiple minisatellites give individual specific DNA 'fingerprints'. These DNA fingerprints are now commonly used in paternity assays (e.g. Wetton et al. 1992), individual identification (e.g. Jeffreys et al. 1985a) and the quantification of within-population genetic variance (e.g. Arden & Lambert 1997). The protocols used to produce DNA fingerprints for kokako are explained in detail by Murphy (1998). Two probes were used to produce the multilocus DNA fingerprints, 33.15 (Jeffreys et al. 1985b; Jeffreys et al. 1985c) and *per* (Shin et al. 1985).

Band scoring and data analysis

Minisatellite DNA can be visualised as a series of bands somewhat like a supermarket barcode (Fig. 7). Each band position is given a number; bands at the exact same position in different individuals (homologous bands) were given the same number. Two techniques were used to analyse band scores and subsequently determine the levels of genetic variation within both populations: (1) average genetic distance between individuals within populations, and (2) average heterozygosity within populations.

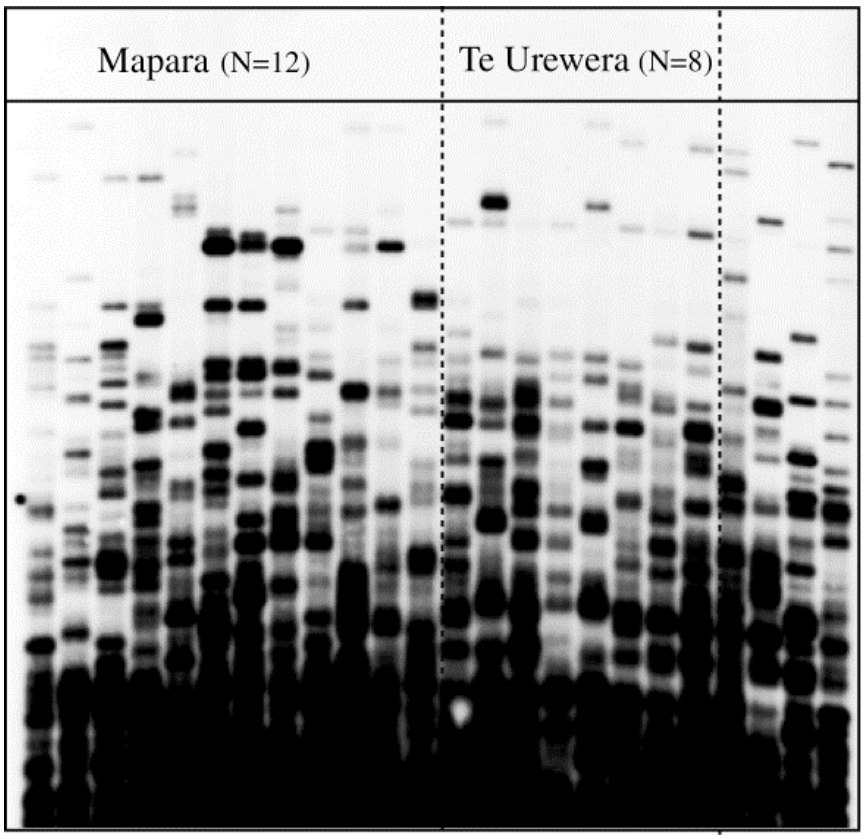


Figure 7. Multilocus DNA fingerprint of North Island kokako. DNA fragments have been separated by size (largest at the top). Fragments which contain minisatellites have been illuminated (seen as bands above) using a radioactively labelled probe (33.15; Jeffreys et al. 1985c) which targets minisatellite DNA.

Two different pairwise genetic distances were calculated in this study. Both require the following parameters:

n = total number of band positions

n_{11} = the number of times individual x and y had a band at the same position

n_{00} = the number of band positions where neither x or y had a band

The first, originally developed by Jaccard (Jaccard 1912) divides the number of shared bands in a pair-wise comparison by the total number of banding positions in that comparison to give genetic similarity as a proportion of shared bands. This can be converted to a measure of genetic distance (D_{JAC}) by subtracting from one:

$$D_{JAC} = \left(1 - \frac{n_{11}}{n - n_{00}}\right) \qquad \text{Equation 5}$$

This calculation only considers shared bands within each pairwise comparison. Therefore, this estimation of genetic distance assumes that homologous band absences are biologically uninformative.

By contrast, the second method calculates genetic distance between individuals (Euclidean distance, D_{EUC}) as a tally of band differences between individuals and therefore includes band absences in the calculation. For example, if neither of two individuals have a band at position *Q*, this is as informative as them both having a band at position *R*. It was originally proposed by Excoffier et al. (1992) in matrix notation, and then developed in the form shown here by Huff et al. (1993):

$$D_{EUC} = n\left(1 - \left(\frac{2(n_{11} + n_{00}}{2_n}\right)\right) \qquad \text{Equation 6}$$

Both the Jaccard or Euclidean methods were used for calculating the mean pairwise genetic distance between two randomly chosen individuals from the same population. Therefore, by comparing these two estimates, it was possible to determine if individual North Island kokako within the Mapara population are genetically more similar (i.e. have less genetic variation) than individual North Island kokako within the Te Urewera population.

A non-parametric permutation-based method was used to test for statistical significance. Pairwise comparisons were made between all individuals from both populations. This produced a distribution of genetic similarity assuming individuals from Mapara and Te Urewera were drawn from the same population. If Mapara's actual '*observed*' mean genetic distance fall outside the 95% confidence interval of this distribution then this would indicate that the average genetic distance between individuals within Mapara are closer than expected given no difference in the variation between two populations. If the Te Urewera population has not experienced a bottleneck then it would be expected to fall within the 95% confidence interval.

Each permutation followed the following procedure:

1. All the possible pair-wise genetic distances within and among the Mapara and Te Urewera data were calculated and pooled,
2. Ten genetic distances from the pooled data were randomly chosen to constitute one hypothetical population (ten being the average sample size),
3. The mean genetic distance was calculated for that group of 10 individuals and was recorded.

One thousand permutations were calculated to estimate the distribution of genetic distances assuming all individuals were from the same population.

Empirical estimation of heterozygosity to infer inbreeding

Inbreeding increases the probability that genes inherited by an offspring will be identical by decent i.e. have a common ancestor (Ayala 1982; Frankel & Soulé 1981). At the population level, this has the effect of increasing the frequency of homozygous individuals. From multilocus DNA fingerprints it is not possible to directly distinguish homozygotes and heterozygotes. However, heterozygous frequency can be calculated as described by (Jeffreys et al. 1985c) using the equation:

$x = (1 - p^2)$

whereby x is the proportion of individuals which have a band at position Q and p^2 is the frequency of individuals without a band at position Q. From Hardy-Weinberg Theory p = 1 – q; we can therefore calculate q by substituting for p:

$$q = 1 - \sqrt{(1 - x)}$$ (Jeffreys et al. 1985c)

From this, an estimate of heterozygosity (H_Q) for band position Q is:

$$H_Q = 2pq$$ (Hartl & Clark 1989)

H_Q was calculated for all band positions within a population and an average taken to determine the frequency of heterozygosity at each band position. This procedure assumes the population is in Hardy-Weinberg equilibrium which could introduce bias under some circumstances (see 2.3 Results).

To test the significance of any observed differences between the mean heterozygosity values for the Mapara and Te Urewera populations a permutation procedure similar to that presented for the Jaccard and Euclidean measures of genetic distance was used:

1. The banding profiles of all individuals were shuffled (permuted),
2. Ten profiles were randomly chosen to make up a hypothetical population,
3. The average heterozygosity value was calculated for the hypothetical population.

This permutation was repeated 1000 times to establish a distribution of heterozygosities (assuming all individuals were from the same population). If the '*observed*' mean heterozygosity for Mapara was significantly reduced by a bottleneck then it would lie outside the 95% confidence interval for the permuted distribution.

2.3 RESULTS

2.3.1 Theoretical prediction of the genetic variation remaining at Mapara

The maximum amount of variation that could have been lost from the Mapara population between 1978 and 1992 is 60%. This figure was calculated assuming that the population size plummeted following the 1978 survey and the effective population in the next and all years to 1992 was eight (eight being the smallest effective population size recorded for Mapara). We also assumed the generation time for kokako was one year so the population was bottlenecked for a total of 14 kokako generations.

That only 40% of the 1978 variation remains within the Mapara kokako population is obviously inaccurate and does not reflect the true dynamics of the bottleneck. It does, however, present the worst-case scenario. Similarly we can estimate the least amount of variation lost if we assume that the bottleneck lasted just one kokako generation and the population remained on average relatively large (say 20 breeding individuals). In this case the expected amount of variation remaining in Mapara would be approximately 98%. Clearly the true amount of variation lost must lie between these two estimates.

Although the true generation time for kokako is not known a useful estimate is simply the average age of the breeding population. In 1997 this was estimated to be 6.3 years, assuming all birds banded as adults were 2 years old when banded (Ian Flux pers. comm.). Thus the bottleneck could only have spanned a maximum of three kokako generations. If we now use that figure in our worst-case-scenario then the maximum amount of variation lost could only be 18% (82% remaining). Clearly this figure must still over-estimate the amount of variation lost. It is likely that more than eight birds bred per generation and many individuals will have lived through the bottleneck and bred once the population size began to increase.

2.3.2 Empirical estimate of the genetic variation remaining at Mapara

A total of 119 band positions were scored from the two autoradiographs produced by the probes 33.15 and *per* (Fig. 7). North Island kokako showed an average DNA fingerprint similarity of 18%. This value was the average proportion of bands shared among individuals, based on Mapara and Te Urewera band scores.

Mean genetic distance

The mean Jaccard genetic distance between individuals within Mapara and Te Urewera was 0.85 and 0.89 respectively (Table 3). These values were not significantly different from the permuted means calculated from the pooled dataset (Fig. 8). The mean Euclidean distance was 32.4 and 33.0 for Mapara and Te Urewera respectively (Table 3). Similarly these values were not significantly different from the permuted means calculated from the pooled data (Fig. 9).

TABLE 3. HETEROZYGOSITY AND MEAN GENETIC DISTANCES AMONG INDIVIDUALS WITHIN THE MAPARA AND TE UREWERA KOKAKO POPULATIONS.

	MAPARA	TE UREWERA	% GENETIC VARIATION REMAINING
Sample size	12	8	
Genetic distance (Jaccard)	0.85	0.89	95.5
Genetic distance (Euclidean)	32.4	33.0	98.0
Heterozygosity	0.21	0.19	-

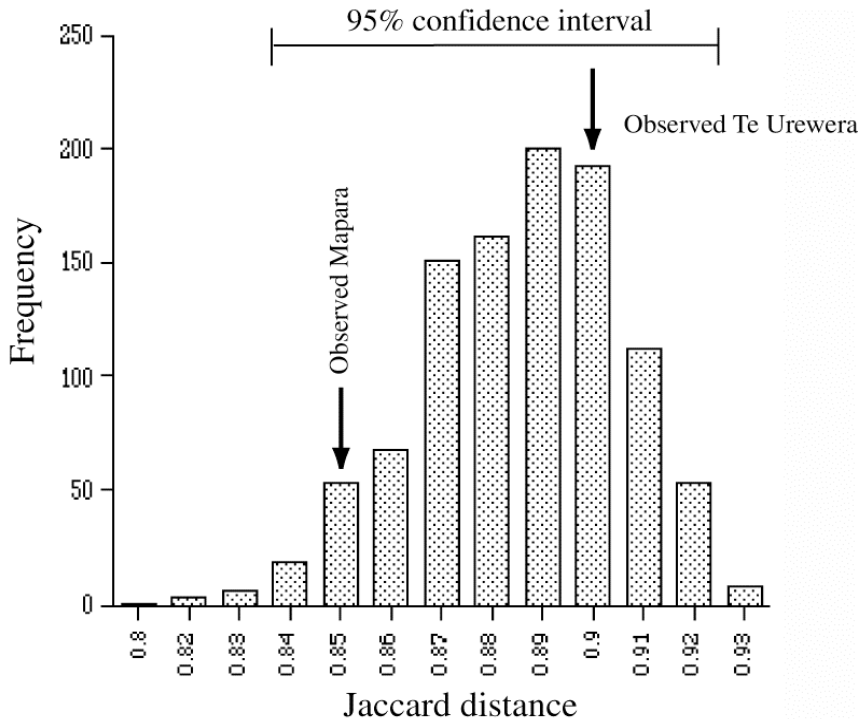


Figure 8. Mean Jaccard genetic distances calculated for 1000 hypothetical populations made up of individuals randomly selected from the sample populations. The observed Jaccard distances for Mapara and Te Urewera lie within the 95% confidence interval.

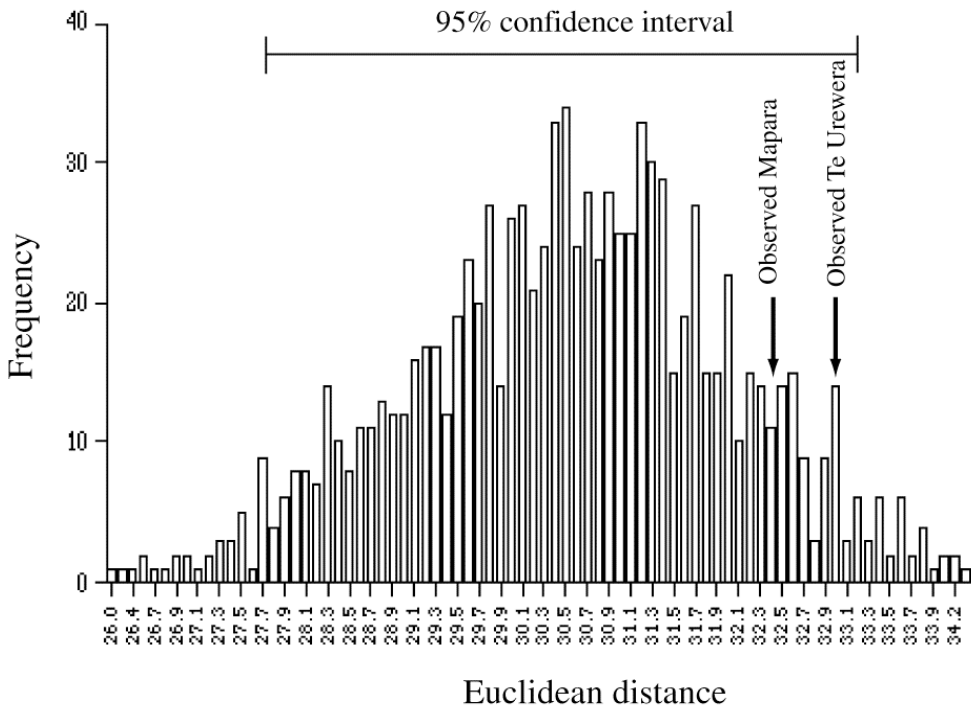


Figure 9. Mean Euclidean genetic distances calculated for 1000 hypothetical populations made up of individuals randomly selected from the sample populations. The observed Euclidean distances for Mapara and Te Urewera lie within the 95% confidence interval.

Heterozygosity

The estimated mean heterozygosity was 0.21 and 0.19 for Mapara and Te Urewera respectively (Table 3). These values were not significantly different from the permuted heterozygosity values calculated from the pooled dataset.

Two assumptions of the Hardy-Weinberg model were violated: that the population should be large and mating should be random. These violations could lead to an inflated heterozygosity figure. However, bias introduced for low estimates of heterozygosity is negligible (see Murphy 1998), therefore the estimated heterozygosity for Mapara (0.21) could only be slightly exaggerated for all levels of inbreeding.

2.3.3 Summary of results

1. The proportion of genetic variation remaining at Mapara post bottleneck was theoretically estimated to lie between 82 and 98%.
2. Based on DNA fingerprint analysis, both the average Jaccard and Euclidean genetic distances among individuals at Mapara and Te Urewera did not significantly differ.
3. Similarly, the average heterozygosity of individuals at Mapara and Te Urewera did not significantly differ.

2.4 DISCUSSION

Te Urewera and Mapara had similar mean genetic distance between individuals based on both Jaccardian and Euclidean measures. Also average heterozygosity per band position did not significantly differ between the two populations. The Jaccard and Euclidean distances suggest that Mapara has the equivalent of 96 to 98% of the variation present at Te Urewera (Table 3), although statistically no

difference was found between the two populations. Theoretical calculations based on demographic data suggest that between 82% and 98% of genetic variation would still be present in the Mapara population after the bottleneck. In summary, all the evidence suggests the bottleneck at Mapara would not have greatly reduced the amount of genetic variation when compared with that within the Te Urewera population.

If the Te Urewera population also lost genetic variation through a decline in population size this could mask any loss of variation within the Mapara population. However, we consider this an implausible scenario for several reasons. Firstly, available evidence suggests that Te Urewera has maintained a large population for a long period (Hudson 1994). Secondly, theoretical calculations based on the demographic data from Mapara suggest only a minor loss of variation given the duration of the bottleneck. Thirdly, if both the Mapara and Te Urewera population had lost variation due to bottlenecks then one would expect low genetic distances within populations and high distances between. This would force the mean genetic distances for each population into the extreme right-hand tail of the permuted distributions. In fact, the genetic distance for each population lies within the 95% confidence intervals of the permuted means. Finally, the level of minisatellite variation within North Island kokako generally is high compared to other avian systems (see Section 3.3.2; Murphy 1998). This evidence suggests that the level of minisatellite variation at Te Urewera is not unusually low and could not mask any bottleneck effects at Mapara.

The sample sizes for each population are small, but even so they show a relatively high level of genetic variability in each population. If low levels of variability had been found then the question of inadequate or biased sampling would have to be investigated before we could conclude that one or both of the populations had lost variation due to a prolonged population bottleneck.

In conclusion, it appears that the kokako population at Mapara still maintains most of the pre-bottleneck levels of genetic variation.

3. References

Arden, S.L.; Lambert, D.M. 1997: Is the black robin in genetic peril? *Molecular Ecology 6*: 502.

Ayala, F.J. 1982: Population and Evolutionary Genetics: A Primer. California: Benjamin/Cummings Publishing Co.

Baillie, J.; Groombridge, B.; Giminez, D.; Gardenfors, U.; Stattersfield, A. 1996: IUCN Red List of Threatened Animals. Washington: IUCN.

Baker, A.; Marshall, H. 1997: Mitochondrial Control Region Sequences as Tools for Understanding Evolution. In: Mindell, D. (Ed.): Avian Molecular Evolution and Systematics. London: Academic Press.

Baker, A.J.; Daugherty, C.H.; Colbourne, R.; McLennan, J.L. 1995: Flightless brown kiwis of New Zealand possess extremely subdivided population structure and cryptic species like small mammals. *Proceedings of the National Academy of Sciences of the United States of America 92*: 8254–8258.

Coker, P. 1978 Internal New Zealand Wildlife Service Report. Wellington (unpublished).

Excoffier, L.; Smouse, P.E.; Quattro, J. M.: 1992 Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics 131*: 479–491.

Falla, R.; Sibson, R.B.; Turbott, E.G. 1981: Collins Guide to the Birds of New Zealand. Auckland: Collins.

Felsenstein, J. 1985: Confidence limits on phylogenies: An approach using the bootstrap. *Evolution 39*: 783–791.

Fleischer, R.; Tarr, C.; Pratt, T. 1994: Genetic structure and mating system in the palila, and endangered Hawaiian honeycreeper, as assessed by DNA fingerprinting. *Molecular Ecology 3*: 383–392.

Fleming, C.A. 1962: New Zealand biogeography. A paleontologist's approach. *Tuatara 10*: 53–108.

Fleming, C.A. 1979: The geological history of New Zealand and its life. Auckland: Auckland University Press.

Frankel, O.H.; Soulé, M.E. 1981: Conservation and Evolution. Cambridge: Cambridge University Press.

Gnanadesikan, R. 1997: Methods for Statistical Data Analysis of Multivariate Observations. Second Edition. New York: John Wiley and Sons.

Gottelli, D.; Sillerozubiri, C.; Applebaum, G.D.; Roy, M.S.; Girman, D.J.; Garciamoreno, J.; Ostrander, E.A.; Wayne, R.K. 1994: Molecular genetics of the most endangered canid - the Ethiopian wolf *Canis simensis*. *Molecular Ecology 3*: 301–312.

Hartl, D.; Clark, A. 1989: Principles of Population Genetics. Sunderland, Massachusetts: Sinauer Associates.

Hasegawa, M.; Kishino, H.; Yano, D. 1985: Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution 21*: 160–174.

Hudson, J. 1994: East Coast Conservancy Report: Department of Conservation (unpublished).

Huff, D.; Peakall, R.; Smouse, P. 1993: RAPD variation within and among natural populations of outcrossing buffalograss [*Buchloë dactyloides* (Nutt.) Englem.]. *Theoretical and Applied Genetics 86*: 927–934.

Jaccard, P. 1912: The distribution of the flora of the alpine zone. *New Phytologist 11*: 37–50.

Jeffreys, A.; Brookfield, J.; Semeonoff, R. 1985a: Positive identification of an immigration test-case using human DNA fingerprints. *Nature 317*, 818–819.

Jeffreys, A.; Wilson, V.; Thein, S. 1985b: Hypervariable "minisatellite" regions in human DNA. *Nature 314*: 67–73.

Jeffreys, A.; Wilson, V.; Thein, S. 1985c: Individual -specific "fingerprints" of human DNA. *Nature 314*: 67–73.

King, C. 1984: Immigrant killers: introduced predators and the conservation of birds in New Zealand. Auckland: Oxford University Press.

Lavers, R.B. 1978: Distribution of the North Island kokako (*Callaeas cinerea wilsoni*): A review. *Notornis 25*: 165–185.

Leathwick, J.R.; Hay, J.R.; Fitzgerald, A.E. 1983: The influence of browsing by introduced mammals on the decline of the North Island kokako. *New Zealand Journal of Ecology 6*: 55–70.

McGlone, M.S. 1985: Plant biogeography and the Cenozoic history of New Zealand. *N.Z. Journal of Botany 23*: 723–749.

Mindell, D.P. 1997: Avian molecular evolution and systematics. San Diego, California: Academic Press.

Murphy, S. 1998: The patterns of genetic variation in North Island kokako *Callaeas cinerea wilsoni*. Australian National University, Canberra.

Palumbi, S.; Martin, A.; Romano, S.; McMillan, W.O.; Stice, L.; Grabowski, G. 1991: The Simple Fool's Guide to PCR, version 2.0. Department of Zoology, University of Hawaii, Honolulu.

Peakall, R.; Smouse, P. 1998: GenAlEx: Genetic Analysis in Excel. Support software for the national workshop 'Genetic Analysis for Population and Biogeographic Studies' July 1998. Canberra.

Primack, R.B. 1993: Essentials of Conservation Biology. Sunderland, Massachusetts: Sinauer Associates, Inc.

Quinn, T.W. 1992: The legacy of Mother Goose - phylogeographic patterns of lesser snow goose *Chen caerulescens caerulescens* maternal lineages. *Molecular Ecology 1*: 105–117.

Sambrook, J.; Fritsch, E.F.; Maniatis, T. 1990: Molecular cloning: A laboratory manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Sherwin, W.; Murray, N.; Marshall Graves, J.; Brown, P. 1991: Measurement of genetic variation in endangered populations: bandicoots (Marsupialia: Peramelidae) as an example. *Conservation Biology 5*: 103–108.

Shin, H.-S.; Bargiello, T.A.; Clark, B.Y.; Jackson, F.R.; Young, M.W. 1985: An unusual coding sequence from a *Drosophila* clock gene is conserved in vertebrates. *Nature (London) 317*: 445–448.

Stevens, G.; McGlone, M.; McCulloch, B. 1988: Prehistoric New Zealand. Heinemann Reed, Auckland, New Zealand.

Stewart, C.B. 1993: The powers and pitfalls of parsimony. *Nature 361*: 603–607.

Swofford, D.L. 2000: PAUP*: Phylogenetic Analysis Using Parsimony. Version 4. Sunderland, Massachusetts: Sinauer Associates, Inc.

Swofford, D.L.; Olsen, G.J.; Wadell, P.J.; Hillis, D.M. 1996: Phylogenetic Inference. Pp. 407–514 in Hillis, D.M.; Moritz, C.; Mable, B.K. (Eds): Molecular Systematics. Sunderland, Massachusetts: Sinauer Associates.

Tarr, C.L. 1995: Primers for amplification and determination of mitochondrial control-region sequences in oscine passerines. *Molecular Ecology 4*: 527–529.

Taylor, A.C.; Sherwin, W.B.; Wayne, R.K. 1994: Genetic variation of microsatellite loci in a bottlenecked species - the northern hairy-nosed wombat *Lasiorhinus krefftii*. *Molecular Ecology 3*: 277–290.

Thompson, J.D.; Higgins, D.G.; Gibson, T.J. 1994: Improved sensitivity of profile searches through the use of sequence weights and gap excision. *Computer Applications in the Biosciences 10*: 19–29.

Wenink, P.W.; Baker, A.J.; Tilanus, M.G.J. 1994: Mitochondrial control region sequences in two shorebird species, the turnstone and the dunlin, and their utility in population genetic studies. *Molecular Biology & Evolution 11*: 22–31.

Williams, G.R. 1976: The New Zealand wattlebirds (Callaetidae). Pp. 161–170 in The Proceedings of the International Ornithological Congress, Canberra.