# A COMPARISON OF MICROSATELLITE ISOLATION TECHNIQUES USING AVIAN GENOMES

# A COMPARISON OF MICROSATELLITE ISOLATION TECHNIQUES USING AVIAN GENOMES

By

**Sean M. Gregory, B.Sc.**

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Science

McMaster University

MASTER OF SCIENCE (2004) McMaster University

(Biology) Hamilton, Ontario

TITLE: A Comparison of Microsatellite Isolation Techniques Using Avian Genomes

AUTHOR: Sean M. Gregory, B.Sc. (McMaster University)

SUPERVISOR: Dr. James S. Quinn

NUMBER OF PAGES: ix, 76.

## Abstract

In the past two decades or so, microsatellites have become a very widely used genetic tool in many disciplines of biology. Their major downfall, however, is that they often need to be isolated *de novo* before they can be applied to molecular studies. Traditional shotgun cloning can be successful, but it is often overly costly and time consuming. Compounding this downfall, isolating microsatellites from some taxa has been shown to be difficult. For example, on average only 0.46% of all clones screened using avian genomes will yield positive clones. This is thought to be a result of a smaller avian genome, a requirement for flight. Several alternative methods have been developed for isolating microsatellites, but the choice as to which isolation method to use is often arbitrary. To address this, four species of birds, the smooth-billed ani (*Crotophaga ani*), herring gull (*Larus argentatus*), yellow-bellied elaenia *(Elaenia flavogaster*), and pukeko (*Porphyrio porphyrio*), representing four different orders were used to compare two alternative isolation methods. Enrichment via selective hybridization versus cloning with Lambda Zap phage vector were compared in terms of monetary requirements (total startup cost as well as per isolation attempt cost), and time requirements (total time from start to finish and hands-on experimentation time). No significant difference was detected in terms of number of polymorphic microsatellite loci isolated by each method ($p = 0.57$), with enrichment yielding more for the anis and elaenias, Lambda Zap yielding more for herring gulls, and both methods isolating equal numbers for pukekos. Nor was any difference found between the methods for dollars spent per sequence with repeat (SWR) using the startup cost ($p = 0.30$). Enrichment, however, proved to be significantly more effective in terms of dollars per SWR isolated using the per use cost ($p = 0.004$) as

well as hands-on minutes per SWR ($p = 0.01$) and total minutes per SWR ($p < 0.01$). Based on these findings, selective hybridization is the better choice for microsatellite isolation.

**Acknowledgements**

I would like to thank my supervisor Jim Quinn for all his help throughout this project. From giving me my first taste of fieldwork to all the editing and everything in between, his efforts were greatly appreciated. Thanks also to J-P Xu for helpful advice and guidance.

Thanks also to everyone at the Cabo Rojo National Wildlife Refuge in Puerto Rico, particularly Joe. The weather was too hot, the hours were too long, and the birds were too elusive, but it was an experience I'm glad I got to have. Thank you to Bridget Stutchbury and Bonnie Wolfenden for elaenia data and DNA samples. Also to the members of the Quinn lab who helped out whenever possible, especially with helpful comments at lab meetings. Thanks especially to Chris Somers, whose deadness inside is an inspiration to us all.

Thank you to all the members of the gang who have come and gone over the many years that I've been here. Without friends like these, it would have seemed like even longer. Thanks Joe, Nikki, Jimmy, Tash, Chris S., Vic, Chris G., Carol, Mark, and especially Mich, Eric, & Kobe. I think I spent more time at their house than my own! A very big thank you to Makiko Kajimura for all of her support over the past couple years.

Finally, thank you to my family for all of their support. Ryan & Sally, Michele and Frank, Bob and Marilyn. I've said it before and I'll probably say it again, but I couldn't have gotten here without them!

## Table of Contents

## List of Tables

## List of Figures

## Introduction:

Microsatellites consist of short sequences of DNA 1-6 base pairs long repeated in tandem. They generally occur in non-coding regions of the genome, however some human genetic diseases have been shown to be the result of trinucleotide microsatellite repeats in coding regions (Fernàndez-López *et al.*, 2004). Despite a lack of consensus on an exact definition (Chambers & MacAvoy, 2000), they continue to increase in popularity for such studies as population genetics (Jarne & Lagoda, 1996; Bernatchez & Duschene, 2000), kinship (Queller *et al.*, 1993; Jones & Ardren, 2003) and genome mapping (Dib *et al.*, 1996; Schuler *et al.*, 1996; Knapik *et al.*, 1998). Microsatellites are of little use for higher-level systematics, since their relatively high mutation rates can result in alleles being the same size from reversions rather than common descent, and because there is variation in mutation rates among species (Schlötterer, 2001; Goldstein *et al.*, 1997). Their utility for shorter-term population studies, however, stems from their ability to be amplified by the polymerase chain reaction, thus only requiring small DNA samples, their often high levels of polymorphism and codominant inheritance, and the fact that they have been found in every eukaryotic genome studied to date (Schlötterer, 2001). A major drawback for microsatellites, however, is that they often need to be isolated *de novo*, particularly if the species of interest is not commonly studied. Cross-species amplification can sometimes be carried out for closely related species (Primmer *et al.*, 1996; Moore *et al.*, 1991), but Schlötterer (2001) showed that a species divergence time of only eleven million years can be enough to reduce cross-species amplification success by more than 50%. Loci isolated from other species often show reduced

variability as well (Schlötterer, 2001). Isolating microsatellites can be an expensive and time-consuming endeavor. The traditional method, labeled "shotgun cloning" for its lack of precision and broad target, has worked for various taxa (e.g. Gibbs *et al.*, 1998a; Rassmann *et al.*, 1991). Briefly, DNA is fragmented using restriction enzymes, size-selected, ligated into a common plasmid, and transformed into bacterial cells. The result is generally thousands of recombinant clones, which are then screened by Southern hybridization using radiolabeled or fluorescent repeat-containing probes. Results are typically around 2% positive for mammals, but less than 0.4% for birds. Microsatellites have been isolated from avian genomes using this method (e.g. Gibbs *et al.*, 1998b), but typically thousands, or even tens of thousands of clones (bacterial colonies) need to be arrayed and screened. Birds have a reduced genome size, generally thought to be a requirement for flight (Neff, 2001; Gregory, 2002) since smaller genomes means smaller cells, and therefore result in faster metabolic rates which are necessary for the high energetic costs of flight. This reduced genome size also means reduced number of microsatellites (Primmer *et al.*, 1997, Longmire *et al.*, 1999). With such a reduction in total numbers of microsatellite repeats, traditional shotgun cloning is generally not an efficient method for isolating these repeats. Several alternative methods have been developed, but the decision as to which method to use is often arbitrary. Zane *et al.* (2002) compared various methods for cost and time requirements by reviewing published data from the journals Molecular Ecology and Animal Genetics, but their data were very general, and had very large spread over the values given. For example, they report enrichment as costing anywhere between $1000 and $4000 US, presumably because different kits and screening techniques between methods. Here we compare two of these

alternative methods in terms of monetary and time costs, as well as success rates, and report them in very specific values. The two methods are enrichment by selective hybridization using biotinylated repeat probes versus cloning using the Lambda Zap Express bacteriophage. Enrichment differs from shotgun cloning in that genomic DNA fragments containing microsatellite repeats are selected while DNA not containing microsatellites is removed before cloning, resulting in a higher proportion of positives. Lambda Zap on the other hand uses a bacteriophage, allowing for hundreds of times more inserts than with shotgun cloning to be screened on a single agar plate. Both these protocols have been successful at isolating variable microsatellite repeats in taxa that have proven troublesome using shotgun cloning (e.g. Blanchard & Quinn, 2001; Hughes & Deloach, 1997 for Lambda Zap; Perrin & Roy, 2000; Schlosser *et al.*, 2003 for enrichment). This direct comparison of the two methods is meant to act as a guide for a researcher wanting microsatellite loci from any given taxonomic group, but particularly from one that is low in microsatellite frequency such as birds. It will not only lay out the steps required for the two methods, but also the time and financial commitments needed for each.

Four species of birds belonging to four different orders were used so that the effectiveness of each microsatellite isolation method could be assessed across a wide range of bird taxa, and that any differences between the methods could be shown to exist across the entire class.

Smooth-billed ani (*Crotophaga ani*)

The smooth-billed ani (order Cuculiformes) is a communally breeding species found throughout the Caribbean Islands, South America, Costa Rica, and Florida (Quinn & Startek-Foote, 2000). Previous studies on this species have used microsatellite primers for preliminary analyses of parentage and relatedness (Blanchard & Quinn, 2001; Blanchard, 2000; G. Schmaltz, pers. com.). Because of the limited number of loci available and their often low levels of variability, however, additional loci would greatly increase the discrimination ability in relatedness comparisons.

Herring gull (*Larus argentatus*)

The herring gull (order Charadriiformes) is often used as a sentinel species for chemical exposures and effects (e.g. Pekarik & Weseloh, 1998; Keith, 1966; Gilbertson & Hale, 1974), particularly in the Great Lakes, and in behaviour studies of parentage and parental investment (e.g. Davis & Quinn, 1997). It is a colonial nesting, fish-eating predator, and stays within the Great Lakes area year-round. Despite the importance of this species, no microsatellites have yet been developed for their genetic analysis. Genetic studies thus far on herring gulls have used multi-locus minisatellite profiles (Yauk & Quinn, 1996, 1999; Yauk *et al.*, 2000), which generally cannot provide the fine-scale details obtained using single-locus microsatellites. In addition, different populations of gulls in the Great Lakes show morphological differences, suggesting some degree of non-random mating leading to population differentiation (unpublished). However, no genetic differences have been detected to date (Yauk & Quinn, 1999). Microsatellites could be used for finer-scale resolution on population structure.

Pukeko (*Porphyrio porphyrio*)

The pukeko (order Gruiformes) is a communal joint-nesting bird found widespread over tropical Africa, and from southern Asia to Australia and New Zealand, with small, scattered populations in the Mediterranean basin in Europe (Tucker & Heath, 1994). Groups often differ in their relatedness, from outbred to highly inbred. Multilocus minisatellite profiling has been used to examine the effect of relatedness with group members on helping behaviour and parental effort (Jamieson *et al.*, 1994; Jamieson & Quinn, 1997). Minisatellite analysis does not give the fine scale resolution that can be achieved from microsatellites, however, and exact levels of relatedness are often difficult or impossible to measure. Microsatellites would allow for better statistical analyses of relatedness, and thus aid in providing more insight into the role relatedness plays on cooperation.

Yellow-bellied elaenia (*Elaenia flavogaster*)

Yellow-bellied elaenias (order Passeriformes) belong to the family Tyrannidae, or the flycatchers. They are generally found in pairs widespread from central Mexico to southern Brazil. Though these birds are generally thought of as socially monogamous, behavioural data suggest that at least some effort is spent on attempts at extra-pair copulations. Microsatellite markers will prove useful in studying the success of these attempts and the frequency of extra-pair matings.

## Methods:

*Sample Collection and DNA Extraction*

Smooth-billed anis were captured in mist nets and nest traps at Cabo Rojo National Wildlife Refuge and La Cartajena NWR in Puerto Rico, USA. 150-200 μL of blood from adults or 50-100 μL from chicks was collected from the jugular vein and stored in lysis buffer. Measurements of mass, beak length and depth, head-to-tip length, and tarsus length were also taken. Nests were checked daily, and tissue was collected from buried or unhatched eggs once the brood had fledged or the adults had abandoned the nest. DNA was extracted from blood using a saturated salt extraction protocol, and from tissue using a standard phenol/chloroform procedure (Sambrook *et al.*, 1989).

Pukeko samples were provided by J. Haselmayer or I.J. Jamieson from funnel-trapped adults or recently hatched nestlings near Dunedin, New Zealand (Jamieson *et al.*, 1994). Herring gull samples were provided by J.S. Quinn and members of the field team. Blood was collected by brachial-vein puncture from gull colonies in Hamilton Harbour, Ontario. Elaenia samples were collected from nest-trapped adults and newly hatched young by B. Stutchbury and E. Morton in Gamboa in Panama from January to March 1996 and 1997. DNA was extracted using a standard phenol/chloroform procedure (Sambrook *et al.*, 1989).

*Cross-Species Amplification*

Cross-species amplification was attempted for smooth-billed anis using three microsatellite primers isolated for Guira cuckoos (*Guira guira*) (Muniz *et al.*, 2003), and for yellow-bellied elaenias using two primers isolated for least flycatchers (*Empidonax minimus*) (Tarof *et al.*, 2001). All Guira cuckoo primer pairs amplified a product for the anis, but all were monomorphic or contained too many non-specific bands to be scored. One of the two flycatcher pairs amplified a product, but again it was monomorphic. The other failed to amplify a product.

***Enrichment***

The enrichment protocol used in this study is based largely on that of Glenn & Schable (2004) with slight modifications. Glenn and Schable's (2003) protocol is an updated and simplified amalgamation of the earlier enrichment protocols of Hamilton *et al.* (1999) and Paetkau (1999).

*Restriction Enzyme Digest*

For each species, 1 μg of DNA from each of five unrelated adult individuals was combined and digested with Rsa I and Xmn I (New England Biolabs, Inc.) in a final volume of 25 μL for 2 hours at 37°C. Samples were chosen so that at least two individuals of each sex were used. To ensure complete digestion had occurred, 2 μL

were run on a 1.5% agarose gel and visualized using UV light following ethidium bromide staining.

*Linker Ligation to DNA Fragments*

The SuperSNX24 double stranded linker (Sigma) was ligated to each end of the DNA fragments (Hamilton *et al.*, 1999; Glenn & Schable, 2004). This linker acts as the primer binding site for subsequent PCR steps and also aids in cloning the fragments into vectors (see below). The double stranded linker was prepared by mixing 100 µL of each 10 µM primer, and 100 mM NaCl. Linkers were ligated to the DNA fragments by combining 7 µL of the double stranded linkers from above, 1x ligase buffer (New England Biolabs), and 800 U DNA Ligase (New England Biolabs), then adding this to the aforementioned digested DNA. Incubation was carried out overnight at 16°C. Self-ligation of the linkers does not occur since any linker dimerization forms an Xmn I recognition site and is therefore cleaved. To ensure that the ligation worked, a PCR reaction was run in a final volume of 25 µL using 2 µL linker-ligated DNA from above, 1x PCR buffer, 0.625 µg BSA, 0.5 µM SuperSNX24-F, 150 µM dNTPs, 2 mM $MgCl_2$, and 1 U Taq DNA Polymerase. Cycling conditions were as follows: 95°C for 2 minutes; 20 cycles of 95°C for 20 seconds, 60°C for 20 seconds, 72°C for 1.5 minutes. 4 µL of PCR product were run on a 1.5% agarose gel with a 100 bp ladder. All four species showed the desired smear centred around the 500 bp region.

*Dynabead Enrichment for Microsatellite-Containing DNA Fragments*

Two oligonucleotide mixes were used for isolating microsatellite repeats: one containing dimers [$(AC)_{12}$ and $(AG)_{12}$] and one with trimers and tetramers [$(AGG)_8$, $(AGGG)_6$, $(AAAG)_6$] (1 μM of each repeat). These probes were chosen because of their relatively higher frequency (Primmer *et al.*, 1997) and published isolation success from avian genomes (e.g. Primmer *et al.*, 1995; McConnell *et al.*, 1999). In a 200 μL PCR tube, 25 μL 2x Hyb Solution (12x SSC, 0.2% SDS), 10 uL of the biotinylated microsatellite probe mix (with oligos at 1 μM each), 10 μL of linker ligated DNA from above, and 5 μL $dH_2O$ were combined and run on a PTC-200 Peltier Thermal Cycler under the following reaction: 95°C for 5 minutes, quickly ramped down to 70°C, decreased by 0.2°C every 5 seconds down to 50°C, then held at 50°C for 10 minutes. It was then ramped down 0.5°C every 5 seconds down to 40°C, and then quickly lowered to 15°C. This programme denatures the DNA, and then goes to a temperature slightly higher than the annealing temperature ($T_m$) of the oligonucleotides. The oligos are allowed to hybridise with the DNA fragments as the temperature slowly decreases to the oligo's $T_m$. While the DNA was in the thermocycler, 50 μL of Dynabeads were washed twice with Tris/EDTA (TE) and twice with 1xHyb solution, with the beads captured using the magnetic particle collecting (MPC) unit after each wash. The washed beads were resuspended in 150 μL TE. The DNA/probe mix was then added to the washed Dynabeads and incubated on a rotator at room temperature for 1 hour. The beads were captured with the MPC, and the supernatant drawn off. The beads were washed two times with 400 μL 2xSSC, 0.1% SDS, then twice with 1xSSC, 0.1% SDS in order to

remove any DNA sequences not bound to the streptavidin coated beads, and therefore not containing microsatellite sequences (Glenn & Schable, 2004). 22 µL of 3M NaOAc was added and the solution was gently mixed. 444 µL of 95% ethanol was added, the tubes were mixed by inversion several times, and placed at -20°C overnight. They were then centrifuged at 14,000 G for 10 minutes. The supernatant was discarded and 500 µL of 70% ethanol added. The tubes were centrifuged for an additional minute, then the supernatant was removed and the pellet left to air-dry until no trace of ethanol remained. The pellet was then resuspended in 25 µL of Tris/Low-EDTA (TLE) (10 mM Tris, 0.1 mM EDTA).

To increase the amount of enriched DNA, a PCR reaction was performed on the eluted DNA fragments. In a 25 µL final volume, 5 µL of eluted fragments were combined with 1x PCR buffer (New England Bioloabs), 0.625 µg BSA, 150 µM dNTPs, 0.5 µM SuperSNX24-F, 2 mM $MgCl_2$, and 1.5 U Taq DNA polymerase. Cycling conditions consisted of 2 minutes at 95°C; 25 cycles of 95°C for 20 seconds, 60°C for 20 seconds, 72°C for 90 seconds; then 72°C for 30 minutes. 4 µL were run on a 1.5% agarose gel to see if DNA was successfully recovered from the enrichment. A smear centred around 500 bp was observed. If the smear was not present then the enrichment eluted all of the DNA, and enrichment was attempted again using less stringent washes. This was sufficient for obtaining the desired smear in this study, however Glenn & Schable (2003) suggest that if bands rather than a smear are visible, it means that there are very few loci recovered. They recommend trying multiple PCRs to see if others can be amplified. If that fails, they suggest enriching with a different set of oligonucleotide

probes. Amplified DNA was then cloned using the Invitrogen TOPO TA Cloning kit according to the instruction manual.

*Testing Transformed Plasmid DNA*

Colonies of TOP10F' bacterial cells (Invitrogen) with the incorporated enriched DNA from above were grown on eight separate plates for each enriched product (dimers or tri-/tetramers) for each species. Between 100-180 colonies grew on each plate after overnight incubation at 37°C. The plates were chilled at 4°C for at least 2 hours, then colony lifts were performed using 82 mm diameter nylon membranes (Hybond N-Amersham-pharmacia biotech). The membranes were then placed on a series of saturated filter papers containing denaturing solution (1.5 M NaCl, 0.5 M NaOH for 4 minutes), two of neutralizing solution (1.5 M NaCl, 0.5 M Tris-HCl for 3 minutes each), then rinsing solution (0.2 M Tris-HCl, 2xSSC for 1 minute). The lifts were then baked at 80°C for 2 hours to bind the DNA to the membrane.

*Probing*

Membranes were prehybridized in Westneat's solution (7% SDS, 0.001 M EDTA, 0.25 M $Na_2HPO_4$, 1% BSA) for two hours at 65°C. Probes were labeled using random primer extension with alpha $^{32}P$ dCTP and Klenow fragment (Oligolabeling Kit – Amersham-pharmacia biotech). Hybridization was performed overnight at 65°C. The filters were then washed twice with 2x SSC, 1% SDS at 65°C for 20 minutes, sealed in

seal-a-meal bags, then placed on Kodak Biomax film. Approximately 20% of colonies gave a positive signal for the $(AC)_{12}/(AG)_{12}$ probes, and 10% for the $(AGG)_8/(AGGG)_6/(AAAG)_6$ probes.

*Minipreps*

Minipreps were done by lifting positive colonies using a P200 pipette tip, which was then dropped into a 15 mL Sarstedt conical tube containing 2 mL LB broth with ampicillin at a final concentration of 50 μg/mL and grown overnight at 37°C. The contents were transferred to a 1.5 mL microfuge tube and centrifuged at 14,000 G for two minutes. After discarding the supernatant, the tubes were blotted on paper towel to remove any remaining liquid. 200 μL of Buffer P1 (0.05 M Tris, 0.01 M EDTA, pH 8.0) was added and the pellet resuspended by vortexing. 200 μL of Buffer P2 (0.2 M NaOH, 1% SDS) were then added and the tubes mixed by inverting several times, followed by 200 μL of Buffer P3 (3 M K-Acetate, pH 5.5 with acetic acid). The tubes were vortexed for 10 seconds, then spun at 14,000 G for 10 minutes, and the supernatant carefully transferred to new tubes. 0.5 μL of RNAse A was added. Following one hour of incubation at 37°C, 350 μL phenol/chloroform was added. The tubes were vortexed for 20 seconds and spun for four minutes. 500 μL of the aqueous top phase was transferred to new tubes, making sure the interface between the layers was not disturbed. 350 μL of isopropanol was added, the tubes briefly vortexed, then stored at –20°C for one hour. They were then spun for eight minutes, and the supernatant removed using a P200 pipette. The pellets were washed with 300 μL of 80% ethanol, then spun three minutes

and the supernatant removed. The pellets were left to air dry, and then resuspended in 30 μL $dH_2O$.

To test positives, a PCR reaction was done using the SuperSNX24 primers and the oligonucleotide repeat as a primer. Because the oligonucleotide repeat can have multiple primer binding sites on a tandem repeat, a blurry band or smear resulted. Jordan *et al.* (2002) used a similar method by including both forward and reverse universal primers (M13) as well as the repeat. They identified positives by the presence of two bands, one from the entire insert between the two linkers being amplified, and one between a linker and the repeat.

Sequencing was done by the Natural Resources DNA Profiling and Forensic Centre at Trent University in Peterborough, ON. Primers were selected for sequences containing repeats and with sufficient flanking regions using the programs Primer3 (Rozen & Skaletsky, 2000) and Oligo (Rychlik, 1992). Primers were made by Sigma-Genosys.

## Lambda Zap

### *Constructing & Screening a Library*

2 μg of DNA from each of five individuals were pooled and digested with Sau3AI in a volume of 100 μL. 4 μL was run on a 1.5% agarose gel to check for complete

digestion. 1/10 of the total volume of 10M Ammonium Acetate was added to stop the reaction. Two times the volume of –20°C ethanol was added, the solution vortexed, and incubated at -20°C overnight. It was then spun at 14,000 rpm for 20 minutes, and the supernatant removed. The pellet remaining was washed with 500 µL 70% ethanol, gently mixed, and then briefly centrifuged. Following removal of the ethanol, the pellet was left to air-dry for 10 minutes, after which time it was resuspended in 15 µL $dH_2O$, and 3 µL orange g dye was added. The entire volume was loaded into a well of a 1% agarose gel with 100 bp ladder size reference in the wells on each side of the sample, and the gel was run at 4 V/cm for approximately 2 hours. Once the orange G was near the end of the gel, the lane of the gel containing sample DNA was cut out, and the outer-portion of the gel was stained in ethidium bromide. The gel was removed from the stain, and the portion of the gel that was not stained was put back into the gap. Visualization was done using a UV light table, and the areas of the sample corresponding to 200-350 bp and 350-500 bp were cut out. These two size ranges were selected so that concatamers, or vectors containing more than one insert, could be easily identified. Removing the portion of the gel containing the DNA prior to staining is necessary with Lambda Zap to avoid contaminating the cloned DNA with ethidium bromide but not with enrichment since the gels from enrichment are simply for testing and are discarded following visualization, whereas with Lambda Zap the DNA is extracted from the gel and used throughout the protocol. The DNA was purified from the agarose gel plug and quantified using a DyNAQuant 200 fluorometer (Hoefer). Ligation into Lambda Zap Express (Stratagene) was done using Stratagene's DNA ligation kit at a 2:1 insert to vector molar ratio in a 5 µL volume using 1 µL λZap vector, 1x ligase buffer (New England Biolabs), 1 mM ATP,

and 200 U ligase (New England Biolabs). To calculate this molar ratio, Lambda Zap is roughly 38,900 base pairs, and 1 μg was used in this reaction. For ligating fragments between 200-350 base pairs (average size 275 bp), the following calculation is used to determine the molar ratio: 38,900bp/275bp = 141.5. 1μg/141.5 = 0.007 μg, or 7 ng of sample DNA for a 1:1 molar ratio. Therefore 14 ng would be added for 2:1 molar ratio. Gigapack III Gold packaging extract (Stratagene) was used to package the ligation for infection.

The XL1-Blue MRF strain of *E. coli* (Stratagene) was used for all infections prior to the excision stage. The library was plated at approximately 20,000 plaque-forming units (pfu) per 150 mm plate with NZY Agar (Sigma). Secondary screening was plated at approximately 100 pfu on 100 mm plates. Plaque lifts were done using 132 mm (for 150 mm plates) or 82 mm (for 100 mm plates) diameter Hybond-N nylon membranes (Amersham-Pharmacia Biotech). Filters were marked with three non-symmetrical cuts from a razor blade so that the plates could be lined up with the autoradiographs. Filters were denatured, neutralized, rinsed, and baked at 80°C as above.

*Probing*

Membranes were probed in the same manner as in the enrichment protocol (see above), however the same membranes were used for probing for both mixes. Lifts were first probed for dimer repeats, then stripped by incubating in 0.4N NaOH at 42°C with shaking for 20 minutes, then another 20 minutes with shaking in 100°C 1% SDS, and the probing process repeated for trimers and tetramers together. Positive plugs were cut out

using the wide end of yellow pipette tips and placed in 2 mL SM buffer with two drops of cholorophorm, which was then used for secondary screening. Since plaques were too dense to isolate individual signal-causing plaques on primary screening, secondary screening was done, and individual plaques isolated from there. Each plug from primary screening contained roughly 10 plaques, and the secondary plating then resulted in approximately 10% of all plaques giving a positive signal when probed. Since the secondary plating was done at a much lower concentration, individual positive plaques could be isolated without also getting negative plaques. Tertiary screening of four test samples showed 100% of plaques gave positive signals.

Inserts were recovered from the plasmid using Ex-Assist helper phage (Stratagene) and excision protocol (Short *et al.*, 1988), and were prepared for sequencing using the miniprep protocol of above. Size and quality of inserts were tested by digesting DNA samples with EcoR I and Pst I, which have restriction sites on opposite sides of the insert, and running the product on a 1% agarose gel. Concatamers were identified and not sequenced. Sequencing and primer selection were the same as for the enrichment protocol above.

## Genotyping

Primer3 (Rozen & Skaletsky, 2000) and Oligo (Rychlik, 1992) were used to design primer pairs that would not hybridize with each other, and the Oligonucleotide

Properties Calculator (Cao *et al.*, 2004) was used to confirm that primers had annealing temperatures within 2°C of each other.

Ten presumed unrelated individuals (i.e. no two adults nesting together) were used to assess allele frequencies, heterozygosity, and polymorphic information content (PIC) of the microsatellite loci using the genetic analysis program CERVUS (Marshall *et al.*, 1998). Initial screening was done by performing a PCR reaction in a 10 μL final volume using 50 ng DNA, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 1x buffer, 0.26 μM each primer, & 0.25 U Taq. Cycling conditions were 94°C for 2 minutes; 30 cycles of 94°C for 20 seconds, 55°C for 30 seconds, 72°C for 30 seconds; then a final 72°C for 2 minutes. 2 μL Orange G was added and the sample was run on a 1.5% agarose gel. If the product was faint or not present, the reaction was repeated using a lower annealing temperature (51°C instead of 55°C). If the product did not improve in appearance then the primers were not used for further analysis. Primers for one herring gull dimer locus from enrichment and one elaenia locus from Lambda Zap were dropped at this stage for failing to amplify a product on the agarose gels.

For the primers which did show a good product on the agarose gels, 10 μM of the forward primer was endlabeled using 12.5 U T4 polynucleotide kinase and gamma $^{33}P$ dATP. The mix was put in a PTC-200 Thermocycler at 37° for 30 minutes, then 64°C for 10 minutes. A PCR reaction was done in a 10 μL volume using 50 ng DNA, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 10 μM unlabeled reverse primer, 10 μM labeled forward primer, 0.05 U Taq DNA polymerase, and 1x buffer. The cycle conditions were 94°C for 2 minutes; 30 cycles of 94°C for 20 seconds, annealing temperature (52-55°C) for 30 seconds, 72°C for 30 seconds; then a final 72°C for 2 minutes. Amplification products

were resolved on a 6% denaturing polyacrylamide gel (Biorad Laboratories) run at 70 watts for 1.5-3 hours, depending on product size known from the sequenced clones. Gels were dried then placed on phosphorimager screens and exposed overnight. Screens were visualized using the PhosphorImager and ImageQuant software (Molecular Dynamics). Loci statistics were calculated using Cervus v2.0 (Marshall *et al.*, 1998).

Some of the primers amplified a product that contained more than two discreet bands. For these, a PCR gradient was run on four individuals using the same PCR reaction as above, but with the gradient function on the PTC-200 Thermocycler so that the annealing temperature ranged from 50 to 65°C. These were visualized as above.

**Statistics and Formulae**

Comparisons were made between the two methods using two-tailed paired t-tests for dimers and tri-/tetramers from the four species (N = 8).

Expected heterozygosity is calculated using the formula

$$He = 1 - \sum_{i=1}^{n} (f_i)^2$$

where n is the number of alleles and the i allele occurs with frequency fi (Nei, 1987).

Polymorphic information content is calculated by the formula

$$PIC = 1 - (\sum_{i=1}^{n} p_i^2) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_i^2 p_j^2$$

where $p_i$ is the frequency of the *i*th allele and n is the number of alleles (Botstein *et al.*, 1980).

Individual exclusion probability is calculated according to Marshall *et al.* (1998) using the formula

$$P_I = 1 - \{ \sum_{i=1}^{n} p_i^2 . p_i (2 - p_i) + \sum_{i \neq i}^{n} p_i p_j (p_i + p_j)(2 - p_i - p_j) \}$$

where $p_i$ is the frequency of allele *i* with a total of n alleles.

**Results:**

For all four species, each isolation method was successful in isolating at least one variable dimer locus, with enrichment isolating more for anis (4/5 loci) and elaenias (4/5), Lambda Zap more for herring gulls (5/8) and both isolating equal numbers of loci for pukekos (2/4). The number of alleles detected per locus ranged from 3-5 for anis, 2-7 for herring gulls, 2-6 for elaenias, and 3-4 for pukeko (Table 1).

It is generally thought that loci with heterozygosities below 0.5 are not very useful for parentage analysis (Marshall *et al.*, 1998). Three herring gull loci (LARSNX10B, LARZAP03, LARZAP19) and two elaenia loci (EFLSNX06, EFLSNX13) were well below this value. While heterozygosity is simply the proportion of individuals in a population that have 2 alleles at a given locus, polymorphic information content (PIC) gives an idea of allelic diversity, and is often more informative than heterozygosity. A population with most members sharing the same two alleles would have high heterozygosity but a low PIC. A locus with a PIC value below 0.25 is said to be only slightly informative, whereas a PIC greater than 0.5 is highly informative (Botstein *et al.*, 1980). All five ani loci have PIC above 0.5, and all but one pukeko locus (PPOSNX19) do as well. PPOSNX19 has a PIC value of 0.489, and so is still relatively informative. Two herring gull loci (LARSNX10B, LARZAP03) and one elaenia locus (EFLSNX13) have a PIC value below 0.25, with another two herring gull loci (LARZAP19, LARZAP26) and one elaenia locus (EFLSNX06) have a value between 0.25 and 0.5.

The average exclusion probability is the average probability of excluding a single unrelated candidate parent from parentage of a given offspring at one or more loci, assuming no typing errors occur (Marshall *et al.*, 1998). If one parent is known, the exclusion probability is calculated taking account of alleles that are unambiguously descended from the known parent. The total exclusionary power for the five ani microsatellites isolated in this comparison study is 0.850 if neither parent is known and 0.966 if one parent is known meaning 85% of all candidate parents can be eliminated if neither parent is known, and 96.6% is one parent is known. For the eight gull loci, the exclusionary power is 0.899 for the first parent and 0.985 for the second, while the power is 0.709 and 0.893 for the first and second parents respectively for the five elaenia loci, and 0.515 and 0.771 respectively for the four pukeko loci.

When a mutation occurs in one or both of the priming regions of a microsatellite, amplification might not occur. The result is called a null allele. An individual heterozygous with a null allele would appear homozygous for the allele that does amplify, and an individual homozygous for the null would have no bands at all (Callen *et al.*, 1993; Holm *et al.*, 2001). None of the loci differed statistically in their heterozygosities from the expected values calculated using the program Cervus v.2.0 (Marshall *et al.*, 1998), and therefore null alleles are not thought to be present for any locus in the individuals sampled. A small number of loci had only 9 of the 10 individuals amplify, but since the heterozygosities were not different from Hardy-Weinberg, this is likely a problem with the DNA rather than the result of null alleles. The locus CANSNX14 had 22% of the individuals not amplify, so new primers were designed that did not overlap with the previous pair. None of the individuals that failed to amplify with the original

primers amplified with this new set, while all the individuals that did amplify originally had the same genotypes with the new primers, suggesting that the non-amplifying individuals were not the result of null alleles. A small number of loci had observed heterozygosities higher than expected, which can be attributed to the small sample size and not a population-wide phenomenon. Because of the small sample size, again no significant differences were detected (all $p > 0.10$).

Loci isolated by the enrichment method were slightly more informative for smooth-billed anis, whereas Lambda Zap gave slightly better results for the other three species. However, because of the small sample sizes, no statistical differences were observed between enrichment and Lambda Zap in terms of number of alleles per locus ($p=0.48$, $\beta=0.8043$; figure 1a), mean expected heterozygosities ($p=0.24$, $\beta=0.7564$ ; figure 1b) or mean PIC ($p=0.29$,$\beta=0.8994$; figure 1c). The mean number of alleles was 3.67 and 4.1 for loci isolated from enrichment and Lambda Zap respectively. PIC was slightly higher for loci from Lambda Zap at a mean of 0.543 compared to 0.507 for enrichment, but again the difference was not significant. There were also no statistical differences between the number of alleles ($p=0.84$) or PIC ($p=0.43$) for the different species.

My original objective was to compare the two microsatellite isolation techniques in terms dollars per polymorphic microsatellite and hours per polymorphic microsatellite isolated. Because both methods failed to isolate any tri-/tetramers, and with the relatively small number of dimers isolated, a comparison using working loci yields insignificant

and inconclusive results ($p = 0.474$). Therefore comparisons focused on a stage prior to final establishment of polymorphic loci. Sequences that contained a repeat, even if it was too short to be variable (Lambda Zap) or lacked a flanking region (enrichment) were counted and used as a data point. The percentage of sequences containing a sequence repeated minimum three times in tandem was used to estimate the total number of positive signals that contain repeats. The comparison was therefore between the two methods for dollars per sequence with repeat (SWR) and hour per SWR. The total number of sequences with repeats was estimated from the total number of positive signals after screening by Southern blot hybridization, and multiplying by the proportion of positive signals sequenced that contain repeats. For example, ani dimer SWR were estimated at 56 since there were 220 positive signals, 25% of the minipreps showed the smear or banding pattern thought to indicate a microsatellite repeat, and 64% (7/11) of those sequenced contained a repeat ($220 \times 25\% \times 64\% = 56$). The numbers used for the following comparisons can be found in Table 1.

The monetary costs for each method are shown in Table 2. The prices given refer to the price at the time of purchase, which for all reagents was between January 2002 and August 2003. All prices are in US dollars. "Startup" refers to the initial purchase price of each item, whereas "per use" is the total cost divided by the number of times that item can be used. With the TOPO-TA cloning kit, the competent cells were the limiting component, so cost per use is calculated as the total cost ($310) divided by the number of vials of competent cells (20). This enrichment protocol can therefore be used 20 times before new supplies need to be purchased. Similarly, the λZAP phagemid runs out

before the other components of the Lambda Zap kit, and so cost per use is the number of aliquots of the λZAP vector. Enough phagemid for 10 uses of this protocol are included in the kit, so after 10 tries new supplies will need to be purchased. Enrichment has a total startup cost of $2179.00 and costs $84.62 per use of the protocol for isolating microsatellites. Lambda Zap on the other hand costs $2080.00 to startup, and $285.90 for each use.

Enrichment takes much less time to complete than does Lambda Zap (Table 3). Since several steps in both methods involve incubation periods or "wait steps", the total hands-on experimentation time as well as the total time from start to finish, including incubations, are given. These times only consider the steps that are different between the two methods, and hence start assuming the experimenter already has genomic DNA of known concentration, and do not include sequencing, primer selection, or microsatellite characterization. Without unexpected delays (see Appendix), enrichment can be completed in approximately eight days, requiring slightly more than 44 hours of hands-on experimentation. Lambda Zap takes more time to complete, roughly 19 days with 72 hours and twenty minutes of hands-on time.

*Enrichment*

A colony was considered to give a positive signal when a blot showed a dark spot on Kodak Biomax film following hybridization with radiolabeled probe. In total, enrichment yielded 220 positive signals for smooth-billed anis, 240 for herring gulls, 170 for yellow-bellied elaenias, and 200 for pukekos after probing the enriched clones with

radiolabeled (AC) and (AG) repeats. Positive signals were also visible when screened with the radiolabeled trimer and tetramer probes, with 120, 110, 130, and 110 positive signals for ani, gull, elaenia, and pukeko respectively (results were rounded down to the nearest 10 to avoid over-estimating based on the small number that literally fell in the gray area – signals that were intermediate between faint and bright). From these, 40 minipreps were done for each species for both dimer repeats and the tri-/tetramer repeats. A PCR reaction was run on the minipreps using the SNX primers and the repeat probe mix as primers. When these PCR products were visualized after being run on agarose gels, 16 ani, 16 gull, 18 elaenia, and 15 pukeko enrichment dimer samples showed a smear or second band, suggesting they contained a microsatellite repeat. All of the tri-/tetramers had faint smearing, so this method was not effective at screening for those. Ten of the minipreps that showed the smear or band pattern were sequenced for each species for dimers, with the exception of smooth-billed anis which had eleven sequences done. Ten tri-/tetramer minipreps were also chosen for sequencing. Two ani, two gull, and one elaenia sample that did not show the extra band or smear were also sequenced to confirm that lack of a smear or extra band was indicative of lack of repeat sequence. Of the minipreps for dimer repeats from enrichment, 64% (7/11) of the ani, 50% (5/10) of the gull, 60% (6/10) of the elaenia, and 40% (4/10) of the pukeko sequences contained a segment repeated at least six times in tandem (Table 1). None of the five minipreps that did not have the band or smear contained a repeat, suggesting that this is a good test to eliminate samples that do not contain repeats. 16 of the 22 dimer repeats isolated were (TG)/(AC). Five of the remaining six repeats were $(AG)_n$, and one gull locus was an (AT) repeat. This large bias in (AC) repeats isolated is consistent with estimates of (AC)

microsatellites accounting for over 60% of all dinucleotide repeats in avian genomes (Primmer *et al.*, 1997).

Many sequences for trimer and tetramer repeats isolated by enrichment contained one of the target repeats but with only one flanking region (N = 9). From the Lambda Zap sequences, several also contained repeats of one of the trimer or tetramer probe sequences, but repeated too few times to be variable (N = 14). (AGGG) was the most common tri- or tetrameric repeat (78.3%, N = 18), with (AGG) appearing on occasion (21.7%, N = 5) and no instances of (AAAG). In two separate cases one ani primer was designed using a sequence that had only one flanking region. A sequencing reaction was then performed using the one primer and sequencing dinucleotides (Invitrogen) in a PTC-200 Thermocycler and run on a polyacrylamide gel, but these attempts resulted only in a black smear throughout the entire gel.

Primers were designed for all the sequences with a dinucleotide repeated at least six times except for one gull sample. This gull dimer sequences did not have primers made because although the flanking regions were adequate in size, the flanking sequence itself was not suitable for primer design since primers could not be designed that would not hybridize with each other.

Four ani, three gull, four elaenia, and two pukeko enrichment dinucleotide loci were found to be polymorphic (Table 4).

Using the startup monetary cost for enrichment and the number of sequences with repeats, enrichment costs 38.89 dollars per microsatellite repeat sequence ($/μ) for ani dimer repeats, 45.37 $/μ for gulls, 47.34 $/μ for elaenias, and 72.59 $/μ for

pukeko, and 24.20 $/μ, 49.49 $/μ, 41.88 $/μ, and 98.98 $/μ for trimer and tetramer repeats respectively (Figure 2a).

Once the materials have all been purchased, however, they can be used multiple times, so an important value is also the cost per use. Despite the slightly higher startup cost, enrichment is much cheaper per use than Lambda Zap, at only $84.62 per use. This translates into a cost of only 1.50 $/μ for ani dimers, 1.76 $/μ, 1.84 $/μ, and 2.82 $/μ for gull, elaenia, and pukeko dimers respectively, and 0.94 $/μ, 1.92 $/μ, 1.63 $/μ, and 3.84 $/μ for trimers and tetramers for the four species respectively (Figure 2b).

Enrichment takes 44 hours and 15 minutes of hands-on time over 8 days when everything works properly to get from genomic DNA to plasmids ready for sequencing. Using the SWR values, this translates to 47.59 minutes per microsatellite (min/μ) for ani, 55.52 min/μ for gull, 57.93 min/μ for elaenia, and 88.83 min/μ for pukeko dimers, and 29.61 min/μ, 60.57 min/μ, 51.25 min/μ, and 121.14 min/μ for trimers and tetramers for the four species in the same order for hands-on time (Figure 3a). The total time from start to finish, including incubation periods and wait steps for enrichment is approximately 219 min/μ, 255 min/μ, 266 min/μ, and 408 min/μ for dimers for the four species, and 136 min/μ, 278 min/μ, 235 min/μ, and 556 min/μ for trimers and tetramers respectively (Figure 3b).

*Lambda Zap*

The Lambda Zap method yielded 100, 120, 120, and 90 positives for ani, gull, elaenia, and pukekos when screened with radiolabeled dimer probes. When they were screened with the radiolabeled trimer and tetramer probes, 33, 41, 36, and 32 positives were seen for the four species respectively. 40 minipreps were done for each species for the dimer repeats, while only 30 were done for the tri-/tetramers because of their smaller number of positive signals. Using these minipreps as template DNA in a PCR reaction with the Lambda Zap primers as well as the repeat probe was not effective. None of the samples showed the extra band or smear that was visible with the enrichment samples and the SNX primers. Therefore this method of screening is not effective with minipreps from Lambda Zap. Ten of these minipreps for dimers and five for tri-/tetramers were sequenced for each species. 30% (3/10) of the ani, 80% (8/10) of the gull, 50% (5/10) of the elaenia, and 50% (5/10) of the pukeko dimer samples contained sequences repeated six times in tandem (Table 1). Again, most of the dimer repeats isolated were (AC)/(TG) repeats, with only one exception, a herring gull (AG) repeat. Some of the sequences for tri-/tetramers contained the probed sequences (Table 1), but always repeated too few times to be variable. All of the sequences with repeats contained either (AGG) or (AGGG) repeated three or four times, which are usually not long enough to show polymorphism (Tautz, 1989; Tautz & Schlötterer, 1994).

All sequences that had a dimer repeated at least six times in tandem had primers designed for them, with the exception of one elaenia sequence, whose flanking sequences would not support primer design. One pair for a pukeko trinucleotide $[(AGG)_4]$ and one for a pukeko tetranucleotide $[(AGGG)_4]$ were also designed.

From the primer pairs designed, the pukeko trinucleotide and tetranucleotide both amplified a product but were monomorphic, as was expected for these shorter repeats. Microsatellites are believed to mutate via strand-slippage (Schlötterer & Tautz, 1992; Goldstein & Schlötterer, 1999), with shorter repeats being less prone to this mechanism than longer ones (Pupko & Graur, 1999).

Lambda Zap gave only one polymorphic dimer locus for both anis and elaenias, five for gulls, and two for pukekos (Table 4).

Using the startup cost of $2080.00, Lambda Zap costs 69.77 $/μ, 21.80 $/μ, 34.89 $/μ, and 46.51 $/μ for ani, gull, elaenia, and pukeko dimer repeats respectively, and 80.51 $/μ, 63.43 $/μ, 95.14 $/μ, and 110.17 $/μ for trimer and tetramers (Figure 2a).

The per use cost of Lambda Zap is much higher than it is for enrichment, at $285.90 per use. Lambda Zap is therefore much more expensive per use than enrichment for microsatellites isolated, costing 9.53 $/μ, 2.98 $/μ, 4.77 $/μ, and 6.35 $/μ, for dimers, and 11.00 $/μ, 8.66 $/μ, 13.00 $/μ, and 15.05 $/μ for trimers and tetramers for anis, gulls, elaenias, and pukekos respectively, an average of nearly seven times as much for Lambda Zap as for enrichment (Figure 2b).

Lambda Zap hands-on time is approximately 71 hours and 20 minutes over a period of 19 days, which is equivalent to 142.67 min/μ, 44.58 min/μ, 71.33 min/μ, and 95.11 min/μ for dimers, and 164.62 min/μ, 129.70 min/μ, 194.55 min/μ, and 225.26 min/μ for trimers and tetramers for anis, gulls, elaenias, and pukekos respectively (Figure 3a). Lambda Zap, using total time, takes 888 min/μ, 278 min/μ, 444 min/μ, and 592

min/μ for dimers, and 1025 min/μ, 807 min/μ, 1211 min/μ, and 1402 min/μ for trimers and tetramers respectively for the four species (Figure 3b).

*Comparison*

Using a two-tailed paired t-test, there is no significant difference between enrichment and Lambda Zap in terms of number of microsatellite repeats isolated ($t = 0.59$, $t\ crit = 2.36$, $p = 0.57$) (Figure 4). With this sample size and standard deviation, we would have been able to detect a difference of 1.7-fold with a power of 0.700, which is the accepted convention for minimum power. Since each method is statistically as good as the other at isolating sequences with repeats, monetary and time investments should be the main factors in deciding which method to use in order to isolate microsatellites. Using the \$/μ calculated from the startup cost, there is no statistical difference between the two methods ($t = -1.13$, $t\ crit = 2.36$, $p = 0.30$). Enrichment, however, is significantly cheaper per microsatellite repeat isolated using the per use cost ($t = -4.19$, $t\ crit = 2.36$, $p = 0.004$). Therefore if several isolation runs are going to be performed, either for multiple species or to isolate many microsatellites for the same species, perhaps of different sequences or lengths, then enrichment is the better method.

Hands-on time is much lower for isolating microsatellites by enrichment versus Lambda Zap, which results in enrichment taking significantly less time per microsatellite isolated ($t = -3.27$, $t\ crit = 2.36$, $p = 0.01$). If a technician is being paid by the hour, then enrichment is a much more cost effective method for isolating microsatellites. Total time

is also much lower for enrichment, which again results in enrichment taking significantly less total time per microsatellite isolated ($t = -4.13$, $t\ crit = 2.36$, $p < 0.01$).

It is important to note however that comparisons between the actual number of polymorphic microsatellite loci isolated, rather than the sequences with repeats, yield no significant results in terms of \$/μ or min/μ. These comparisons could not be made for trimeric and tetrameric repeats since none was isolated, but for dimeric repeats the comparisons are non significant using paired t-tests for \$/μ for startup cost ($t = -1.37$, $t\ crit. = 3.18$, $p = 0.26$) and per use cost ($t = -2.45$, $t\ crit. = 3.18$, $p = 0.09$), or for min/μ using hands-on time ($t = -2.11$, $t\ crit. = 3.18$, $p = 0.13$) or total time ($t = -2.87$, $t\ crit. = 3.18$, $p = 0.06$). With these sample sizes and standard deviations, we would need a 2.96-fold difference in start-up \$/μ and a 6.41-fold difference in total time for a power of 0.700. Since the number of microsatellites isolated was so small, ranging from two to four for enrichment dimers and from one to five for Lambda Zap dimers, comparing the actual number of variable microsatellite loci isolated would not yield meaningful results since the numbers are small and they do not represent typical values that can be achieved using these protocols. Using the sequences with repeat (SWR) values instead of the total number of polymorphic microsatellites isolated should still provide a non-biased means of comparing the methods, because if the troubleshooting steps (appendix 1) were used to improve the isolation capabilities of the protocols then the number of sequences with repeats would likely be a more accurate estimate of total microsatellites that could be isolated.

## Discussion:

On average, nearly 20% of the colonies following enrichment contained a microsatellite repeat. This is comparable to other published values (Zane *et al.*, 2002). The high number of positives in this study shows why an enrichment protocol such as selective hybridization or a method that allows for the screening of hundreds of thousands of clones ,such as Lambda Zap, is so important. The typical range of positives for non-enriched DNA, or using the traditional shotgun cloning method, is 0.13-4.5% for mammals, and an even less productive 0.025-1.7% for birds. Gibbs *et al.* (1998b) found only 42 positive clones out of 40,000 colonies (0.1%) when isolating microsatellites from the common cuckoo, and of these, only seven gave polymorphic microsatellite loci (0.018% of colonies). While isolating microsatellites from the ostrich genome, Tang *et al.* (2003) found that an enrichment protocol increased the proportion of clones containing (CA) repeats from 0.4% for shotgun cloning to 78.8%.

There exist the possibility of over-estimating the number of microsatellite repeats using the number of sequences with repeats rather than total number of unique polymorphic microsatellite loci, including not accounting for redundant inserts, ones that gave a positive signal because of biased nucleotide composition rather than a microsatellite repeat (Chenuil *et al.*, 2003), and concatamers. These problems will be ignored however for this study, since there is no reason why one method should have a higher degree of these biases than the other. Based on the results using the number of sequences with repeats, enrichment using selective hybridization appears to be the better

option for isolating microsatellites from avian genomes. Although the startup cost is slightly more expensive for this enrichment protocol than for Lambda Zap, the lower per use cost and shorter time frame make it a better alternative. Also, nearly 20% of the startup cost for enrichment is spent on the magnetic particle collection unit. This unit can be used over and over again, however, and thus this cost needs only to be spent once for unlimited number of runs with the enrichment protocol. The most expensive perishable component of the enrichment protocol is the competent cells in the TOPO-TA cloning kit, which will need to be replaced after 20 enrichments. Competent cells can be produced in-house, although the One Shot competent cells provided with the kit (Invitrogen) have a transformation efficiency of approximately $1x10^9$ colony forming units/μg DNA, much higher than the normal for homemade cells (Shuman, 1994). The kits from both methods should provide more than enough material to isolate microsatellites for multiple species of interest, or repeated cloning attempts for fewer species. If using Lambda Zap, however, additional Lambda vector will be needed after only 10 uses, and all components of the startup cost are consumed during the procedure. Therefore if microsatellites are to be isolated more than 10 times using the same protocol, then the "material replacement cost" of enrichment is much lower than that of Lambda Zap. Lambda Zap's per use cost is also increased by the need to do two, or often three screenings using radiolabeled probes. Enrichment is also a significantly quicker method, taking only about eight days versus nineteen days for Lambda Zap to go from genomic DNA to minipreps ready for sequencing. If the steps that were not included in the comparison are taken into consideration, namely DNA extractions, sequencing, primer design, and characterization of loci, in less than two weeks it is possible to go from tissue or blood samples to

polymorphic microsatellite loci with enrichment, whereas it would take over three weeks with Lambda Zap. Enrichment also is much more straightforward and simple to follow, and therefore more suited towards students with little or no cloning experience. Several steps in the Lambda Zap method require testing different components to make sure they are at the proper concentration (the optical density (OD) of the cells needs to be tested and then diluted on four separate occasions in one run of λ Zap), and the plates often need to be checked to avoid overgrowing the plaques. Enrichment is a more forgiving method, where the cells can be left overnight without worry of overgrowing. I found the protocol of Glenn & Schable (2003) much easier to follow and far less confusing than the λ Zap manual. The mathematical calculations of enrichment are also much simpler than those of λ Zap. Enrichment does not have any calculations beyond basic dilutions. Lambda Zap, however, requires some more advanced calculations on top of the dilutions, such as a 1:1 molar ratio of sample DNA to λ. This varies depending on the size of the insert being used (7 ng of DNA for 200-350 bp inserts versus 10.9 ng for 350-500). Enrichment therefore is better suited for researchers with less lab experience, whereas Lambda Zap is much more geared towards those already familiar with cloning techniques.

An additional advantage for the enrichment protocol rather than Lambda Zap is the proven and consistent success of enrichment, particularly from taxa that are often troublesome including birds (e.g. Tang *et al.*, 2003; Schlosser *et al.*, 2003; Martínez-Cruz *et al.*, 2002). The protocol of Glenn & Schable (2003) or similar enrichment protocols (see below) have been used to isolate polymorphic microsatellite loci applicable to population genetic studies from several avian orders, including different species from the

four orders used in this study: Gruiformes (e.g. Chbel *et al.*, 2002; Lieckfeldt *et al.*, 2001), Cuculiformes (e.g. Muniz *et al.*, 2003); Passeriformes (e.g. Kawano, 2003; Frentiu *et al.*, 2003), and Charadriiformes (e.g. Given *et al.*, 2002; Crochet *et al.*, 2003). Microsatellites have also been isolated for most other avian orders using an enrichment protocol (e.g. Tang *et al.*, 2003; Schlosser *et al.*, 2003). Lambda Zap, on the other hand, has only been used successfully for very few avian species (Blanchard & Quinn, 2001; Hughes & Deloach, 1997), however both published uses were successful at isolating trinucleotide repeats, which are generally more polymorphic than dinucleotides (Schug *et al.*, 1998), while most enrichment publications report isolating dimer repeats, usually $(AC)_n$, with longer repeats being rare. This increased number of successes using enrichment over Lambda Zap could simply be the result of a bias in terms of greater use of enrichment. Based on the cost and time investments, however, a bias towards enrichment is justified.

Both methods were successful in isolating dinucleotide repeats, however in this study they both failed to isolate any variable tri- or tetranucleotide loci. The reasons for the failure to isolate suitable trimer and tetramer repeats are thought to differ depending on the isolation method. Enrichment isolated several long trimer or tetramer repeats (9 of 21 sequences from the four species), but often there was no flanking region on one end of the repeat, and thus no primer could be synthesized. This lack of flanking sequence may be because some of the repeat probe used during the enrichment acted as a primer in the PCR reaction prior to cloning. If the enriched DNA is amplified using the repeat probe as one of the primers, the insert in the cloning reaction would not have any of the unique

flanking sequence on one end since the PCR amplification would have begun at the repeat rather than at the linker upstream. Our results were consistent with this. Additionally, Koblížková *et al.* (1998) showed that if the probes were biotinylated on the 3' end rather than the 5', thereby prohibiting the probe from extending along the 3' end since the biotin is in the binding site where new nucleotides would be added during replication, then this problem is prevented. This is thought to only have happened with the longer repeats because the shorter (TG) and (TC) repeats would have a much lower annealing temperature due to their lower G/C content when compared to the trimer (AGG) and the tetramer (AGGG), and thus would not anneal during the PCR reactions using the annealing temperature of 60°C (Cao *et al.*, 2004).

Another possible cause for this lack of flanking region on one side of the microsatellite repeat would be if the restriction enzyme recognition sequence was too similar to the repeat sequence. This is most likely to be an issue if searching for tetramers and using a four-cutter restriction enzyme. Brad White (pers. com.) found that using Sau 3A1 which recognizes ^GATC, and probing for a (GATA) repeat, often only one flanking region was present. After switching to Hae III, an enzyme which recognizes GG^CC, a sequence very different from the probe sequence, this problem was eliminated. In theory, one quarter of all microsatellite repeats of the (AGAT) variety will have a cytosine as the first nucleotide in the flanking region following the repeat, thereby creating a Sau 3A1 recognition site. This recognition site immediately following the repeat was therefore most likely the reason for the problem in their case. This would predict ¼ of the repeats would be affected. This is not a likely explanation for the poor results in this comparison study, however, since the restriction enzymes used in this

experiment were Xmn I and Rsa I, which cut GAANN^NNTTC and GT^AC respectively. Xmn I in theory could cause this problem with one out of every 64 (GAAA) repeats, which was a sequence included in the probe mix, but no GAAA repeats were isolated in the limited number of samples sequenced, and the problem was present for two repeats very different from the recognition sites: (AGGG) and (AGG). The easiest way to avoid this problem would therefore be to use probes biotinylated on the 3' end, and to use restriction enzymes that recognize sequences very different from the probe repeat sequences.

The repeats isolated from Lambda Zap had flanking sequence on each end of the repeat, but the repeat was usually too short to be of any use. Most repeats were present only three or four times in tandem, which are rarely if ever variable. The reason such short repeats were isolated is believed to be that the stringency of the probing was too low. All probing was done at 60°C with 2xSSC, 0.1%SDS. These conditions are stringent enough to remove shorter dimer repeats so that (AC) and (AG) repeats that were obtained were long enough to be useful, but a repeat such as (AGGG) has a much higher G/C content, and therefore a higher annealing temperature. The Oligonucleotide Properties Calculator (Cao *et al.*, 2004) shows that $(AGGG)_6$ has an annealing temperature of 75°C, so at the stringency used in this experiment, both longer and shorter repeats were bound to the radiolabeled probe. Therefore it is likely that some of the sequences isolated do in fact contain longer trimer and tetramer repeats, but because these are less common than shorter repeats, only shorter runs were found in the few sequences done. A method to test this hypothesis would be to take the Hybond filters, strip them,

and then probe for the repeats using a hybridization temperature closer to 75°C. In theory many of the positives signals from the 60°C probing should no longer be present, but a smaller number should remain. The plaques corresponding to these positives could then be sequenced. Unfortunately, since this test was conceived several months after the cloning, the agar plates and Hybond filters had all been discarded. More sequencing could be done, however, to determine if longer repeats are also in fact present.

Both methods used would realistically yield far fewer variable microsatellites than the estimated number of sequences with repeats used in this study. Redundancy in repeats (i.e. the same stretch of DNA being sequenced from more than one clone) is a common occurrence in both methods (Leanne Blanchard, pers. com; Blanquer *et al.*, *in press*). The general rule of thumb for microsatellites is the rule of half (L. De Sousa, pers. com.; Casey & Burnell, 2001). Briefly, for every positive signal, only half will contain a microsatellite repeat of adequate length. Of those, only half will have decent flanking regions, and of the primer pairs that are designed, only half will show variability. The result of all these halves is approximately one out of every eight positive signals will result in a primer pairs that amplify a polymorphic microsatellite, which is similar to what was found in this study. Using the Lambda Zap protocol to isolate microsatellite repeats from the smooth-billed ani genome, Blanchard (2000) found roughly 400 variable intensity TG plaques. After secondary and tertiary screening, this number was down to 73 positives. 41 of these were prepared for sequencing using a standard alkaline lysis protocol, and after determining size of inserts, 31 were eventually sequenced. Only 12 of these were adequate for primer design, and of these, only three

were polymorphic, or 4% of the positives after tertiary screening and fewer than 10% of the total number of samples sequenced.

Andrea Blanquer (pers. com.) and Cecile Perrin (pers. com.) both found that enrichment had very low success on species low in microsatellites. Blanquer found that very few positives were isolated from two species of sponges (*Scopalina sp.*), and of those that were isolated many were the same locus. Perrin & Roy (2000), using sea urchins (*Evechinus chloroticus*) as their focal species, also found very few loci with enrichment, even after trying different probe cocktails and restriction enzymes. They eventually developed a new protocol by modifying an earlier version of Glenn & Schable's (2003) protocol and adding an extra PCR step. To hybridize the DNA to the probes, 100-500 ng of amplified DNA was mixed with 5'-biotinylated repeat probes in 20 μL extension solution (0.2 mM each dNTP, 2 mM $MgCl_2$, 1 x Taq buffer, one probe, and 0.5 U Taq DNA polymerase). The mixture was subjected to one round of PCR at 94°C for 5 minutes, 1 minute at 55°C, then 1-10 minutes at 72°C. Products were purified with High Pure PCR product purification kit (Roche) and eluted in 30 μL $dH_2O$. This greatly improved the success rate of the enrichment protocol (Perrin & Roy, 2000).

Many of the enrichment clones that do not contain microsatellite repeats were still likely carried through the enrichment because of their sequences had at least some binding affinity to the biotinylated probes. Many of the clones displayed traces of sequence simplicity, or sequences of highly biased nucleotide composition (Tautz, 1989). These biases would result in the probes binding to the sequence non-specifically, but enough for them to remain bound during the reduced stringency of the washes in the

enrichment protocol. Partial binding of the repeat probe to these sequences likely explains why they were not removed during the washes. This partial binding in enrichment could also explain why a sequence that was not in the probe mix was isolated. One instance of an (AT) repeat was found, which was unexpected since there was no probe for this repeat in the enrichment mix. Chenuil *et al.* (2003) also reported this phenomenon. They followed an enrichment protocol using a probe mix containing dimers, trimers, and tetramers, and found that only one of the 26 clones containing microsatellites had a repeat motif that was used for enrichment. They were even successful in isolating repeats of a size that was not included in the mix. The same partial binding can still explain why these were found. For example, the probe $(AAAT)_8$ could have 62.5% complementary binding sites with the repeat (TTTTG), which was a repeat isolated by Chenuil *et al.* (2003). The (AT) repeat isolated in this study likely bound to one of the dimer probes since every other nucleotide matched (the 'T' of the 'AT' repeat bound to the 'A' of the 'AG' or 'AC' probe). If washes were done at a higher stringency then this repeat likely would not have been isolated, however at higher stringencies there is also a higher likelihood of washing away desired sequences. The stringency was chosen because after more stringent washes, no DNA was recovered.

Selective hybridization, the form of enrichment used here, has become the most commonly used method for microsatellite isolation. As of March, 2001, 25% of all reviewed primer notes from the journals Molecular Ecology and Animal Genetics employed this technique (Zane *et al.*, 2002). There are several other enrichment protocols that have been shown to be quite successful at isolating microsatellites. The

following is a brief list of some of the more common protocols, and a small description as to what makes each of them different from the others. The method of Glenn & Schable (2003) is the enrichment method used in this experiment, so further detail will not be given regarding it.

**Primer extension** is a method that has been proposed by Ostrander *et al.* (1992) and Paetkau (1999). Both involve the construction of a primary genomic library where DNA fragments with an average size of less than 500 bp is inserted into a phagemid. It is then amplified to obtain closed circular single-stranded DNA. This ssDNA then acts as a template using repeat-specific oligonucleotides as primers, thus generating a double stranded product only for vectors that contain the repeat. These primer-extended products are then recovered either by transforming into a specific strain of *E. coli* (*dut*+ *ung*+) (Ostrander *et al.*, 1992) or by performing a second primer extension using 5' biotinylated oligonucleotides and Klenow DNA polymerase followed by recovery with straptavidin-coated magnetic beads (Paetkau, 1999), resulting in an enriched library. Primer extension is rarely used to isolate microsatellites, likely a result of its high number of steps, but it is worth noting that it has been successful at isolating microsatellites from avian genomes (e.g. Maak *et al.*, 2003).

Shibata *et al.* (2003) describe a much simpler method that requires only three basic techniques: PCR, cloning, and sequencing. Briefly, 500 ng of high molecular weight genomic DNA is amplified by PCR using 250 pmol of AC-repeat primer $(AC)_{10}N_4$. The product is then ligated into a cloning vector (pDrive, Qiagen), and the

ligation products transformed into *Epicurian coli* XL2-Blue MRF' ultra-competent cells (Stratagene). Plasmid DNAs are then amplified using the Templiphi DNA amplification kit (Amersham Biosciences) in order to generate template DNA for sequencing reactions. Sequencing is then done using M13 universal primers. This results in sequences containing two microsatellite repeats with unique (i.e. non-repetitive) DNA sequence between them. A pair of locus-specific primers is designed in order to amplify the sequence flanking the repeats. Template DNA is made by digesting 100 ng of genomic DNA with a restriction enzyme that will not cut the sequence between the pair of repeats (one of Sau3AI, TaqI, PstI, or EcoRI). This cut DNA is then self-ligated using 50 U of T4 DNA ligase at 16°C for 18 hours. 10 ng of this circularized genomic DNA is then amplified with the Templiphi amplification kit to yield more than 1 μg of the template DNA. Following the PCR, approximately 1 μL of the product is ligated into pDrive and transformed into *Escherichia coli* DH5α competent cells. Plasmid DNA is then sequenced using either the M13-20 or the M13-RV primer to determine the 5'-upstream sequences of the repeat region of each microsatellite locus. Locus-specific forward primers can then be designed. This method is was highly effective at isolating microsatellites from the Japanese squirrel (Shibata *et al.*, 2003), with 20 of the 22 clones sequenced containing microsatellite repeats, and 11 of those 20 being variable (2-7 alleles, mean $H_O = 0.45$, mean $H_E = 0.62$ for 10 presumably unrelated individuals). This method has the added advantage of not requiring any probing, and estimates on the cost for startup are right around $1000 US. This price does not include sequencing, however, which needs to be done on two separate occasions for this method. This isolating technique also has the added advantage of being relatively quick method, going from

genomic DNA to sequenced DNA in less than one week. Surprisingly, Shibata *et al.* (2003) also isolated repeats that differed from the (AC) repeats used as probes. Ten of the 20 clones sequenced contained (CT)n, (GA)n, $(GGT)_n$, or a compound microsatellite. Some of these unexpected repeats were shown to be variable, including a $(CT)_{21}$ repeat and a compound pentameric repeat $(CCAGG)_3$(GGCAGG)(GGCAAGG). One downfall of this technique, however, is its rather expensive per use cost. It only costs slightly more than $1000 CDN for startup, but the Templiphi DNA amplification kit (Amersham) that it uses can only be used twice, or sometimes only once depending on the number of microsatellites required, before needing to be reordered. Shibata *et al.* (2003) used the entire kit in one run of this protocol, and were successful in isolating 11 polymorphic microsatellite loci with a range of two to seven alleles and an average observed heterozygosity of 0.455. This method is therefore not necessarily the best choice if microsatellites for many species are to be isolated, and it has yet to be shown how well it works on taxa with low microsatellite frequency, which would have on average a much greater distance between microsatellite repeats. Another possible downside of this technique is that because it is such a new technique, as of June 2004 not a single paper published in Molecular Ecology, Molecular Ecology Notes, or Animal Genetics, three journals most commonly used for publishing microsatellite primers (Zane *et al.*, 2002), has cited this paper. It is therefore impossible to show how effective it is across different taxa. This is important since results for microsatellite isolation success are often sporadic even within classes, and there can be even more variability across taxa such as birds and mammals, even when using the same technique (Zane *et al.*, 2002; Primmer *et al.*, 1996).

For now, however, it appears to be a good alternative for isolating microsatellites from taxa abundant in microsatellite repeats.

Traditional shotgun cloning is not a viable choice when microsatellites are needed from species whose genomes are low in microsatellite abundance, as it generally has low efficiency and can be quite time consuming. This comparison study shows that an enrichment procedure like that of Glenn and Schable (2002) requires less time and is less expensive than using a Lambda Zap phage vector for cloning, yet yields as many sequences with repeats. Enrichment also has the advantage of requiring only basic skills in molecular biology. New microsatellite isolation techniques are becoming increasingly common in the published literature, yet most involve some form of enrichment using biotinylated probes. Whether this trend continues or if another more efficient method comes along is yet to be seen, but for now the enrichment protocol of Glenn and Schable (2002) appears to be the best option.

**Literature Cited:**

Armour, J.A., R. Neumann, S. Gobert, & A.J. Jeffreys. 1994. Isolation of human simple repeat loci by hybridization selection. *Genomics* 29: 136-144.

Bernatchez, L. & P. Duschene. 2000. Individual-based genotype analysis in studies of parentage and population assignment: how many loci, how many alleles? *Canadian Journal of Fisheries and Aquatic Sciences* 57: 1-12.

Blanchard L. & J.S. Quinn. 2001. The characterization of microsatellite loci in the communally breeding smooth-billed ani (*Crotophaga ani*). *Molecular Ecology Notes* 1: 152-154.

Blanchard, L. 2000. An investigation of the communal breeding system of the smooth-billed ani (*Crotophaga ani*). PhD Thesis. McMaster University, Hamilton.

Botstein, D., R.L. White, M. Skolnick, & R.W. Davis. 1980. Construction of a genetic linkage map in man using restriction length polymorphisms. *American Journal of Human Genetics* 32: 314-331.

Callen, D.F., A.D. Thompson, Y. Shen, H.A. Phillips, R.I. Richards, J.C. Mulley, & G.R. Sutherland. 1993. Incidence and origin of "null" alleles in the (AC)n microsatellite markers. *American Journal of Human Genetics* 52: 922-927.

Cao, Q., W.A. Kibbe, E. Buehler, & B. Somers. 2004. Oligonucleotide Properties Calculator version 3.06. http://www.basic.northwestern.edu/biotools/oligocalc.html.

Casey, D.G. & A.M. Burnell. 2001. The isolation of microsatellite loci in the Mediterranean fruitfly *Ceratitis capitata* (Diptera: Tephritidae) using a biotin/streptavidin enrichment technique. *Molecular Ecology Notes* 1(3): 120-122.

Chambers, G.K. & E.S. MacAvoy. 2000. Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology Part B* 126: 455-476.

Chbel, F., D. Broderick, Y. Idaghdour, A. Korrida, & P. Mccormick. 2002. Characterization of 22 microsatellites loci from the endangered Houbara bustard (*Chlamydotis undulata undulata*). *Molecular Ecology Notes* 2: 484-487.

Chenuil, A., M. Le Gac, & M. Thierry. 2003. Fast isolation of microsatellite loci of very diverse repeat motifs by library enrichment in echinoderm species, *Amphipolis squamata* and *Echinocardium cordatum*. *Molecular Ecology Notes* 3: 324-327.

Crochet, P.-A., J.Z. Chen, J.-M. Pons, J.-D. Lebreton, P.D.N. Hebert, & F. Bonhomme. 2003. Genetic differentiation at nuclear and mitochondrial loci among large white-headed gulls: sex-biased interspecific gene flow? *Evolution* 57(12): 2865-2878.

Davis, J. N. & J.S. Quinn. 1997. Distribution of parental investment and sibling competition in the herring gull, *Larus argentatus*. *Behaviour* 134: 961-974.

Dib C., S. Faure, C. Fizames, D. Samson, N. Drouot, A. Vignal, P. Millasseau, S. Marc, J. Hazan, E. Seboun, M. Lathrop, G. Gyapay, J. Morissette, & J. Weissenbach. 1996. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380: 152-154.

Fernàndez-López, L., E. Piñeiro, R. Marcos, A. Velázquez, & J. Surrallés. 2004. Induction of instability of normal length trinucleotide repeats within human disease genes. *Journal of Medical Genetics* 41: e3.

Frentiu, F.D., C.L. Lange, T. Burke, & I.P.F. Owens. 2003. Isolation of microsatellite loci in the Capricorn silvereye, *Zosterops lateralis chlorocephalus* (Aves: Zosteropidae). *Molecular Ecology Notes* 3: 462-464.

Gibbs, H.L., K. Prior, & C. Parent. 1998a. Characterization of DNA microsatellite loci from a threatened snake: the eastern massasauga rattlesnake (*Sistrurus c. catenatus*) and their use in population studies. *Journal of Heredity* 89(2): 169-173.

Gibbs, H.L., L. De Sousa, K. Marchetti, & H. Nakamura. 1998b. Isolation and characterization of microsatellite DNA loci from an obligate brood parasitic bird, the common cuckoo (*Cuculus canorus*). *Molecular Ecology* 7: 1437-1439.

Gilbertson, M. & R. Hale. 1974. Pollutants in breeding herring gulls in the Lower Great Lakes. *Canadian Field-Naturalist* 88: 354-356.

Given, A.D., J.A. Mills, & A.J. Baker. 2002. Isolation of polymorphic microsatellite loci from the red-billed gull (*Larus novaehollandiae scopulinus*) and amplification in related species. *Molecular Ecology Notes* 2: 416-418.

Glenn & Schable, 2003. http://www.uga.edu/srel/DNA_Lab/protocols.htm.

Goldstein, D.B. & C. Schlötterer. 1999. Microsatellites: evolution and applications. Oxford University Press.

Gregory, T.R. 2002. A bird's-eye view of the C-value enigma: genome size, cell size, and metabolic rate in the class Aves. *Evolution* 56(1): 121-130.

Hamilton, M.B., E.L. Pincus, A. Di Fiore, & R.C. Fleischer. 1999. Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *BioTechniques* 27: 500-507.

Holm, L.E., V. Loeschcke, & C. Bendixen. 2001. Elucidation of the molecular basis of a null allele in a rainbow trout microsatellite. *Marine Biotechnology* 3(6): 555-560.

Hughes, C.R. & D.M. Deloach. 1997. Developing microsatellites when they are rare: trinucleotide repeat loci in the northern mockingbird. *Molecular Ecology* 6: 1099-1102.

Jamieson, I.G., J.S. Quinn, P.A. Rose & B.N. White. 1994. Shared paternity among non-relatives is a result of an egalitarian mating system in the communally breeding bird, the Pukeko. *Proceedings of the Royal Society of London, Series B* 257: 271-277.

Jamieson, I.G. & J.S. Quinn. 1997. Problems with removal experiments designed to test the relationship between paternity and parental effort in a socially polyandrous bird. *Auk* 114: 291-295.

Jarne, P. & P.J.L. Lagoda. 1996. Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution* 11(10): 424-429.

Jones, A.G. & W.R. Arden. 2003. Methods of parentage analysis in natural populations. *Molecular Ecology* 12: 2511-2523.

Kawano, K.M. 2003. Isolation of polymorphic microsatellite markers in the great tit (*Parus major minor*). *Molecular Ecology Notes* 3: 314-315.

Keith, J.A. 1966. Reproduction in a population of herring gulls (*Larus argentatus*) contaminated by DDT. *Journal of Applied Ecology* 3: 57-70.

Knapik E.W., A. Goodman, M. Ekker, M. Chevrette, J. Delgado, S. Neuhauss, N. Shimoda, W. Driever, M.C. Fishman, & H.J. Jacob. 1998. A microsatellite genetic linkage map for zebrafish (*Danio rerio*). *Nature Genetics* 18: 338-343

Koblížková, A., J. Doležel, & J. Macas. 1998. Subtraction with 3' modified oligonucleotides eliminates amplification artifacts in DNA libraries enriched for microsatellites. *BioTechniques* 25: 32-38.

Lieckfeldt, D., A. Schmidt, & C. Pitra. 2001. Isolation and characterization of microsatellite loci in the great bustard, *Otis tarda*. *Molecular Ecology Notes* 1: 133-134.

Longmire, J.L., D.C. Hahn, & J.L. Roach. 1999. Low abundance of microsatellite repeats in the genome of the brown-headed cowbird (*Molothrus ater*). *The Journal of Heredity* 90(5): 574-578.

Maak, S., K. Wimmers, S. Weigend, & K. Neumann. 2003. Isolation and characterization of 18 microsatellites in the Peking duck (*Anas platyrhynchos*) and their application in other waterfowl species. *Molecular Ecology Notes* 3: 224-227.

Marshall, T.C., J. Slate, L.E.B. Kruuk, & J.M. Pemberton. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* 7: 639-655.

Martínez-Cruz, B., V.A. David, J.A. Godoy, J.J. Negro, S.J. O'Brien, & W.E. Johnson. 2002. Eighteen polymorphic microsatellite markers for the highly endangered Spanish imperial eagle (*Aquila adalberti*) and related species. *Molecular Ecology Notes* 2: 323-326.

McConnell, S.K.J., D.A. Dawson, A. Wardle, & T. Burke. 1999. The isolation and mapping of 19 tetranucleotide microsatellite markers in the chicken. *Journal of Avian Biology* 32(1): 76-78.

Moore S.S., L.L. Sargeant, T.J. King, J.S. Mattick, M. Georges, & D.J. Hetzel. 1991. The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. *Genomics* 10(3):654-660.

Muniz, L.S.B., R.H. F. Macedo, & J. Graves. 2003. Isolation and characterization of dinucleotide microsatellite loci in communally breeding Guira cuckoos (Aves: Cuculidae). *Molecular Ecology Notes* 3: 209-211.

Neff, B.D. & M.R. Gross. 2001. Microsatellite evolution in vertebrates: inference from AC dinucleotide repeats. *Evolution* 55(9): 1717-1733.

Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Ostrander, E.A., P.M. Jong, J. Rine, & G. Duyk. 1992. Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences. *Procedings of the National Academy of Sciences of the United States of America* 89: 3419-3423.

Paetkau, D. 1999. Microsatellites obtained using strand extension: an enrichment protocol. *BioTechniques* 26: 690-697.

Pekarik, C. & D.V. Weseloh. 1998. Organochlorine contaminants in Herring Gull eggs from the Great Lakes, 1974-1995: change point regression analysis and short term regression. *Environmental Monitoring and Assessment* 53: 77-115.

Perrin, C. & M.S. Roy. 2000. Rapid and efficient identification of microsatellite loci from the sea urchin, *Evechinus chloroticus*. *Molecular Ecology* 9: 2221-2223.

Primmer, C.R., A.P. Møller, & H. Ellegren. 1995. Resolving genetic relationships with microsatellite markers: a parentage testing system for the swallow *Hirundo rustica*. *Molecular Ecology* 4:493-498.

Primmer, C.R., A.P. Møller, & H. Ellegren. 1996. A wide-range survey of cross-species microsatellite amplification in birds. *Molecular Ecology* 5: 365-378.

Primmer, C.R., T. Raudsepp, B.P. Chowdhary, A.P. Møller, & H. Ellegren. 1997. Low frequency of microsatellites in the avian genome. *Genome Research* 7: 471-482.

Pupko, T. & D. Graur. 1999. Evolution of microsatellites in the yeast *Sacharomyces cerevisiae:* role of length and number of repeat units. *Journal of Molecular Evolution* 48: 313-316.

Queller, D.C., J.E. Strassmann, & C.R. Hughes. 1993. Microsatellites and kinship. *Trends in Ecology and Evolution* 8(8): 285-288.

Quinn, J.S. & J.M. Startek-Foote. 2000. Smooth-billed ani (*Crotophaga ani*). In *The Birds of North America*, (A. Poole and F. Gill, eds.). The Birds of North America, Inc., Philadelphia, PA.

Rassmann, K., C. Schlötterer, & D. Tautz. 1991. Isolation of simple-sequence loci for use in polymerase chain reaction-based DNA fingerprinting. *Electrophoresis* 12: 113-118.

Rozen, S. & H.J. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp 365-386

Rychlik, W. 1992. Oligo 4.06 Primer Analysis Software. National Biosciences Inc. Plymouth, MN. USA.

Sambrook, J., E.F. Fritsch, & T. Maniatis. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor. Laboratory Press, New York.

Schlosser, J.A., T.W.J. Garner, J.M. Dubach, & A.G. McElligott. 2003. Characterization of microsatellite loci in Humboldt penguin (*Spheniscus humboldti*) and cross-species amplification in other penguin species. *Molecular Ecology Notes* 3: 62-64.

Schlötterer, C. & D. Tautz. 1992. Slippage synthesis of simple sequence DNA. *Nucleic Acids Research* 20(2):211-215.

Schlötterer, C. 2001. Geneological inference of closely related species based on microsatellites. *Genetical Research* 78: 209-212.

Schug, M.D., C.M. Hutter, K.A. Wetterstrand, M.S. Gaudette, T.F.C. MacKay, & C.F. Aquadro. 1998. The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. *Molecular Biology and Evolution* 15(12): 1751-1760.

Schuler, G.D., M.S. Boguski, E.A. Stewart, L.D. Stein, G. Gyapay, *et al.* 1996. A gene map of the human genome. *Science* 274: 540-546.

Shibata, K., K. Bandoh, N. Yaekashiwa, T. Matsuzaka, & H.B. Tamate. 2003. A simple method for isolation of microsatellites from the Japanese squirrel, *Sciurus lis*, without constructing a genomic library. *Molecular Ecology Notes* 3: 657-658.

Short, J.M., J.M. Fernandez, J.A. Sorge, & W.D. Huse. 1988. λ ZAP: a bacteriophage λ expression vector with *in vivo* excision properties. *Nucleic Acids Research* 16(15): 7583-7600.

Tang, B., Y.H. Huang, L. Lin, X.X. Hu, J.D. Feng, P. Yao, L. Zhang, & N. Li. 2003. Isolation and characterization of 70 novel microsatellite markers from ostrich (*Struthio camelus*) genome. *Genome* 46: 833-840.

Tarof, S.A., L.M. Ratcliffe, & P.T. Boag. 2001. Polymorphic microsatellite loci for assigning parentage in least flycatchers (*Empidonax minimus*). *Molecular Ecology Notes* 1: 146-148.

Tautz, D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* 17: 6463-6471.

Tautz, D. & C. Schlötterer. 1994. Simple sequences. *Current Opinion in Genetics and Development* 4(6): 832-837.

Tucker, G. M. & M.F. Heath. 1994. Birds in Europe: their conservation status. BirdLife Conservation Series No. 3. Cambridge: BirdLife International.

Yauk & Quinn. 1996. Multilocus DNA fingerprinting reveals high rate of heritable genetic mutation in herring gulls nesting in an industrialised urban site. *Proceedings of the National Academy of Sciences of the United States America* 93: 12137-12141.

Yauk, C.L. & J.S. Quinn. 1999. Genetic structure among breeding herring gulls (*Larus argentatus*) from the Great Lakes and Eastern Canada. *Journal of Great Lakes Research* 25(4): 856-864.

Yauk, C.L., G.A. Fox, B.E. McCarry, & J.S. Quinn. 2000. Induced minisatellite germline mutations in herring gulls (*Larus argentatus*) living near steel mills. *Mutation Research* 452(2):211-218.

Zane, L., L. Bargelloni, & T. Patarnello. 2002. Strategies for microsatellite isolation: a review. *Molecular Ecology* 11: 1-16.

**Table 1.** Number of colonies, positive signals after probing, proportion of positives following PCR on minipreps, proportion of positives that contain a sequence repeated at least six times in tandem (except for Lambda Zap tri-/tetramers, which were considered positive with a target sequence repeated at least three times in tandem), proportion of those that were polymorphic, and estimated total number of sequences with repeats (SWR). (CAN = *Crotophaga ani*; LAR = *Larus argentatus*; EFL = *Elaenia Flavogaster*; PPO = *Porphyrio porphyrio*).

| | | CAN Enrichment | CAN λZap | LAR Enrichment | LAR λZap | EFL Enrichment | EFL λZap | PPO Enrichment | PPO λZap |
|---|---|---|---|---|---|---|---|---|---|
| No. Colonies/Plaques | dimers | 1,400 | 150,000 | 1,240 | 150,000 | 800 | 150,000 | 920 | 150,000 |
| | tri-/tetramers | 1,200 | 150,000 | 1,080 | 150,000 | 1,120 | 150,000 | 880 | 150,000 |
| No. Positive Signals | dimers | 220 | 100 | 240 | 120 | 170 | 120 | 200 | 90 |
| | tri-/tetramers | 120 | 33 | 110 | 41 | 130 | 36 | 110 | 32 |
| % of +ve Preps following PCR | dimers | 40 (16/40) | -- | 40 (16/40) | -- | 45 (18/40) | -- | 37.5 (14/40) | -- |
| | tri-/tetramers | -- | -- | -- | -- | -- | -- | -- | -- |
| % +ve Signals Having Repeats | dimers | 64 (7/11) | 30 (3/10) | 50 (5/10) | 80 (8/10) | 60 (6/10) | 50 (5/10) | 40 (4/10) | 50 (5/10) |
| | tri-/tetramers | 75 (4/6) | 80 (4/5) | 40 (2/5) | 80 (4/5) | 40 (2/5) | 60 (3/5) | 20 (1/5) | 60 (3/5) |
| % Variable Repeats | dimers | 57 (4/7) | 33 (1/3) | 75 (3/4) | 62.5 (5/8) | 75 (4/6) | 25 (1/4) | 50 (2/4) | 40 (2/5) |
| | tri-/tetramers | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Estimated Total Number SWR | dimers | 56 | 30 | 48 | 96 | 46 | 60 | 30 | 45 |
| | tri-/tetramers | 90 | 26 | 44 | 33 | 52 | 22 | 22 | 19 |

**Table 2.** Startup and per use monetary costs for enrichment via selective hybridization and Lambda Zap.

a) Selective Hybridization

| *Material* | Startup (US $) | Per use (US $) |
|---|---|---|
| Probes | 404.00 | 1.35 |
| Restriction enzymes | 105.00 | 4.65 |
| Linkers | 31.00 | 0.08 |
| Ligase | 67.00 | 2.57 |
| Taq | 52.00 | 0.26 |
| Straptavidin beads | 141.00 | 3.45 |
| Magnetic Particle Collecting Unit | 347.00 | *0.00* |
| TOPO-TA cloning kit | 310.00 | 15.50 |
| Buffer Ingredients | 94.00 | 0.39 |
| $^{32}P$ | 217.00 | 21.70 |
| Hybond-N 82 mm nylon membranes | 70.00 | 11.16 |
| Ampicillin | 40.00 | 0.16 |
| Agar | 49.00 | 1.95 |
| Agarose | 233.00 | 2.79 |
| Plates | 19.00 | 18.60 |
| | | |
| Total | 2179.00 | 84.62 |

b) Lambda Zap

| *Material* | Startup (US $) | Per use (US $) |
|---|---|---|
| Lambda Zap Express kit | 969.00 | 96.89 |
| Ligase | 67.00 | 0.64 |
| Restriction enzymes | 169.00 | 7.44 |
| Probes (for screening) | 35.00 | 0.12 |
| $^{32}P$ | 217.00 | 43.41 |
| Hybond-N 82 mm nylon membranes | 70.00 | 20.93 |
| Hybond-N 132 mm nylon membranes | 70.00 | 70.00 |
| Buffer ingredients | 138.00 | 3.18 |
| Tetracycline | 26.00 | 0.03 |
| Agar | 49.00 | 5.37 |
| Agarose | 233.00 | 0.93 |
| Plates | 37.00 | 37.20 |
| | | |
| Total | 2080 | 285.90 |

*Does not include DNA extractions, sequencing, or primer design and testing since these are common to both isolation methods.

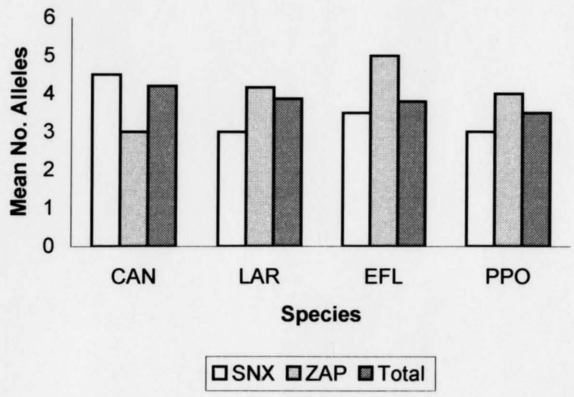
**Table 3.** Summary of the hands-on time required for the major steps of microsatellite isolation for enrichment via selective hybridization and cloning using Lambda Zap.

| | Selective Hybridization (minutes) | Lambda Zap (minutes) |
|---|---|---|
| Preparing plates, bacterial stocks, digests | 235 | 390 |
| Agarose Gels | 165 | 335 |
| PCR | 40 | 0 |
| Dynabead enrichment | 185 | 0 |
| Packaging | 0 | 240 |
| Cloning (without phage) | 240 | 0 |
| Plaque/Colony Lifts | 60 | 240 |
| Probing | 130 | 290 |
| Excision | 0 | 515 |
| Minipreps | 680 | 680 |
| Test Restriction | 0 | 50 |

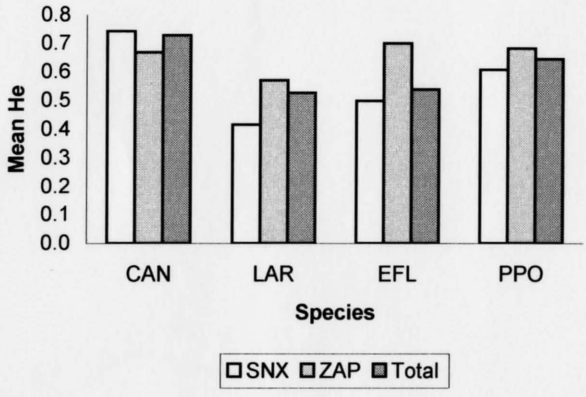
**Table 4**. Microsatellite markers isolated from the four species with observed heterozygosity ($H_O$) and expected heterozygosity ($H_E$).

| Locus | Primer Sequences | Repeat Motif | Clone Size (bp) | No. of Alleles | $H_O$ | $H_E$ |
|---|---|---|---|---|---|---|
| CANSNX07 | F: GCCGTCTTCACTGACTCGTT<br>R: ATCAGACAGAAGAGGCATAA | $(TG)_{15}$ | 156 | 5 | 0.900 | 0.647 |
| CANSNX14 | F: TGCCAGAATTAAGATGCCAAC<br>R: CTCCTAAACACTGGGCTGAA | $(AC)_{21}(TC)_3$ | 308 | 6 | 0.524 | 0.717 |
| CANSNX17 | F: GTCTCTGGCCGTCTTCACTG<br>R: GGTAAGTTTCCCACAAGATCA | $(TG)_{15}$ | 102 | 4 | 0.900 | 0.695 |
| CANSNX18 | F: AACACGTTTGCTGTCCCTTC<br>R: TGATCTTGTGGGAAACTTACC | $(TG)_{14}$ | 269 | 5 | 1.000 | 0.784 |
| CANZAP22 | F: ATGGACTCATGCCAAAGTAAGC<br>R: TTGTCTGGTTTTTCATCACAGG | $(AC)_{12}(CA)_2$ $C(CA)_9$ | 209 | 4 | 0.313 | 0.619 |
| LARSNX01 | F: GCTGATAGTCCCTTTAGTAGAC<br>R: TATTTTTAGCCGTCCTCAATGG | $(AC)_n$, $(TG)_n$, $(AT)_n$ | 266 | 3 | 0.300 | 0.279 |
| LARSNX10B | F: ATCAATTTCCTGTTGCTGGTCT<br>R: GCTTGGTGTGCATATTTGTGAT | $(AT)_7$ | 133 | 2 | 0.200 | 0.189 |
| LARSNX24 | F: GGCAGGATTGGTCTTGAAAAC<br>R: TAGCCGGGACCACGATACAAC | $(GT)_9$ | 235 | 4 | 0.800 | 0.689 |
| LARZAP11 | F: TTGGACTCAGTGCCCTCTTC<br>R: GGGCAGAAGCTCATGTGTCTA | $(GT)_{13}$ | 211 | 4 | 0.700 | 0.658 |
| LARZAP12 | F: CAGAATTGAAAATGTACAGCTC<br>R: TGGAAGTATGTGGGTTCCTGT | $(GA)_{12}$ | 196 | 6 | 0.889 | 0.850 |
| LARZAP14 | F: TTGTATCAAAAATCCATTAAAAGG<br>R: TTCCCGAGTAAACATGGCTTT | $(TG)_{15}$ | 181 | 7 | 0.800 | 0.853 |
| LARZAP19 | F: AGGAAACGAACTCCCTGACAT<br>R: TTCTGGCTTTAATTCTCAGTCTT | $(GT)_{10}$ | 217 | 3 | 0.300 | 0.358 |
| LARZAP26 | F: CCCCTCTTTGCCAGCATTG<br>R: TGGTTGTCTTTTGTCCCATGTG | $(GT)_{12}$ | 168 | 4 | 0.500 | 0.500 |
| EFLSNX06 | F: TCAGCGTTGGAGCTAGGAATA<br>R: CAGAAATGAAACTGTGAGGAAG | $(TC)_7$ | 224 | 3 | 0.300 | 0.279 |
| EFLSNX13 | F: TCTCTTGATTCATTCAGTGGACACC<br>R: AACTCTTTTGCTCTCTCCCTATACC | $(AC)_8$ | 156 | 2 | 0.111 | 0.111 |
| EFLSNX16 | F: CCTTTGCAAAACCGGGTCTG<br>R: TTTTCTTATATCTATTGAGAGATGGT | $(TG)_2TT$ $(TG)_8$ | 185 | 3 | 0.900 | 0.647 |
| EFLSNX22 | F: CCCGGGAAAGGCTTCGTCTTC<br>R: GGAGATTTTATATCGGTGGC | $(AG)_{13}$ | 283 | 5 | 0.556 | 0.752 |
| EFLZAP27 | F: GTGTCAGAGCAAGGCAGT<br>R: GTGCTCACGTGCAGATCA | $(TG)_9$ | 102 | 5 | 0.900 | 0.742 |
| PPOSNX18 | F: CAAGAATGTGGATGAGGAGACA<br>R: TTTCTTCATTTCAGAATGCCAG | $(CT)_8$ | 198 | 3 | 0.556 | 0.582 |
| PPOSNX19 | F: TGTGGGAACTCAGTGTTACAGG<br>R: AGCCAACAGAATTAAAGGGACA | $(AC)_9$ | 173 | 3 | 0.500 | 0.426 |
| PPOZAP04 | F: CAGCAGCTGATGTGCCGTGAAG<br>R: CGATTCCCTGCCTGATTCACA | $(AC)_9$ | 217 | 4 | 0.400 | 0.647 |
| PPOZAP15 | F: TTTAACCTCATCAGGATTGT<br>R: GTAACATCAGCAGTTCTTCAC | $(AT)_4(AC)_7$ | 204 | 4 | 0.500 | 0.711 |

a)




b)




c)

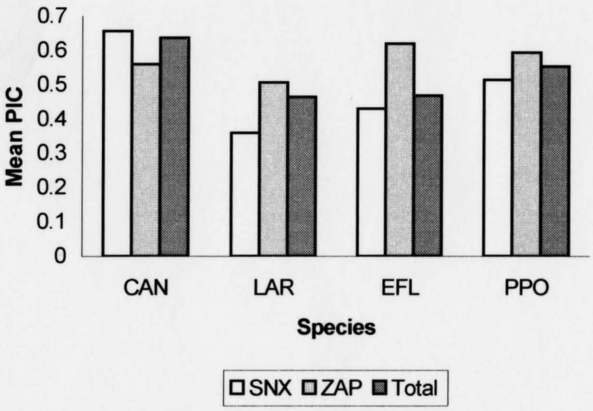
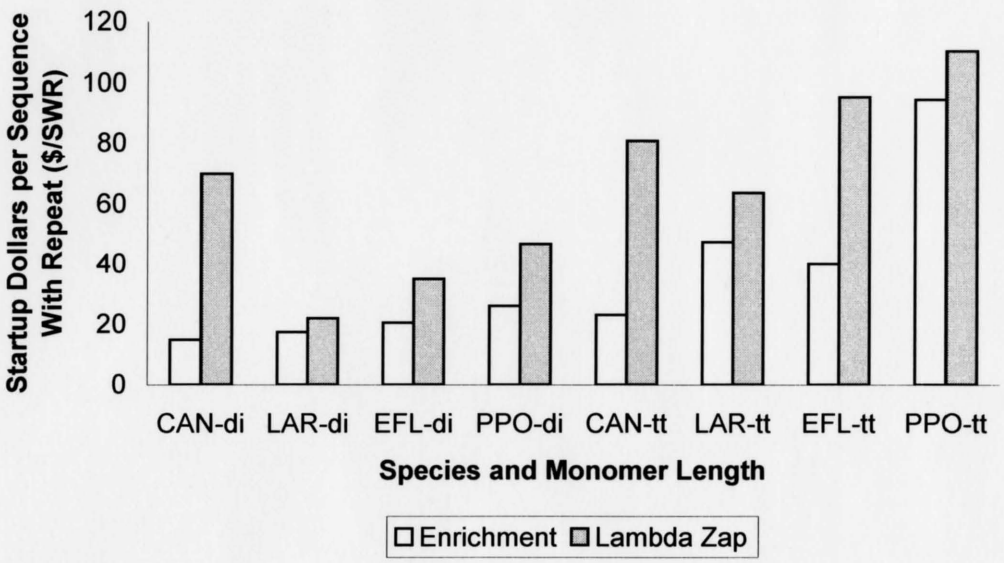



Figure 1. Mean number of alleles (a), expected heterozygosity (b) and polymorphic information content (c) for all polymorphic microsatellite loci isolated by enrichment (SNX) and Lambda Zap (ZAP) and the two methods combined (Total) compared between species. (CAN = *Crotophaga ani*; LAR = *Larus argentatus*; EFL = *Elaenia Flavogaster*; PPO = *Porphyrio porphyrio*).

a)




b)

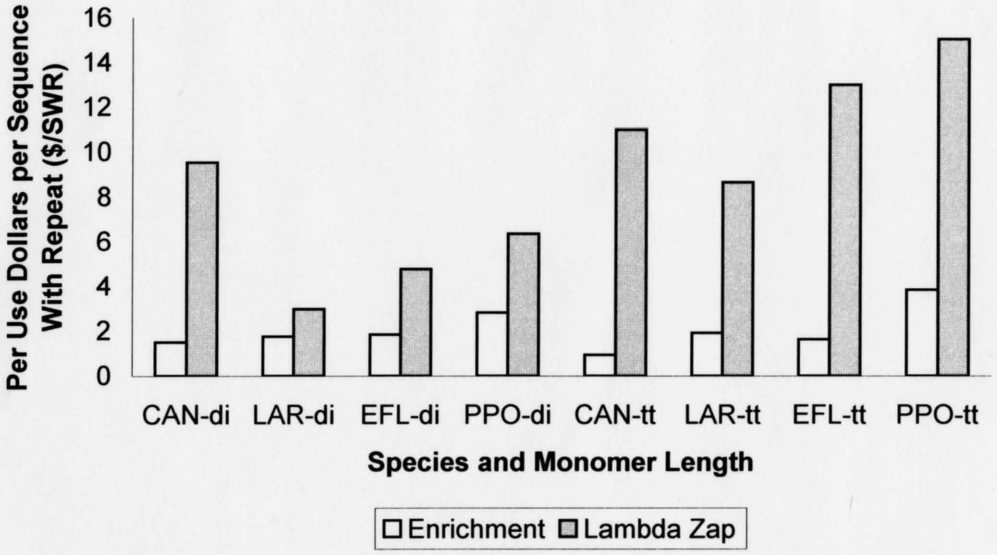
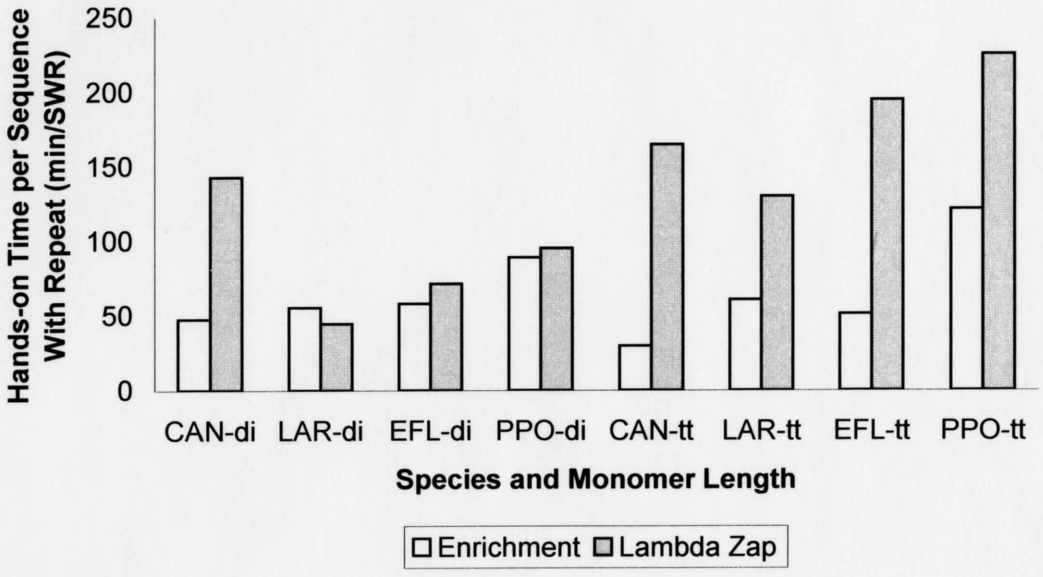



Figure 2. Dollars per microsatellite for each species and both the dimer (di) and tri-/tetramer (tt) probe mixes using a) the startup costs, and b) the per use costs. (CAN = *Crotophaga ani*; LAR = *Larus argentatus*; EFL = *Elaenia Flavogaster*; PPO = *Porphyrio porphyrio*).

a)




b)

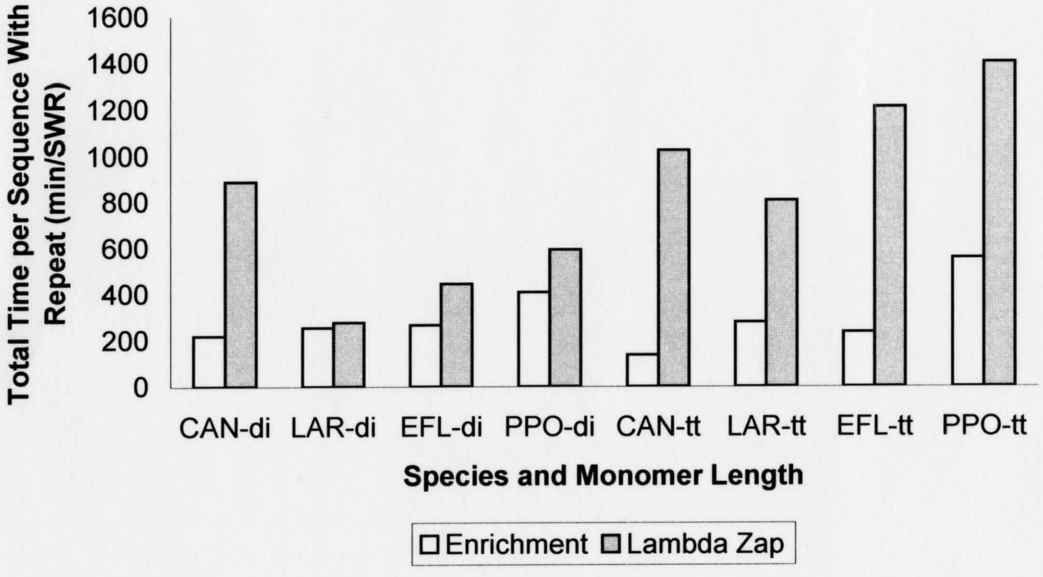



Figure 3. Minutes per Sequence With Repeat for each species and both the dimer (di) and tri-/tetramer (tt) probe mixes using a) the hands-on time, and b) the total time, including incubation periods. (CAN = *Crotophaga ani*; LAR = *Larus argentatus*; EFL = *Elaenia Flavogaster*; PPO = *Porphyrio porphyrio*).

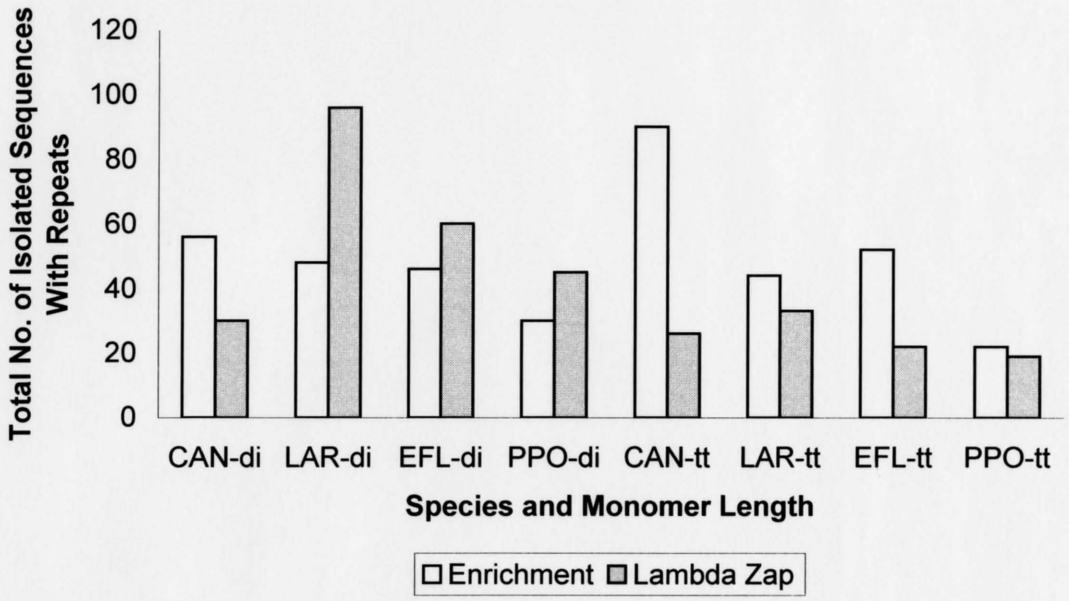
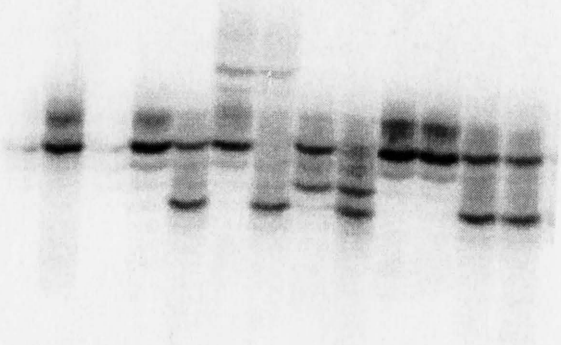


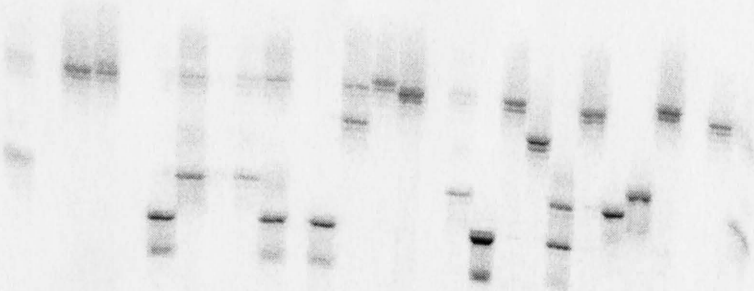

Figure 4. The total number of sequences with repeats isolated for each species using the two methods for dimer (di) and tri/tetramer (tt) repeats. (CAN = *Crotophaga ani*; LAR = *Larus argentatus*; EFL = *Elaenia Flavogaster*; PPO = *Porphyrio porphyrio*).

a)



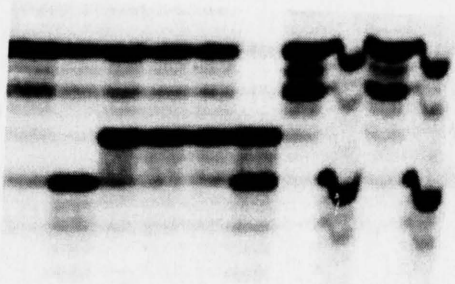
CANSNX07

b)



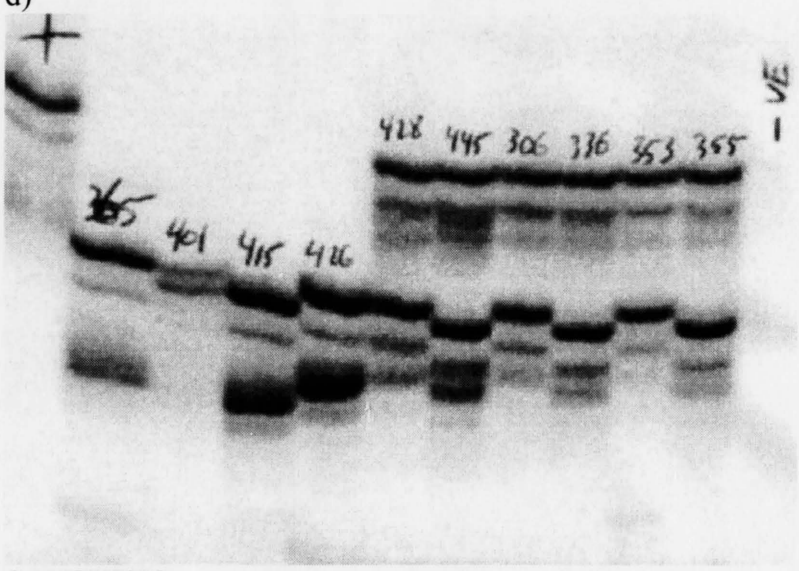
CANSNX14

c)



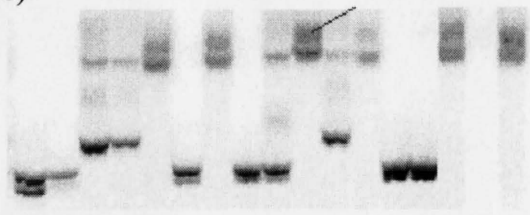
CANSNX17

d)




CANSNX18

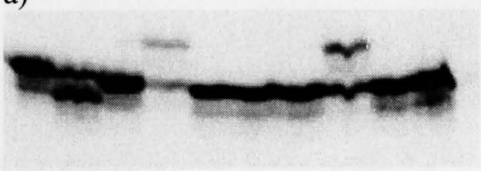
e)



CANZAP22

Figure 5. Scans of polyacrylamide gels showing the five polymorphic microsatellites isolated for the smooth-billed ani.

a)



LARSNX01

b)

LARSNX10B

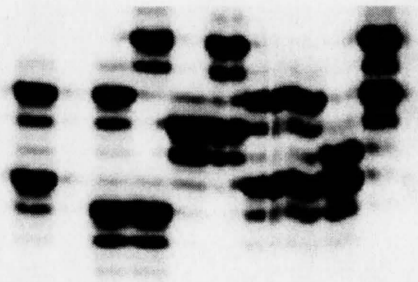
c)

LARSNX24

d)

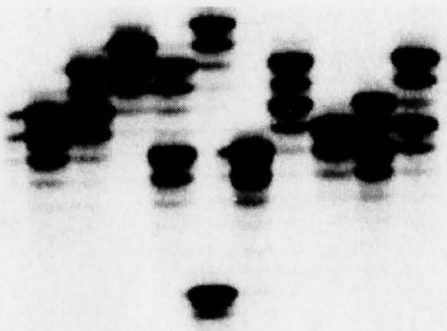
LARZAP11

e)



LARZAP12

f)

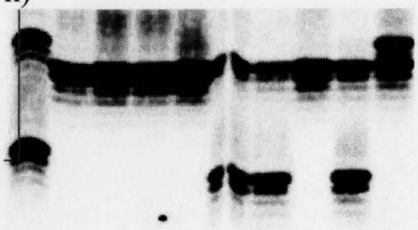


LARZAP14

g)

LARZAP19

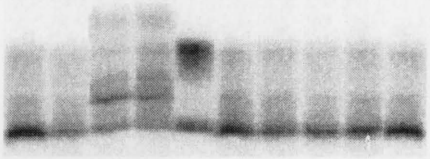
h)



LARZAP26

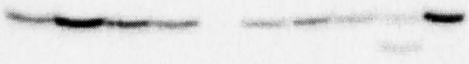
Figure 6. Scans of polyacrylamide gels showing the eight polymorphic microsatellites isolated for the herring gull.

a)



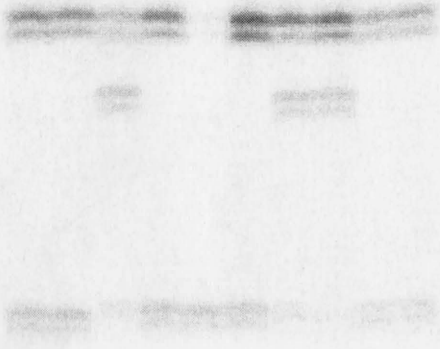
EFLSNX06

b)



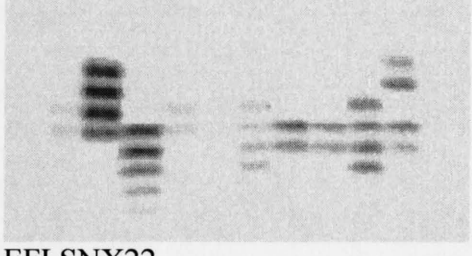
EFLSNX13

c)



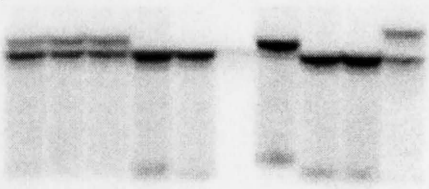
EFLSNX16

d)



EFLSNX22

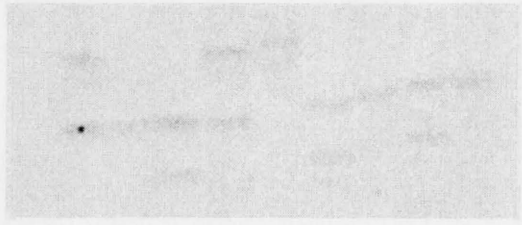
e)



EFLZAP27

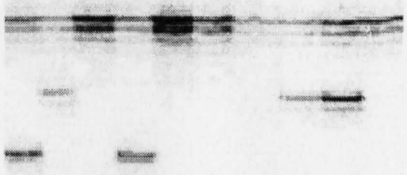
Figure 7. Scans of polyacrylamide gels showing the five polymorphic microsatellites isolated for the yellow-bellied elaenia.

a)



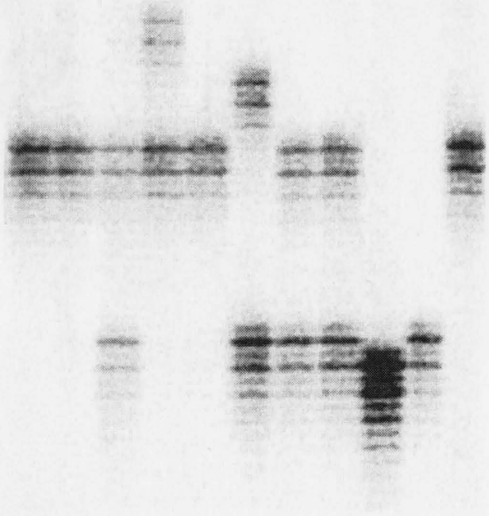
PPOSNX18

b)



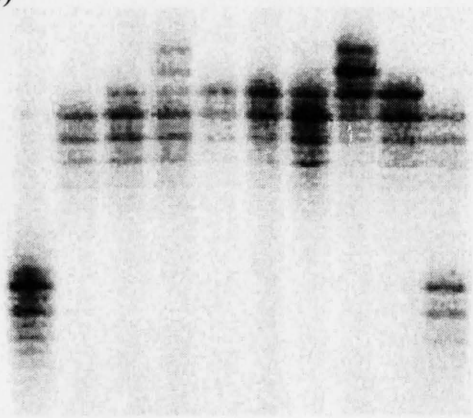
PPOSNX19

c)



PPOZAP04

d)



PPOZAP15

Figure 8. Scans of polyacrylamide gels showing the four polymorphic microsatellites isolated for the pukeko.

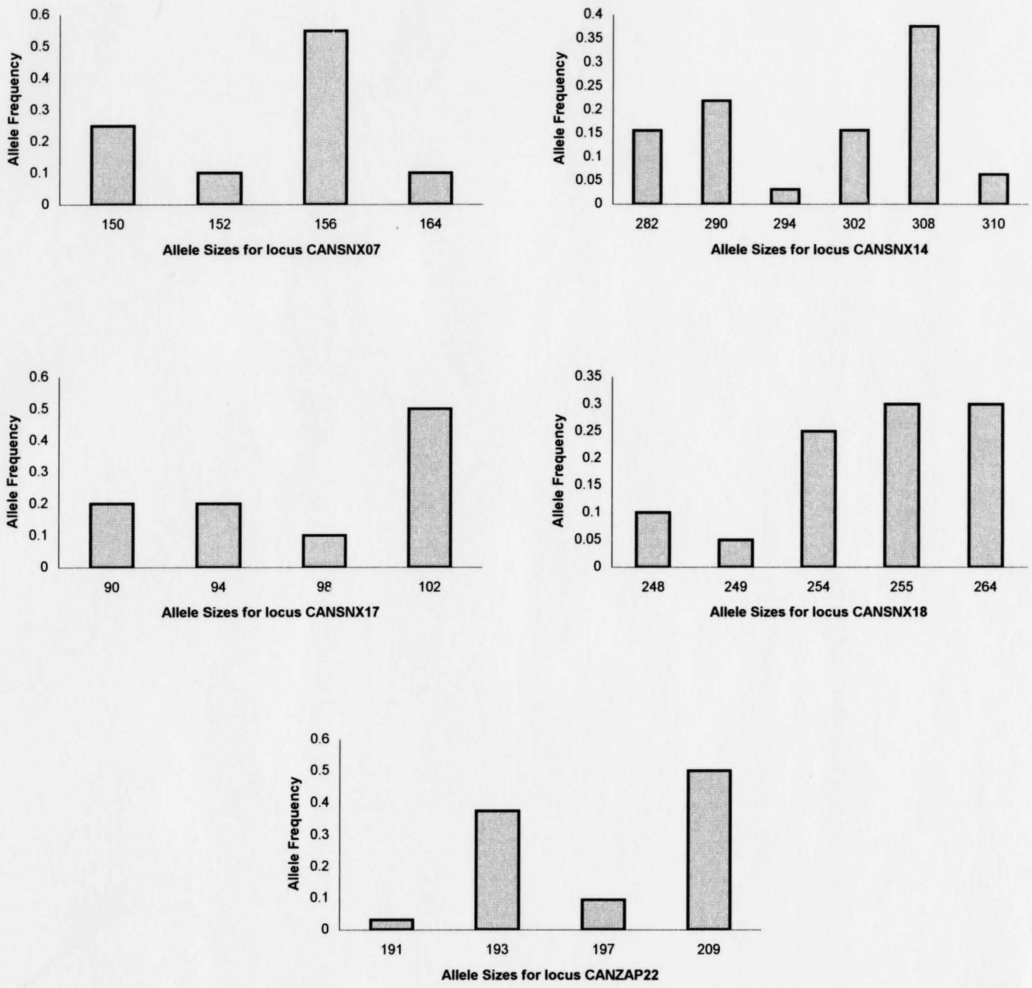
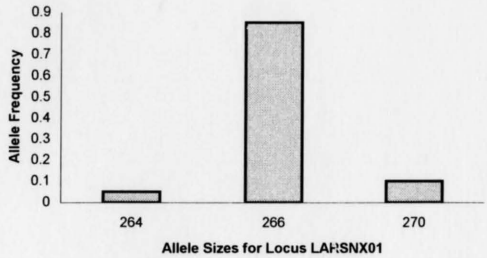
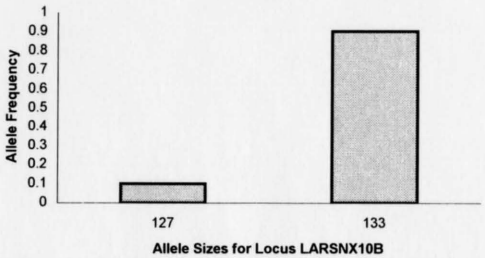
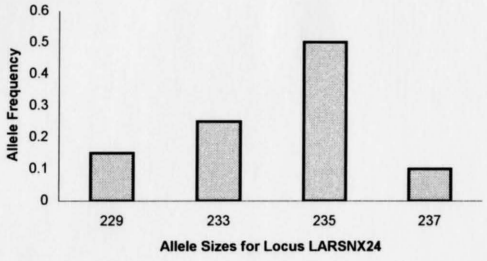
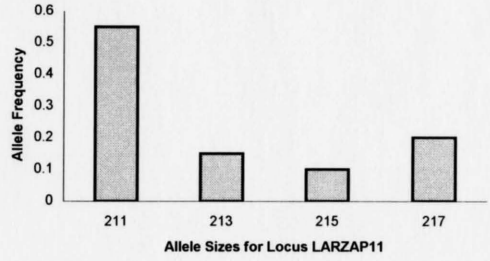


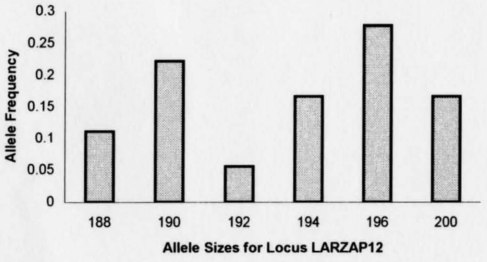

Figure 9. Allele frequency graphs for each of the five smooth-billed ani loci.

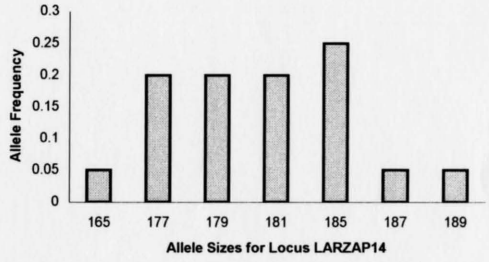


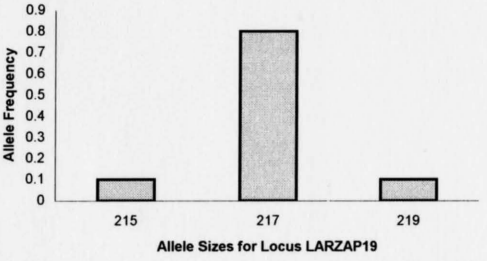



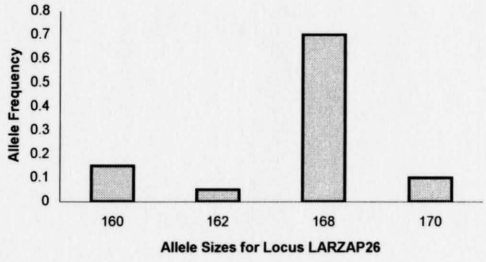




Figure 10. Allele frequency graphs for each of the eight herring gull loci.

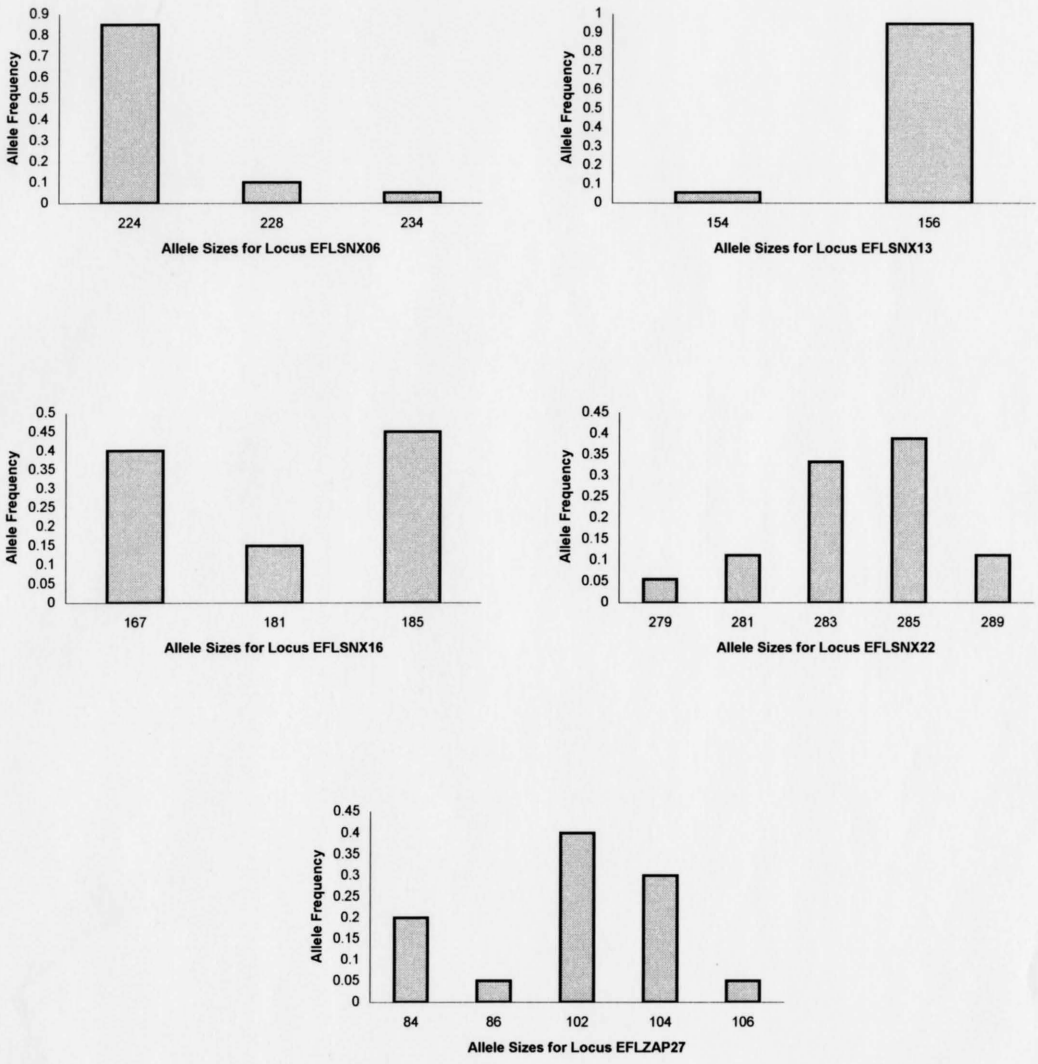



Figure 11. Allele frequency graphs for each of the five yellow-bellied elaenia loci.

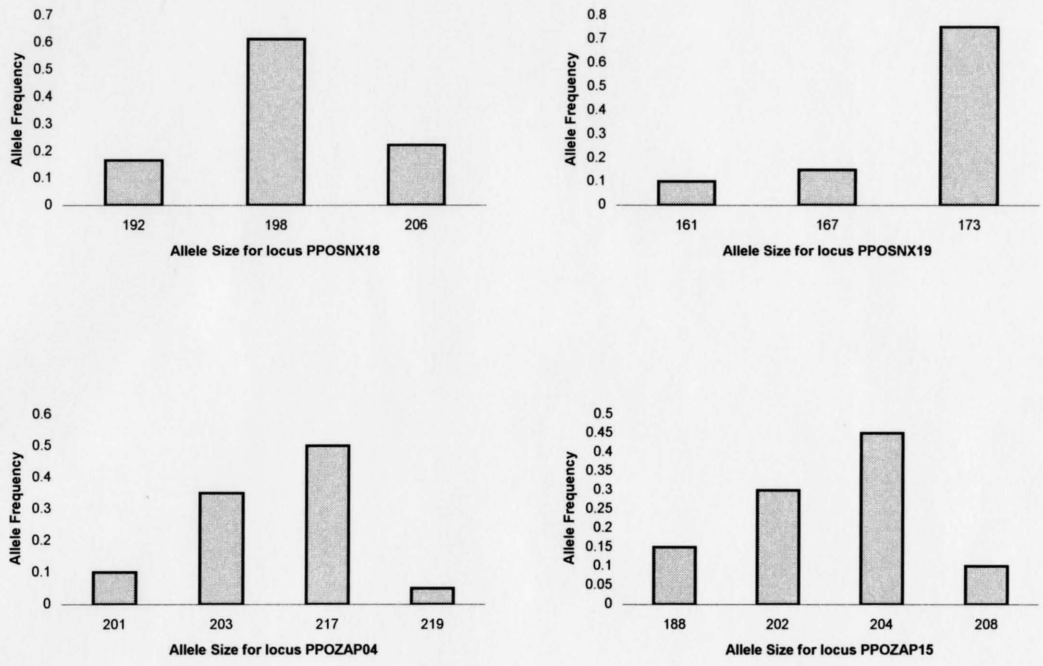



Figure 12. Allele frequency graphs for each of the four pukeko loci.

# Appendix 1

*Troubleshooting*

Throughout both methods, there exist areas in the protocol where problems may arise. We encountered many delays because of problems ranging from a defective TOPO-TA cloning kit to contamination. The following is a summary of possible problems, positive controls to test the problem, and where possible, solutions.

*Enrichment*

An important step to do with the enrichment protocol is to test the TOPO-TA cloning kit before using it for any cloning reactions. Unfortunately each of the following controls uses a shot of competent cells, which are expensive, but if there are problems with the kit then it is better to learn it early rather than proceed with cloning and waste cells as well as extra time. The TOPO-TA kit comes with its own positive controls. To test the transformation efficiency of the One Shot competent cells, pUC19 plasmid is included. To perform the control, transform with 10 pg of pUC19 per 50 μL of cells, then plate 10 μL plus 20 μL room temperature SOC medium. It is also recommended to plate a second more dilute mix by diluting 10 μL of the mix with 90 μL SOC medium. The transformation efficiency of the cells should be approximately 1 x $10^9$ cfu/μg DNA. Another test of the kit involves performing a control PCR using the control DNA and PCR primers included in the kit. This should test the TOPO vector. Perform the control PCR under the reaction conditions given in the manual, then set up two 6 μL cloning reactions: one with the vector only and one with the vector plus the PCR insert. These

are then transformed in separate vials and spread onto LB plates containing kanamycin and X-Gal (the control template is a plasmid that encodes ampicillin, so using ampicillin plates will result in far more white colonies in colour selection and an apparent increase in cloning efficiency, despite the fact that the colonies do not contain the desired construct). After overnight 37°C incubation, hundreds of colonies from the vector plus PCR insert reaction should be visible. At least 95% of these should be white, and at least 90% of these will contain the 750 bp control insert. The vector-only reaction should produce relatively few colonies, with most of them being blue.

Although both of the above positive controls are meant to test the competence of the cells, as well as the TOPO vector in the second control, a lack of colony growth could also be the result of a less likely, yet entirely possible (since it happened to us) scenario: that both the TOPO vector *and* the pUC19 plasmid are defective, while the cells are fine. We had concluded that both controls failed because of the common component in both reactions: the competent cells. However, when replacements were sent from Invitrogen (and replacements for the replacements) they also failed to give the expected results. To test the other hypothesis, that the two other components were both defective, pUC18 was obtained from a different supplier and transformed into the TOPO competent cells. Also, a PCR reaction using the contents of the cloning kit was done, and transformed into "homemade" competent cells using either the TOPO vector or a Bluescript vector. The pUC19 from the TOPO-TA cloning kit was also transformed into the homemade competent cells. The pUC18 produced colonies in the expected numbers with the competent cells, as did the reaction using Bluescript and the homemade competent cells, whereas the reactions using the TOPO vector and the homemade competent cells failed,

as did the reaction with the kit pUC19 and the homemade cells. From these results it was concluded that the One Shot competent cells in the kit were not the problem, but that both the control pUC19 plasmid and the TOPO vector from the kit were defective. The replacement from Invitrogen worked as expected based on information in the manual on both the new cells that came with the replacement kit, as well as on the cells from the original kit.

Once the components of the TOPO kit are known to be working properly, there are other steps that may go awry. It is important to do all of the test polymerase chain reactions listed in the methods, since sometimes even the simpler steps can prove troublesome. For example, on more than one occasion the ligation of the double stranded linkers did not work even though the exact same protocol was followed using the exact same reagents as times it did work. When the ligation was repeated immediately following the failed attempt, it worked properly. It is not known why this step failed at times.

*Lambda Zap*

Contamination was the main problem with Lambda Zap. When plating with NZY top agar, particularly but not exclusively when amplifying the Zap Express library, plates often were overgrown with contamination, even when using freshly made and autoclaved NZY top agar. The best way to ensure that the top agar is not contaminated is to prepare it several hours, or better yet one day before using it, and storing it at 48°C prior to use. This way contaminants will grow in the top agar *before* being plated, thus avoiding costly delays.

Another possible area where delays could occur with Lambda Zap involves extracting the DNA from the agarose gel following the size selection. Several companies offer kits for extracting DNA from agarose, but the ones tried failed to isolate enough of the DNA. The method that proved to be the most successful did not involve the use of any kits. A small amount of polyester stuffing (Stearns Canada) was placed halfway down a 1000 µL pipette tip so that it fills roughly 1 cm of the tip vertically. A small slice of the gel was then placed in the tip, and the tip placed inside a 1.5 mL eppindorf tube. It was then spun at 14,000 rpms for 7 minutes. A pipetteman was then used to squirt out any remaining liquid. Very low fluorometry readings are expected following the gel extraction. Readings below 20 ng/µL are not accurate, and in fact cannot with any certainty be taken as different from zero. A test to make sure that there is DNA with these low readings is to do the initial readings using 2 µL of DNA as is normally done, and then to do a reading using 10 µL of DNA. The reading should increase 5-fold. If it stays at the same level as the first reading, then no DNA is present.

## Appendix 2

*Glossary of Abbreviations*

CAN.....Smooth-billed ani (*Crotophaga ani*)

EFL.....Yellow bellied elaenia (*Elaenia flavogaster*)

$H_e$.....Expected heterozygosity

$H_o$.....Observed heterozygosity

LAR.....Herring gull (*Larus argentatus*)

MPC.....Magnetic particle collecting unit

OD.....Optical density

PIC.....Polymorphic information content

PPO.....Pukeko *(Porphyrio porphyrio*)

SWR.....Sequence With a Repeat

TE.....Tris/EDTA

TLE.....Tris/Low-EDTA

$T_m$.....Annealing temperature