Lincoln University Digital Thesis

Copyright Statement

Development and evaluation of a protocol to identify individuals of *Trichosurus vulpecula* with non-invasively recovered DNA

A thesis
submitted in partial fulfilment
of the requirements for the Degree of
Master of Applied Science
at
Lincoln University

by

Juan F. Dueñas-Serrano

Lincoln University
2012

a mis hermanos: Francisco, Nicolás y Pamela.
Los llevo en mi corazón...

Abstract of a thesis submitted in partial fulfilment of the
requirements for the Degree of M.Appl.Sc.

Development and evaluation of a protocol to identify individuals of *Trichosurus vulpecula* with non-invasively recovered DNA

by

Juan F. Dueñas-Serrano

The Australian brushtail possum (*Trichosurus vulpecula*) is a pervasive marsupial pest of New Zealand. Impacting on the native flora and fauna and the nation's livestock industry by its vectoring of bovine turberculosis, *T. vulpecula* is a priority for control and eventual eradication. Current pest control initiatives involve aerial deployment of chemical poisons, baiting and trapping. To establish the success of such control operations, estimates of possum population size pre- and post culling are required. Currently several monitoring methodologies — requiring the detection and trapping of individuals — are available to estimate indices of abundance (e.g. the residual trap-catch index). But these monitoring protocols are constrained by logistical and analytical considerations. The necessity to overcome the limitations of traditional monitoring schemes presents the opportunity to develop and evaluate the implementation of non-invasive genetic monitoring systems for possums. This thesis aimed to optimise an efficient amplification system for a panel of eight microsatellite loci that allow the identification of individual possums, characterise the occurrence of genotyping error across

a range of conditions, and evaluate the use of salivary DNA retrieved from interference devices as template for amplification.

Optimisation of amplification conditions for all loci in the panel was evaluated with DNA extracted from possum tissue collected at three localities in Canterbury region. Allele polymorphism was analysed by capillary electrophoresis and fluorescence based detection of fragments. After optimisation, locus Tv16 was discarded from the panel due to its linkage with locus Tv27 and amplification of unspecific fragments. Microsatellite diversity patterns of the seven remaining loci revealed moderate to high polymorphism and heterozygosity, no evidence of genetic structuring between localities across Canterbury (F_{st} = 0.03), and a sufficiently low overall probability of identity adjusted for siblings (PI_{sib}) (3 $\times 10^{-3}$) to ensure a robust identification of individual possums based on their multi-locus genotype. Further exclusion of locus Tv54 was recommended based on its high PI_{sib} (0.63–1.00) and incidence of genotyping error.

Amplification of template DNA extracted from tissue was not exempt from genotyping error (mean error rate per locus (e_l) = 4,8% and observed error rate per multi-locus genotype (e_{obs}) = 33.3%), these errors being associated in equal measure with stochastic causes (e.g. allele drop-out and false alleles) and systematic causes (e.g. scoring errors, sample swapping or contamination). No evidence of null alleles was detected.

Six loci were successfully assembled into a multiplex PCR assay. The implementation of mutliplex PCR had no significant effects on the incidence of genotyping error or the consistency of allele size estimation compared to standard PCR, and represented a substantial reduction in labour and resources needed to obtain a genotype (92% cost reduction relative to singleplex).

While 1:6 dilution of DNA extracted from tissue did not show significant effects on the amplification success and the mean genotyping error rate per locus, the use of template DNA retrieved from saliva decreased the performance of the microsatellite amplification system significantly. Only 18 of 24 samples were able to generate positive or partially positive genotypes, loci with amplicons $>$ 200 bp being the most affected, while the mean error rate per locus increased to 45%

Altogether, these results indicate that locus characteristics (i.e. amplicon size) and quality of template DNA are crucial factors affecting the sensitivity and reliability of the protocol developed. Potential ways to improve the remote collection of DNA from saliva are recommended.

Keywords: *Trichosurus vulpecula*, microsatellite markers, probability of identity, multiplex PCR, amplification success, genotyping error, saliva, pest control, population size, non-invasive genetic monitoring.

Acknowledgements

I would like to express my sentiments of gratitude to numerous colleagues, institutions and friends for supporting me throughout the duration of this work.

First, I would like to thank my supervisory team of Dr. Rob Cruickshank and Dr. James Ross for their guidance, disposition and input during the development and final accomplishment of this work. Rob patiently introduced me to the fascinating world of DNA, offering advice on molecular techniques and giving me much appreciated freedom to venture into the depths of the molecular ecology laboratory at Lincoln University. James provided invaluable assistance with statistical analyses and his ideas enriched the design of the study. Dr. Hayley Ridgeway assessed the proposal and provided suggestions that greatly improved the study design. Dr. Adrian Paterson read and commented on early drafts of this thesis and provided relevant suggestions to improve my writing skills. I am especially grateful to Dr. Marie Hale (Canterbury University, New Zealand), who generously shared the secrets of microsatellite data analysis and donated her time for an informative induction to the mechanics of capillary electrophoresis.

This study would not be possible without the generous collaboration of the Centre for Wildlife Management and Conservation staff, who helped me with data collection at the various stages of this study. My gratitude goes to

Brent Barrett, Matt Kavermann, Shona Sam, Lee Shapiro and Tim Sjöberg. I am also deeply grateful to Mr. Tim Kelly, who assisted with the collection of tissue samples at Lewis Pass. I wish to also thanks my colleagues at the molecular ecology laboratory; in particular Jagoba, Emily and Ben, for sharing their expertise and advice when needed. I am especially indebted to Norma Merric, who operates the sequencer at the Bio-Protection Research Center, for selflessly sharing her ample experience in fragment analysis; and to Candice Barclay, for serving as the bridge between the bio-technology industry and me.

Several colleagues provided generous assistance with data analysis and presentation of results. Sam Brown and Ruppert Collins inspired me to try LaTeX and offered advice on how to operate it. Ferran Palero (Institute of Science and Technology, Austria) provided feedback for the operation of MICROSATELIGHT. Dr. Paul Johnson (University of Glasgow, United Kingdom) offered timely suggestions for the quantification and analysis of genotyping error. Dr. Paul Moran (NOAA fisheries service, USA) granted permission for the reproduction of figure 3.1. I would like to express my debt of gratitude to all the programmers that share the philosophy of open software.

I gratefully acknowledge the financial assistance the Faculty of Agriculture and Life Sciences at Lincoln University and the Animal Health Board, through a research grant awarded to Dr. James Ross. I also wish to extend my debt of gratitude to the people at New Zealand Aid, in particular Angela Williams and Mandy Buller, for offering me the financial support to complete my degree in New Zealand, and for timely processing all my requests.

My love and thanks go to my family, for sparing me the time and support

to pursue my goals, and for offering me their unlimited love in return. To my friends, for sharing unforgettable adventures through land and sea, and for providing the necessary humour to bear with the odds of life. I hope I can repay in some way the time I have been away from you. Finally, to my dearest Annabel, my partner and love, for all the happiness you bring to my life, for providing the inspiration to continue against all odds and for spending painstaking hours correcting my 'spanglish'.

Contents

List of Figures

List of Tables

Abbreviations

NIGM non-invasive genetic monitoring

SSR short sequence repeats

NPCA National Possum Control Agency

RTCI residual trap catch-index

CWMC Centre for Wildlife Management and Conservation

MHC major histocompatibility complex

HW Hardy-Weinberg

ANOVA analysis of variance

CO-I barcoding region of the mitochondrial cytochrome oxydase sub-unit I gene

ANCOVA analysis of covariance

Chapter 1

Introduction

1.1 Non-invasive genetic monitoring

In the past decade molecular ecology has become established as an important sub-discipline of ecology (Rieseberg et al., 2010). Molecular ecology applies popular molecular biology techniques, such as the polymerase chain reaction (PCR) or DNA sequencing, to quantify the genetic diversity of organisms, populations and communities while answering questions about their ecology (Freeland et al., 2011; Beebee & Rowe, 2008). The link between genetic diversity and ecology provides new insights to topics that have been traditionally addressed with field techniques, such as population demography (Banks et al., 2003), host-parasite (Luikart et al., 2008a) or predator-prey interactions (Blejwas et al., 2006), to name a few. An increasing number of reviews across several sub-disciplines of ecology examine the potential of molecular approaches to further enrich their varied list of methodological strategies (Waits & Paetkau, 2005; Sarre & Georges, 2009; Schwartz et al., 2007). In many situations the creative use of molecular ecology approaches can infer answers to questions that would be impractical to address using

traditional techniques (e.g. the inference of historical population expansion, Curtis et al., 2009). The identification of individuals within a population — in order to determine the census population size — is one of the most widely studied applications of molecular ecology and constitutes the focus of this work.

Census population size (N_c) is defined as the number of individuals constituting a population at any given point in time (used synonymously with abundance in some texts, Freeland et al., 2011; Morrison, 2009; Mills, 2007). Scientists estimating N_c must be able to reliably detect and identify individuals of the targeted species. This can involve conducting surveys that target and count individuals, made difficult and expensive if the species under study is elusive or distributed over a large area. Data generated from such surveys can be analysed with rarefaction curves, capture-mark-recapture experiments or occupancy models to gain an estimate of N_c (Borchers et al., 2002; Sinclair, 2006; MacKenzie et al., 2006). A complication to the estimation of N_c arises because the concept of *population* has many definitions (Morrison, 2009), and surveys can only cover limited geographic regions.

To overcome some of the logistical and conceptual issues of estimating N_c, a set of alternative parameters known as indices of abundance have been defined (Williams et al., 2002). For example, density is an index that counts the number of individuals per unit of area. One assumption of indices of abundance is that all individuals in the population have constant detection probabilities (Mills, 2007), but if this is false, the relationship between indices of abundance and N_c is obscured. Despite this limitation, a moderate level of imprecision of population size estimates is tolerated more easily than the impracticalities of measuring N_c with the result that wildlife management monitoring schemes often include indices of abundance rather than N_c

(Warburton, 2000).

Further alternatives to traditional surveying methods and indices of abundance lie in the developing field of non-invasive genetic monitoring (NIGM). NIGM, as the name suggests, involves non-invasively collecting samples such as hair or faeces from which DNA can be retrieved. NIGM uses molecular markers, such as microsatellites or single nucleotide polymorphisms (SNP's), to generate a genotype, or fingerprint, of each individual within a population using the DNA collected. Unique genotypes can be later used to estimate $\hat{N}_c$ using similar analytical approaches to those used for traditional monitoring techniques. For instance, when using genetic data for mark-capture-recapture analysis, a genotype can be considered a capture while a repeated detection of the same genotype is considered a recapture (for a detailed review of N_c estimation using genetic data see Luikart et al., 2010).

NIGM possesses two important characteristics that make it an attractive monitoring scheme for wildlife management: universality and versatility. Universality because virtually all living organisms can be detected or monitored by retrieving their DNA. Versatility because over the last decade progress in this field has made it possible to generate reliable genotypes from minute amounts of template DNA - as little as 1 ng or less (Benschop et al., 2011; Petricevic et al., 2010). With the reduced amount of template DNA required, the number of sources that DNA can be obtained from is virtually endless. Hair follicles and intestinal cells present on fresh faeces of free ranging animals are currently the most commonly used samples for wildlife management applications (Broquet et al., 2007b). These achievements have been made possible by a suite of technical improvements primarily originating from the field of forensic sciences (van Oorschot et al., 2010).

The unique characteristics of NIGM have sparked an interest in evaluating this approach for an increasing range of non-model organisms and under a variety of contexts. While traditionally data from NIGM has been applied to species and individual identification, kinship analysis, dispersal and estimation of population size (DeYoung & Honeycutt, 2005; Luikart et al., 2010); the information generated by this method is also suitable for the study of population genetics and short-term evolution of species (Hamilton, 2009; Palstra & Ruzzante, 2008). For instance, information gathered during NIGM regimes can help in the interpretation of the mechanisms of establishment and spread of founding populations, or to determine if hybridisation counters the deleterious effects of inbreeding or loss of evolutionary potential (Allendorf & Luikart, 2007). Assessment of populations to gain such information is a critical first step in devising management plans of invasive species making NIGM increasingly important in effective monitoring efforts.

1.2 NIGM and Pest Management in New Zealand

The implementation of advanced monitoring methods for improved management of invasive species is of particular importance for New Zealand. One such species, constituting a significant threat to the country's indigenous biota and livestock industry, is the Australian brushtail possum *Trichosurus vulpecula*. Because of the direct and indirect negative interactions of *T. vulpecula* with native flora and fauna and its vectoring of bovine tuberculosis, extensive control operations are commonly undertaken in New Zealand rural areas (O'Reilly-Wapstra & Cowan, 2010; Sweetapple & Nugent, 2009). Control is primarily by means of aerial delivery of sodium fluoroacetate (1080) or ground-based deployment of toxic baits. Once control operations take place, it is necessary to determine their effectiveness

in order to prevent future outbreaks, guide subsequent control efforts, and evaluate the work of private contractors (Warburton, 2000). This requires a monitoring method that accurately estimates population size pre- and post-control. It is desirable that the monitoring method also allows additional information to be collected, such as patterns of aggregation and movement of surviving individuals (Brown et al., 2004; Cowan et al., 2002; Pech et al., 2010), or quantification of the re-invasion rate of control areas (Gleeson et al., 2010). In order to determine these factors, the National Possum Control Agency (NPCA) currently uses the residual trap catch-index (RTCI) (National Possum Control Agencies, 2008a; Warburton et al., 2004). The RTCI is a standardised index of abundance for possums and is relatively robust. However, RTCI suffers from several logistical and analytical limitations. First, the RTCI protocol requires the deployment of bulky equipment — leg-hold traps — and subsequent visits to these devices across extensive areas of often rugged terrain; second, the deployment of devices poses an unintended risk of interaction with endangered flightless birds such as the weka *Gallirallus australis* and kiwi *Apteryx* spp.; third, a recent study suggested that RTCI estimates have a non-linear relationship with population size due to seasonal and density dependent variation in detection probabilities (Forsyth et al., 2005). Despite these limitations, equally robust and more cost effective methods for estimating possum abundance are currently unavailable for widespread adoption.

A possible alternative for replacement of RTCI as the possum monitoring standard makes use of WaxTags and is under development by the NPCA (National Possum Control Agencies, 2008b). Waxtags are interference devices consisting of wax blocks attached to visual lures. When compared to leg-hold traps these devices show increased efficiency at detecting possums,

while the indices of abundance derived from them (i.e. tag station index and the bite-mark index) are consistent with the RTCI (Thomas et al., 2003; Ogilvie et al., 2006). Thomas et al. (2007) also concluded that the bite-mark index robustly reflects changes in possum population size comparable to RTCI. Yet the accuracy of the bite-mark index has been questioned. Indices of abundance derived from Waxtags appear to lose sensitivity in detecting population size changes when possum populations are either extremely large or small (Morgan et al., 2007; Warburton et al., 2004). A common argument against the use of abundance indices derived from Waxtags is that the use of lures to attract possums towards Waxtags triggers a behavioural response known as contagion. Contagion occurs when an individual actively seeks and bites several Waxtags on a given sampling occasion, which if unnoticed can bias the estimate of the bite-mark index (Thomas et al., 2003; Sargeant et al., 1998; Warburton, 2000).

In an attempt to further develop the Waxtags method, Vargas et al. (2009) were able to amplify the barcoding region of the mitochondrial cytochrome oxydase sub-unit I gene (CO-I) from DNA found on traces of saliva collected using Waxtags. Currently, saliva retrieved from Waxtags is an attractive substrate of remotely-collected possum DNA, given that preliminary studies have established major logistical constraints of obtaining DNA of this possum species from hair follicles and faecal samples (Gleeson et al., 2003; Morgan et al., 2007). Saliva presents a challenge and also a new and promising opportunity to test the ability of Waxtags to aid and better inform allocation of possum control operations. If Waxtags allow the retrieval of DNA of sufficient quantity and quality in order to conduct reliable genotyping assays, N_c can be directly determined, potentially offering the possibility of replacing existing indices of abundance.

1.3 Marker selection

In order to answer a variety of different ecological questions with genetic information, marker selection is critical. The progress in molecular biology has increased the availability of markers to choose from, each with limitations and therefore suited for a particular use.

Microsatellites are frequently used in studies involving population size estimation. Microsatellites or short sequence repeats (SSR) are non-coding regions of DNA conformed by a variable number of tandem repeats of a short sequence motif, generally of one to six nucleotides (Allendorf & Luikart, 2007). Variation in the number of repeats gives microsatellites the interesting property of having more than one allele, or polymorphism. It is acknowledged that microsatellite polymorphism is the result of a high mutation rate relative to the rest of the genome (10^{-2} to 10^{-6} nucleotides per locus, per generation, *sensu* Oliveira et al., 2006). However, the proposed causes (e.g. polymerase slippage or unequal crossing over) and accompanying theoretical models explaining the high mutation rate are numerous (Buschiazzo & Gemmell, 2006, and references therein) and ultimately have profound implications on the way genetic parameters of population differentiation, such as F_{st} (Balloux & Lugon-Moulin, 2002), are estimated.

When it comes to discriminating individuals, a task requiring the use of highly polymorphic markers to quantify genetic differences at such fine scales, microsatellite usage is widespread. While there is a variety of similar polymorphic markers such as allozymes or restriction fragment length polymorphism (RFLP's), microsatellites are generally preferred because of their co-dominant nature, reproducibility among studies and increased power for statistical analysis. However, the popularity of microsatellites has been rivalled by the inception of single nucleotide polymorphism markers (SNPs),

which unlike microsatellites have a well defined model of mutation and are therefore easier to score, do not require calibration when transfering between laboratories, and are amenable to different forms of typing (Garvin et al., 2010). Although SNPs have several desirable properties, in some situations microsatellites can prove a superior choice. Unlike SNPs, primers of microstellites can be transferred across taxa and errors associated with genotyping can be identified and tracked (Guichoux et al., 2011a). In the end, the choice of marker will be governed by reasons of practicality and economy of application.

Taylor & Cooper (1998) isolated a panel of eight microsatellite loci for *T. vulpecula*. This work provided future researchers with tools to study the genetic structure of both introduced and native populations of this species. Since then, this panel has been used in studies focused on gaining insights into the reproductive system of *T. vulpecula* in their native (Clinchy et al., 2004) and introduced ranges (Taylor et al., 2000); and understanding the genetic structure and relationships of introduced populations in New Zealand (Taylor et al., 2004). To date, one attempt has been made (using scats) to test these microsatellites in the context of NIGM (Morgan et al., 2007). In 2000, Lam et al. reported a microsatellite locus located within an intron of the major histocompatibility complex (MHC) gene region. This locus has since been used in combination with some of the microsatellites described by Taylor & Cooper (1998), in studies that examine the patterns of dispersal of *T. vulpecula* remnant populations in urban Australia (Stow et al., 2006) and to study the relationship of adult survival and genetic diversity of a closely related species *Trichosurus cunninghami* (Banks et al., 2008). Finally, very recently Sarre et al. (2010) described another set of microsatellite markers for possums. It is clear the potential of microsatellite markers to contribute

to a more effective control allocation is yet to be fully explored.

1.4 NIGM limitations

Genotyping of DNA collected from non-invasive sources poses a series of challenges. The first set of complications occur during allele amplification. Low quantity and quality of DNA collected from sources such as hair or saliva, may generate a unique kind of stochastic error - genotyping error. Genotyping error manifests in a variety of ways of which the most commonly reported are: 1) allele drop-out (i.e. failure to amplify one of the alleles of an heterozygote individual) and 2) false alleles (the erroneous scoring of "stuttering" peaks as true alleles) (Taberlet et al., 1996). Regardless of the type of error, the magnitude of genotyping error can seriously limit the ability of NIGM to yield accurate population size estimates (see reviews by Bonin et al., 2004; Broquet & Petit, 2004; DeWoody et al., 2006; Pompanon et al., 2005). For instance, Waits & Leberg (2000) determined that a per locus error rate as low as 0.05 can introduce a 200% bias on population size estimates. Furthermore, failure to quantify and monitor genotyping error occurrence can confound the conclusions of a broad range of studies. Hoffman & Amos (2005), in their study of paternity assignment of antartic fur seals *Arctocephalus gazella*, determined that error rates as low as 0.13–0.74% per sample resulted in the incorrect assignment of 4.9% of paternities.

The occurrence of stochastic genotyping error is often associated with the characteristics of each locus and the level of optimisation of the reaction conditions at which they were processed. For instance, loci that target relatively large amplicons (> 200 bp) may perform poorly with highly degraded DNA (Broquet et al., 2007b), while poorly optimised reaction conditions may produce genotypes plagued with artefacts such as stuttering peaks, split peaks

and primer dimers (reviewed in Guichoux et al., 2011a). Genotyping error is also often associated with low quantity and quality of template DNA, which is a common occurrence in non-invasively collected samples. But there is uncertainty, as to which of this two characteristics — quantity or quality — influences the occurrence of genotyping error more significantly. It is clear by this evidence, that once a panel of markers has been selected it is critical to optimise the markers individually while quantifying the occurrence of error in order to detect its causes.

A further complication of the genotyping process lies in errors that may arise during two distinct processes leading to the scoring of true or "underlying" alleles. These processes are known as sizing and binning. Both tasks, while conceptually simple, may induce the occurrence of human errors that will eventually have pervasive consequences on the final quality of genotyping data (Pompanon et al., 2005; Guichoux et al., 2011a). For instance, Hoffman & Amos (2005) attributed an important percentage of the total errors detected in their study to human errors (57.4%). The specific problems that allele sizing and binning may cause can be clarified and mitigated by understanding how these processes operate.

In the case of allele sizing, specialised software detects the presence of a peak representing a fragment and estimates its length by implementing a series of complex algorithms. Size-calling algorithms use a size standard — a collection of DNA fragments of known size all labelled with the same fluorocrome that is co-injected into the capillary of a sequencer along with the sample of interest — as a reference to estimate the length of a fragment (e.g. size-matching algorithm, Applied Biosystems 2005). Despite the sizing process being perceived as highly reproducible, in recent years a variety of studies have questioned this assumption. In a study of *Aspergillus fumigatus*,

Pasqualotto et al. (2007) reported differences of up to 5 nucleotides between the size of a fragment determined by sequencing and the size estimated with capillary electrophoresis. Other studies have reported differences in the size of alleles scored using different machines (LaHood et al., 2002; Ellis et al., 2011). Causes for this discrepancies may be varied, but it has been suggested that nucleotide composition of microsatellite flaking sequences, fluorochrome dye characteristics, differences in sequencer platforms, migration deviations of the internal size standard, nature of the gel matrix used and even the room temperature at the moment of electrophoresis can cause shifts in the mobility of alleles, therefore producing slightly different size estimates of the same alleles (Haberl & Tautz, 1999; Pasqualotto et al., 2007; LaHood et al., 2002; Sutton et al., 2011). Mobility shifts can be a serious problem that should not be overlooked, particularly when the goal is to standardise microsatellite typing protocols across laboratories (Pasqualotto et al., 2007; Moran et al., 2006; Ellis et al., 2011).

Binning raises a different set of issues. Because allele length estimates are expressed as decimal values, such data needs further processing to assign discrete allele categories. In well studied organisms, such as humans, this task is performed by comparing allele length estimates to a database of fragments of known size. The estimate length is then simply assigned the integer value of the closest known fragment. Unfortunately for most non-model organisms, like possums, such a database is missing. Accordingly, scientists assign allele length estimates to bins defined either automatically via several existing algorithms (e.g. least square minimisation, Idury & Cardon, 1997), or manually by visually defining bin boundaries based on allele sizes observed in a newly acquired dataset (Amos et al., 2007; Palero et al., 2011). Whichever binning method is considered appropriate, it needs

to be consistent in order to avoid further discrepancies in allele scoring. Correct binning needs to reflect the variety of alleles detected while accounting for the variability of size estimates caused by mobility shifts. This can be hard to achieve in some instances because the nominal repeat unit of a particular locus, for instance a dinucleotide repeat, will not correspond exactly to the observed repeat unit (e.g. in average 1.8 to 2.2 bp, Amos et al. 2007). This pattern has been termed "allele drift" or "size shift". When size shifts occur, allele size estimates located at the extremes of the size range could be out of alignment with the actual repeat length and consequently fall into the boundaries of a different bin (Amos et al., 2007; Ellis et al., 2011).

As some of these issues, in particular problems related to consistent sizing and binning, might only become apparent at advanced stages of the genotyping process their identification during a pilot study is essential to inform subsequent standardisation processes.

1.5 Aims and objectives: the importance of a pilot study

Given the outlined limitations, prior to testing the ability of the panel to identify individuals in field conditions, microsatellites constituting a panel need to be carefully screened for their ability to yield reliable and cost effective genotypes. Reliable genotypes are generated from loci with low susceptibility to genotyping error and carefully refined laboratory protocols. Cost effective genotypes can be easily interpreted without prohibitively increasing the costs and time needed to process them.

Three major issues were addressed via a detailed pilot study. First it was necessary to establish the panel's capacity to identify individuals. Parallel to

this objective, it was advisable to establish an efficient sample processing system that reduced the resources spent on obtaining the desired information. Finally, it was critical to implement a quality control system throughout the study to ensure that the acquired information was reliable.

Once the sample management and processing was refined, the ability to retrieve DNA of sufficient quantity and quality for genotyping from Waxtags was tested. However, since this last objective warranted a dedicated experiment, initial work was limited to refining the laboratory protocols using good quality DNA sources such as tissue, thus avoiding the complications of dealing with poor quality samples.

Chapter 2

Microsatellite panel optimisation

2.1 Introduction

PCR is a powerful technique and also incredibly specific. Since PCR is constituted by a number of cofactors, such as salts, primers, Taq polymerase and DNA template; and these cofactors operate optimally at a specific set of temperatures, some optimisation is generally required for use with specific protocols (Freeland et al., 2011). The amplification conditions vary from locus to locus making it necessary to carefully establish the cofactor concentrations and temperature profiles for each locus individually in order to obtain consistent and strong outputs. To complicate things even further, the relatively low portability of microsatellites across laboratories necessitates the implementation of tight standardisation procedures (see Ellis et al. 2011 for an example). Thus, even when optimal conditions for markers have already been published elsewhere, it may be necessary to re-optimise them in order to adjust for the particular circumstances of a new laboratory.

Once candidate markers have been identified and refined, it is important to evaluate their capacity to identify individuals. This capacity is quantified by the probability of identity (Taberlet & Luikart, 1999; Waits et al., 2001; Waits & Paetkau, 2005). The probability of identity (PI) corresponds to the probability of randomly capturing two different individuals from a population that have the same multi-locus genotype. The probability of identity is determined by the number of microsatellites included in the panel and the amount of variation in the population at each locus. To achieve a sufficiently low level of PI that enables identification of related individuals (i.e. the most genetically similar), numerous microsatellite loci are required. However, the greater the number of loci the greater the probability of genotyping error (Paetkau, 2004; Broquet et al., 2007b; Soulsbury et al., 2009). This trade-off is particularly important to consider when the target population is believed to have depleted genetic diversity. If this is the case, then a larger number of loci will be required to obtain an acceptable probability of identity, thus potentially increasing the occurrence of genotyping error to unacceptable levels.

Optimising a panel also involves quantifying and monitoring the occurrence of genotyping errors and artefactual processes that might be polluting the data. Researchers in the field of NIGM have developed several techniques to tackle these issues, essentially based on replicating PCR reactions for each sample (see Pompanon et al. 2005 for a thorough review). Since the implementation of these error checking methods can greatly affect the final cost of sample processing and are labour intensive, it is desirable to establish baseline information that allows the identification of error prone loci early during the development of a panel.

The first experiment aimed to work as a control and provide context for

the following objectives. The specific objectives were to: 1) quantify the diversity patterns of microsatellites in the panel, 2) establish the panel's overall statistical power to discriminate individuals (i.e. PI), and 3) screen the panel for the occurrence of genotyping error. The opportunity was also taken to gain some insight into the variability of microsatellite diversity patterns across several areas in Canterbury.

2.2 Materials and Methods

2.2.1 Study Area

Between June and July of 2011, ear tissue was collected from possums captured in three distinct geographical areas across Canterbury in the South Island of New Zealand: Banks Peninsula, Lewis Pass and Hororata (Figure 2.1). Prior to tissue collection, all individuals were sacrificed humanely, either on the spot or at the Centre for Wildlife Management and Conservation (CWMC) at Lincoln University, where some of these individuals were relocated for further research purposes. Approval for the removal and further experimentation on possums was granted by the Animal Ethics Committee.

Possums were captured in habitats ranging from pockets of native beech forest (*Nothofagus* spp.) within a tussock matrix to areas of exotic vegetation in the vicinity of farmland. Individuals from Lewis Pass ($n = 9$) were captured in transects set up within a margin of beech forest and tussock grassland, near Boyle village. Possums from Hororata were captured on private farmland and subsequently transported to the CWMC ($n = 19$). Possums from Banks Peninsula were captured at Kaituna Valley ($n = 9$), within pockets of native vegetation featuring stands of kahikatea and matai.

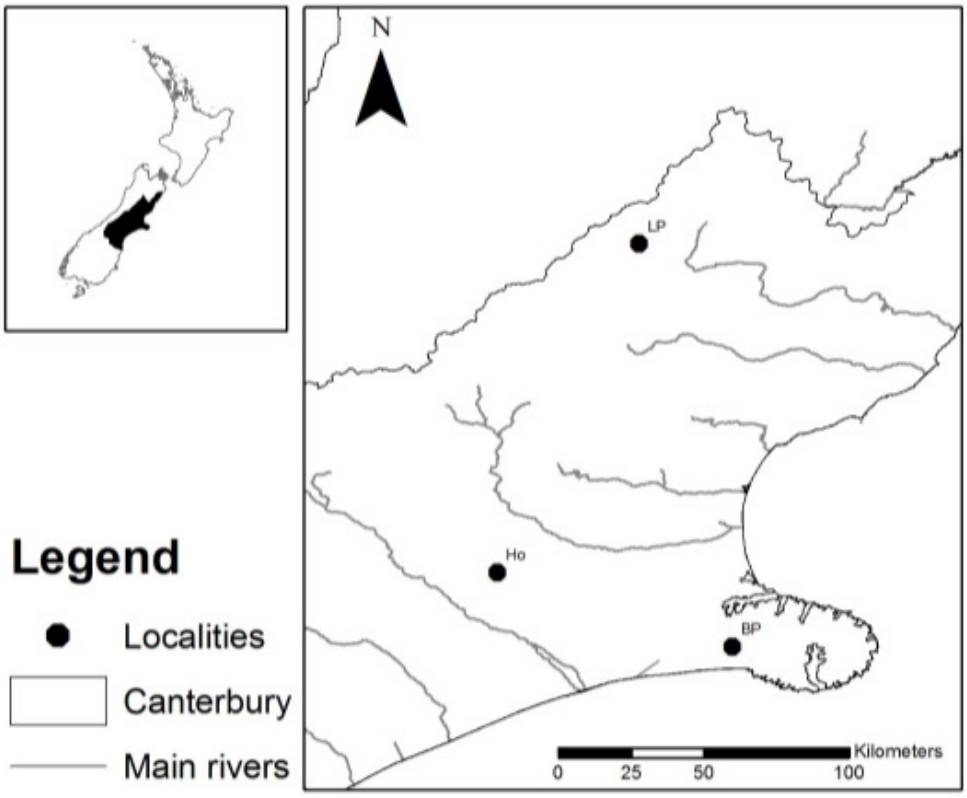

Figure 2.1: Geographic origin of samples collected for this study. BP = Banks Peninsula, Ho = Hororata and LP = Lewis Pass.

2.2.2 Data preparation

Tissue preservation

Tissue was collected by an incision of the distal portion of the ear. A total of two samples per individual were collected and stored separately. Two 10 mm samples of skin and cartilage were preserved in 1.5 mL tubes containing 1 mL of 99% ethanol (EtOH) and labelled with a unique identifier. Scalpel blades were rinsed in bleach after each incision to minimise the possibility of cross contamination. Specimens were transported to the molecular ecology laboratory at Lincoln University and stored at $-20\,^{\circ}\mathrm{C}$.

All tissue samples were collected and preserved within 24 hours of the animal being sacrificed. If samples could not be processed immediately, they were stored in a fridge for no more than 72 hours.

DNA extraction

DNA was extracted using the QIAGEN DNeasy tissue and blood extraction kit following the manufacturer's instructions. A tissue sample of approximately 2 mm was diced into smaller portions and transferred to a 1.5 mL tube with 180 μL of buffer ATL and 20 μL of proteinase K. The mixture was incubated for 24 hours in a dry oven at $56\,^{\circ}\mathrm{C}$. After incubation, a mixture of 200 μL of buffer AL and 200 μL of 100% EtOH was added to each tube, precipitating the DNA. This new mixture was transferred to a spin column and centrifuged at 6000 g for 1 min. Each spin column was fitted with a filter that binds DNA molecules as they are centrifuged. This filter was subsequently transferred to new collection tubes. Two washes, first with 500 μL of washing buffer AW1 followed by centrifugation at 6000 g for 1 min, and a second with 500 μL of washing buffer AW2 followed by centrifugation at

20000 g for 3 min were required to rinse non-target products off the filters.

Collection tubes and flow-throughs were discarded and the columns were finally transferred to a 1.5 mL tube. DNA was eluted by adding 200 μL of buffer AE and centrifuging tubes at 6000 g for 1 min. The elution step was repeated by adding 200 μL of buffer AE. A label containing a unique identification number was printed and fitted to each tube. Finally, the extracts were preserved at $-20\,^{\circ}$C.

Quantification and screening of DNA extracts from tissue samples

To quantify DNA yield, extracts were thawed and 1μL measured using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Inc) to the nearest ng/μL. After initial quantification, samples were screened by amplifying the DNA barcoding region of the mitochondrial cytochrome oxidase sub-unit I (CO-I) gene. All PCRs were performed with a MultiGene TC9600-G thermal cycler (Labnet International, Inc) with the following profile: denaturation at 94 °C for 2 min, then 33 cycles of denaturation at 94 °C for 30 s, annealing at 45 °C for 30 s and extension at 72 °C for 1.30 min, followed by a final extension time at 72 °C for 5 min. The 10 μL reactions contained 1 μL of template, 1 μL of 10x buffer (iNtRON) (final $MgCl_2$ concentration of 2 mM), 0.8 μL of dNTP mixture (final concentration of 0.2 mM of each dNTP, iNtRON), 0.48 μL of primers 0.48 μL of primers MLepF1 (Hajibabaei et al., 2006) and HCO2198 (Folmer et al. 1994, from a stock solution of 10 μM) and 0.08 μL of i-StarTaq (5 U/μL, iNtRON).

PCR products were visualised on 1.5% agarose gels. After a mixture of 5 μL of PCR product and 1 μL of gel loading buffer was loaded into each well, gels were run at 80 volts for 40 min. Gels were visualised under UV light to detect the presence of bands that indicated the successful amplification of a

DNA fragment.

2.2.3 Microsatellite amplification

Fluorochrome dyed primers

A panel of eight microsatellites were selected from the literature. The 5' end of the forward primers was labelled with a specific fluorescent dye (6-FAM, VIC, NED or PET, Applied Biosystems). Dye selection was based on previously reported allele size ranges and on recommendations from colleagues (Vargas, pers. comm) (Table 2.1).

10,000 pmoles of both the forward and reverse primers were supplied dry by the manufacturer. Primers were brought to 200 μM stock solutions by adding 50 μL of 10 nmol/L Tris, 1 nmol/L ethylenediamide tetra acetic acid (TE Buffer) to the fluorescent dyed primers, and 50 μL of deionised Water (dH_2O) to the unlabelled primers. Stock solutions were stored at $-20\,^{\circ}$C. For initial testing purposes, primers of each locus were brought to a working concentration of 2 μM and stored in 100 μL aliquots. Subsequently, optimal working concentrations specific for each primer pair were adjusted empirically.

Amplification optimisation

Initial amplification was performed following protocols in Taylor & Cooper (1998) and Vargas et al. (2009), followed by empirical adjustment of $MgCl_2$ concentration, thermal cycler profile and primer concentration.

Three different DNA polymerase enzymes were tested: conventional *i*-Taq, *i*-StarTaq and *i*-Star MAX II (iNtRON Biotechnology). Each enzyme has slightly different capabilities. While the *i*-Taq is active at room temperature, *i*-StarTaq is designed for hot start PCR, therefore it is in-

Table 2.1: Characteristics of the 8 loci selected to genotype possums

Locus	Sequence 5'-3'	Allele size range	Fluorescent dye	Source
TvM1	GACCACAACCTGGGTCTAACCAACG	195–223	6FAM (Blue)	Lam et al. 2000
	CATGACACCTGGGCACTCAGGACT			
Tv27	AGTGGAACCACATGTCAGGGC	125–168	6FAM (Blue)	Taylor & Cooper 1998
	GGACTGAAATGACTGCACAAC			
Tv53	GGGAGTAGTTGTCTGAGTTCCC	233–263	NED (Yellow)	Taylor & Cooper 1998
	CCCTGGAGTTTGACAACCTG			
Tv54	GGGAGGCATAAAGTGCCAGA	87–119	NED (Yellow)	Taylor & Cooper 1998
	TGACCGACACTGACGACCCC			
Tv5.64	TTTATCCCTACTAGAGGTAGGT	122–168	VIC (Green)	Sarre et al. 2010
	ATTAGCGCTTACCAGAGTGC			
Tv16	GAGGCTACCATTAGACGCAA	83–115	VIC (Green)	Taylor & Cooper 1998
	AGATACTATCTGCATCCAGAG			
Tv19	CCTCCTCCCCATCCTTCCTG	214–254	PET (Red)	Taylor & Cooper 1998
	GTTCAATTGCAGGGCTATGG			
Tv58	GCACCCAAGGACCCCCAAGA	102–158	PET (Red)	Taylor & Cooper 1998
	CCATATCACAGTGCTTGGCG			

active until the temperature of the reaction raises above a certain cut-off value. The *i*-Star MAX II features a combination of *i*-StarTaq polymerase plus a proofreading polymerase that gives it a higher fidelity during amplification. After initial experiments, *i*-StarTaq outperformed *i*-Taq and performed similarly to *i*-Star MAX II. For this reason, and because it is the least expensive enzyme, the remaining PCRs were carried using *i*-StarTaq only.

Microsatellites were amplified in 10 μL standard reactions (hereafter referred as standard PCR) containing 2.5 μL of template, 1 μL of forward and reverse primer solution (see Table 2.2 for final concentrations for each locus), 1–1.6 μL of 10x buffer (supplemented with 0–1.2 μL of Mg^{2+} solution (final $MgCl_2$ concentration of 2–3 mM), 0.8 μL of dNTP mixture (final concentration of 0.25 mM of each dNTP), and 0.08 μL of *i*-StarTaq ($5U/\mu$L). Cycling was performed in a MultiGene TC9600-G thermal cycler with the following profile: denaturation at 94 °C for 3 min, then 25 cycles of denaturation at 94 °C for 30 s, annealing at optimal temperature for 45 s (see Table 2.2) and extension at 72 °C for 45–90 s (depending on the amplicon size, 45 s if $<$ 200 bp, 90 s if larger), followed by a final extension time at 72 °C for 5 min. Locus Tv19, Tv27 and Tv5.64 required of 5 touch down (TD) cycles of denaturation at 94 °C for 30 s, TD annealing temperatures for 45 s (see Table 2.2) and extension at 72 °C for 45 s.

In order to minimise contamination, a series of guidelines were implemented for each experiment. Master mix solutions were prepared under a UV hood to prevent contamination of reagents, negative controls were included in all PCRs to detect possible cross-contamination, and finally PCRs were not performed on the same day as the extractions took place.

Table 2.2: Optimised PCR conditions for seven[a] microsatellite loci

Locus	Annealing temperature (in °C)	Primer concentration (in μM)
Tv19	55–60	0.5
Tv16	–	–
Tv27	60–65	0.5
Tv53	60	0.2
Tv54	60	0.75
Tv58	60	0.5
Tv5.64	55–60	0.5
TvM1	55	0.5

[a] Locus Tv16 was excluded from the panel due to difficulties with PCR optimisation and its linkage to locus Tv27

2.2.4 Fragment Analysis

The products of microsatellite singleplex reactions were visualised on 1.5% agarose gels to screen for positive bands, in the same fashion as described in section 2.2.2. Those products that exhibited strong positive bands were selected for capillary electrophoresis.

Capillary electrophoresis was carried out using an ABI PRISM 3130xl sequencer (Applied Biosystems). Products were submitted on a 96-well micro-plate. Each well contained 0.5 μL of PCR product, 12 μL of Hi-Di formamide (Applied Biosystems) and 0.3 μL of GeneScan–500 LIZ size standard (Applied Biosystems). In some cases, when PCR products showed weak bands on agarose gel electrophoresis, 1 μL of product was added to the mixture. The micro-plate was sealed and the mixture in the wells was denatured for 1 min at 95 °C. In order to prevent re-annealing of DNA strands, the micro-plate was then kept on ice prior to loading samples into

the sequencer.

Allele scoring

Fragments were visualised and their sizes estimated using Peak Scanner v1.0 (Applied Biosystems). A combined results table was then exported in .txt format into the program MICROSATELIGHT (Palero et al. 2011). This program is an open-source platform specifically developed to score microsatellite data and translate its outputs into files that can be interpreted by a variety of commonly used population genetics software.

MICROSATELIGHT offers automated and manual possibilities for defining allele bin boundaries (AlleloBin, Prasanth et al. 2006; Binator, Palero et al. 2011). Binning was performed using the AlleloBin sub-routine. AlleloBin implements the least square minimisation algorithm defined by Idury & Cardon (1997).

In order to detect all of the alleles produced during PCR reactions, allele scoring was completed by visually inspecting all electropherograms and comparing these observations with results from automated allele scoring.

2.2.5 Genotyping error

In order to detect and quantify the magnitude and frequency of genotyping error, a minimum of two samples per locality were randomly selected for replication ($n = 8$). The selected samples represent $\sim 20\%$ of the total dataset as advised by (Pompanon et al., 2005). Microsatellite amplification was undertaken as described in section 2.2.3.

2.2.6 Data Analysis

Microsatellite polymorphism

The number of alleles per locus, number of private alleles, observed (H_o), and the expected heterozygosity (H_e) were estimated for each locus across localities. In order to detect deviations from Hardy-Weinberg expectations (HW), a HW equilibrium test was performed. Because there was no *a priori* expectation of heterozygote excess or deficit, an HW exact probability test was chosen (Guo & Thompson, 1992; Weir, 1996). Wright's F statistics, such as the inbreeding (F_{is}) and subpopulation structure coefficients (F_{st}), were estimated using the analysis of variance (ANOVA) framework defined in Weir & Cockerham (1984) and Weir (1996). F_{st} was favoured over other measures of population structure, such as R_{st}, because it performs relatively well for moderately structured populations, when a small number of loci is analysed, and is independent of microsatellite mutation patterns (Holsinger & Weir, 2009; Balloux & Lugon-Moulin, 2002). An exact test of population differentiation (100000 Markov chain steps Goudet et al., 1996; Raymond & Rousset, 1995) was performed to determine whether F_{st} among localities significantly differs from zero.

Genetic diversity measures were estimated with the program GENALEX (Peakall & Smouse 2006), which is built into Microsoft Exel. The HW exact probability tests and F statistics were estimated using GENEPOP v.4.1.1 (Rousset, 2008). It is worth noting that Rousset (2008) implemented slight modifications to the estimation of F statistics in the latest version of GENEPOP. Details of these modifications can be found in Rousset (2008). Both the F_{is} and F_{st} goodness of fit tests were performed in the program ARLEQUIN (Excoffier & Lischer 2010).

Probability of identity

In order to quantify the panel's ability to discriminate individuals, allele frequencies of each locus were used to estimate the probability of identity (PI). Because the dataset in this study is likely to contain related individuals (a violation of one of the fundamental assumptions for an unbiased calculation of PI) the panel's discriminatory power was estimated using the more conservative probability of identity adjusted for siblings (PI_{sib}). PI_{sib} accounts for potential underestimation of PI and was estimated by the formulae given in Taberlet & Luikart (1999) and Waits et al. (2001).

The per locus PI_{sib} is given by the equation:

$$PI_{sib} = 0.25 + (0.5 \sum p_i^2) + [0.5(\sum p_i^2)^2] - 0.25(\sum p_i^4) \qquad (2.1)$$

where the term p_i is the allele frequency of the ith allele at a given locus.

In accordance with the product rule, the multi-locus PI_{sib} is obtained by multiplying the PI_{sib} values of each individual locus. The effect that removing a particular locus would have on the discriminatory power of the panel was evaluated by re-estimating the PI_{sib} iteratively, excluding the information from one locus on each occasion. These calculations were performed in the program GENALEX.

Genotyping error

Genotyping errors were detected by comparing the multi-locus genotype of blindly replicated samples with their corresponding multi-locus genotype in the original dataset. To avoid further ambiguities only allele size estimates prior to binning were compared.

Errors were further categorised according to their possible causes in two

main classes: 1) stochastic errors, which include allelic drop-out and false alleles; and 2) systematic errors, which cover scoring errors, contamination and sample confusion. False alleles were scored only when differences between replicated alleles were equal to or greater than two base pairs. Data on the prevalence of stochastic errors was used to estimate per locus allele drop-out rate (p_l), per locus false allele rate (f_l), weighted average of allele drop-out rate (p_w) and weighted average of false allele rate (f_w), according to the equations defined in Broquet & Petit (2004).

In order to have a more general metric for error quantification, the mean error rate per locus (e_l) and the observed error rate per multi-locus genotype (e_{obs}) were estimated with the equations given in Pompanon et al. (2005). These equations require counting the frequency of mismatches. A mismatch was scored when a minimum of one allele on each locus differed from the corresponding alleles in the reference genotype. Only mismatches correspondent to stochastic and scoring errors were taken into account to estimate e_l and e_{obs}, since mismatches associated with contamination or sample swapping are not the direct result of locus or reaction characteristics. It was considered that systematic errors can be mitigated by an increased level of training and by enforcing more stringent laboratory and data handling protocols.

To assess if any deficit in heterozygotes is compatible with the presence of null alleles across loci, the tests implemented by the program MICROCHECKER (van Oosterhout et al. 2004) were applied to data from the localities that showed significant departures from HW equilibrium.

2.3 Results

2.3.1 Sample characteristics

A total of 37 individuals were captured and sacrificed. Of these, 20 were male, 14 were female and 3 were pouch-young of undetermined sex. Except for the 3 pouch young individuals, it was impossible to determine *a priori* if all of the adult individuals collected in this study were related.

Fur colouration of 16 individuals was categorised as black while 17 were categorised as grey. There were four individuals who were skinned before fur colouration pattern could be determined. A great deal of variation in fur tones was noted, thus the categorisation is considered subjective and is provided as supplementary information only.

2.3.2 DNA quantification and screening

Tissue samples yielded variable but relatively high DNA concentrations. A total of 71 extractions were necessary to ensure the freshness of template DNA used for optimisation reactions. The first batch of extractions, corresponding to 37 individuals, was performed between June-July of 2011 (median = 59.85 ng/μL, range: 29.11–297.03 ng/μL; Figure 2.2). A second batch of just 34 individuals was extracted during October 2011 (median = 57.22 ng/μL, range: 22.61–168.00 ng/μL). All extracts exhibited positive bands for CO-I region amplification.

2.3.3 Microsatellite amplification

PCR conditions for seven out of eight microsatellite primer pairs were optimised allowing for the generation of multi-locus genetic profiles of the 37 individuals collected (see Table 2.2). Despite considerable effort, PCR con-

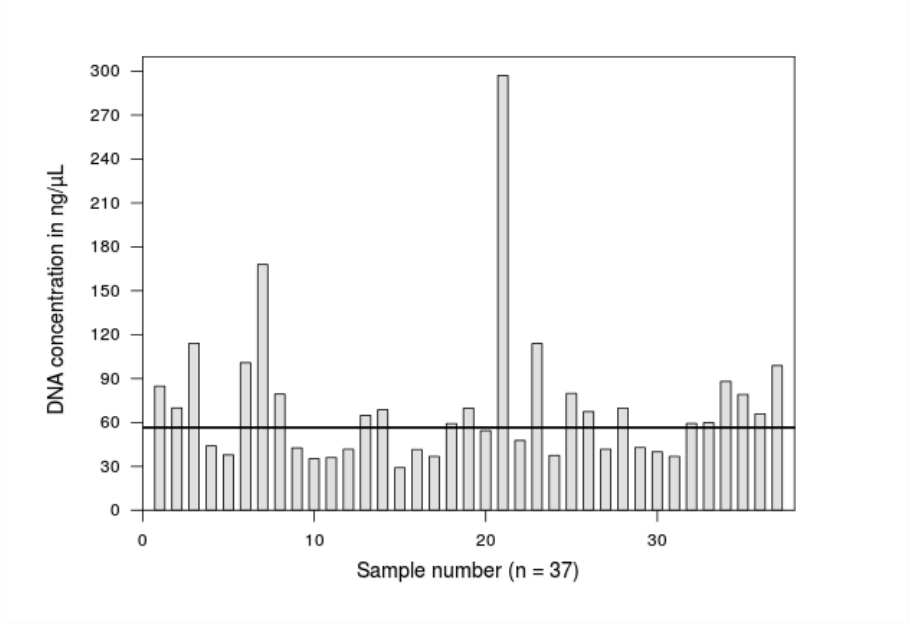

Figure 2.2: DNA concentration measured in ng/μL. Bars represent tissue samples from which DNA was extracted between June-July, 2011. The horizontal line represents the median DNA concentration (59.85 ng/μL)

ditions for marker Tv16 could not be optimised, thus forcing its exclusion from the panel. The automated algorithm implemented to assign raw allele size lengths into discrete size categories performed well for all markers examined in this study, as revealed by the low magnitudes of normalised variance estimated during the binning process (Idury & Cardon, 1997, data not shown).

All refined loci showed moderate levels of allelic diversity except for locus Tv54, which showed a reduced number of alleles across localities (Figure 2.3). Most alleles detected in the study were present across all of the sampled sites, although there were some instances in which private alleles were detected. This is the case for the Lewis Pass sample site, which despite having a reduced sample size compared to Hororata, showed a greater prevalence of private alleles. Lewis Pass was followed closely by Hororata and finally by

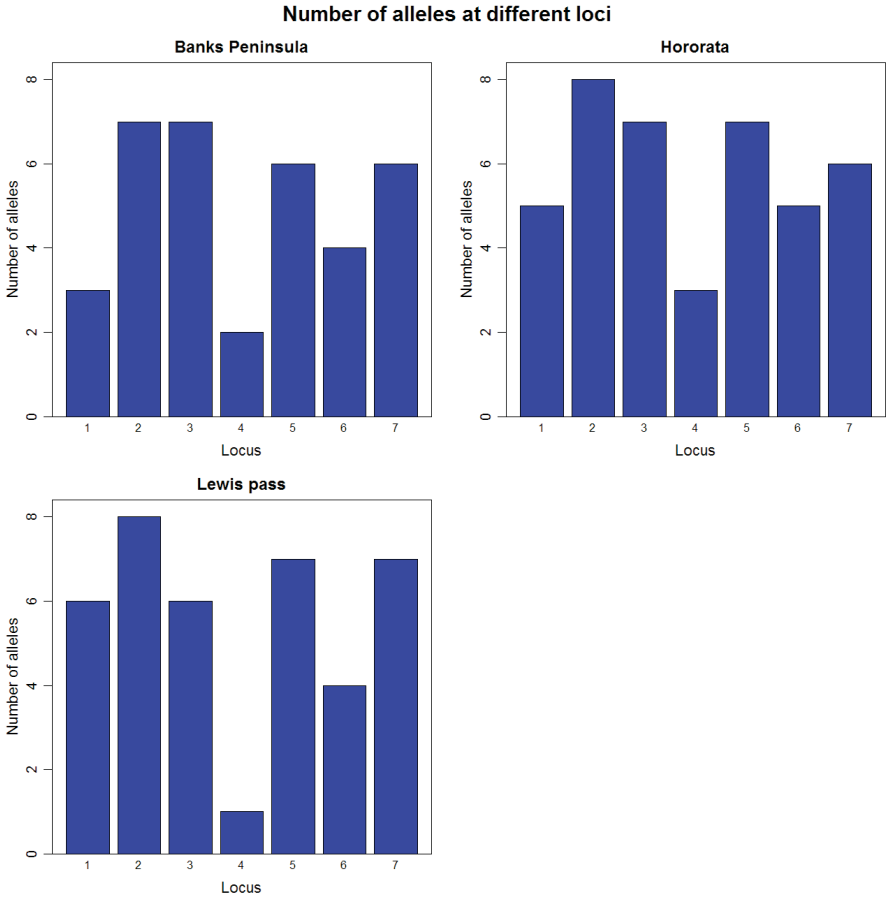

Figure 2.3: Number of alleles observed across loci and location in Canterbury. Numbers on the x-axis represent the different loci in ascending order as follows: 1= Tv19, 2 = Tv27, 3 = Tv53, 4 = Tv54, 5 = Tv58, 6 = Tv5.64 and 7 = TvM1

Banks Peninsula (Table 2.3).

Despite a low mean sample size (mean ± SE: 12.29 ± 1.04), it was possible to detect moderate levels of allelic diversity (5.57 ± 0.5) and observed heterozygosity (0.64 ± 0.06) across localities. As expected by the number of alleles detected, locus Tv54 was the only one in which observed heterozygosity levels were low across all localities. Locus Tv58 exhibited low heterozygosity levels both in Hororata and Banks Peninsula, while locus Tv5.64 showed relatively low heterozygosity only in Banks Peninsula

(Table 2.3).

From the three localities sampled in this study, Hororata ($\chi^2 = 33.98$; d.f. = 14; $P = 0.0021$) and Lewis Pass ($\chi^2 = 23.11$; d.f. = 12; $P = 0.026$) showed significant departures from HW equilibrium. When results of the exact probability tests were partitioned in locus/locality combinations, six out of twenty one combinations showed significant departures from HW expectations (Table 2.4). However, after the sequential Bonferroni correction was applied (Holm, 1979), only locus Tv58 in Hororata exhibited significant departures from HW expectations ($P = 0.019$).

Loci evaluated in this study showed a low proportion of genetic diversity due to allele frequency differences among localities ($F_{st} = 0.03$). However, a detailed partition of F_{st} parameters revealed moderate levels of genetic structure for locus Tv19 (Table 2.3). Evidence of lack of genetic structure was further supported by a pairwise comparison matrix of F_{st} (Table 2.5). From Table 2.5 it is possible to visualise F_{st} magnitudes among localities indicating only moderate to low structuring. Finally, the exact test of population differentiation showed that the overall value of F_{st} found among the three localities does not significantly differ from zero ($P = 0.179$).

2.3.4 Probability of identity

The PI_{sib} estimated on the basis of allele frequencies for all localities sampled in Canterbury was 3.0×10^{-3}. This parameter indicates that the microsatellite panel optimised in the present study provided sufficient information to identify individual possums based on their multi-locus genetic profile. When this parameter was re-estimated by separating allele frequencies per sampled site, some variation in the magnitude of PI_{sib} was observed (Figure 2.4). Despite this variation, the capacity to discriminate individuals remained robust

Table 2.3: Number of alleles (A), number of private alleles (PA), observed (H_o) and expected (H_e) heterozygosity and inbreeding coefficient (F_{is}) for every locus by location sampled in Canterbury. Population structure coefficient (F_{st}) estimated for each locus in Canterbury

	Banks Peninsula						Hororata						Lewis Pass						
Locus	n	A	PA	H_o	H_e	F_{is}	n	A	PA	H_o	H_e	F_{is}	n	A	PA	H_o	H_e	F_{is}	F_{st}
Tv19	9	4	0	0.78	0.64	-0.80	19	5	0	0.79	0.67	-0.04	9	6	1	0.78	0.80	0.34	0.1111
Tv27	9	6	0	0.67	0.77	-0.21	19	10	2	0.79	0.80	0.12	9	9	1	0.89	0.85	0.22	0.03764
Tv53	9	7	0	0.89	0.80	0.17	19	8	0	0.79	0.75	-0.14	9	6	0	0.56	0.80	0.23	0.02993
Tv54	9	2	0	0.11	0.10	-0.14	19	3	1	0.58	0.43	-0.26	9	1	0	0.00	0.00	0.00	0.07092
Tv58	9	4	0	0.22	0.38	0.42	19	8	2	0.42	0.66	0.45	9	8	3	0.67	0.66	0.02	-0.03764
Tv564	9	4	1	0.44	0.60	-0.33	19	4	0	0.79	0.67	0.16	9	4	0	0.78	0.70	-0.20	0.00062
TvM1	9	5	0	0.89	0.77	-0.12	18	7	0	0.83	0.78	0.06	9	6	1	0.78	0.80	-0.03	0.00142

Table 2.4: Summary of statistical parameters estimated per locus at each of the localities sampled in Canterbury. The HW *exact* probability test P-values (HW P)[a] and their standard error (SE) as well as the PI_{sib} for every locus are presented in this table

	Banks Peninsula[b]			Hororata[c]			Lewis Pass[d]		
Locus	HW P	SE	PI_{sib}	HW P	SE	PI_{sib}	HW P	SE	PI_{sib}
Tv19	0.013	0.000	0.476	0.007	0.001	0.459	0.01	0.00	0.369
Tv27	0.310	0.016	0.386	0.279	0.016	0.366	0.16	0.02	0.338
Tv53	0.537	0.022	0.369	0.390	0.021	0.397	0.01	0.00	0.366
Tv54	1.000	0.000	0.899	0.720	0.000	0.638	NA	NA	1.000
Tv58	0.034	0.006	0.665	0.000	0.000	0.462	0.65	0.03	0.454
Tv5.64	0.779	0.000	0.506	0.469	0.012	0.455	1.00	0.00	0.436
TvM1	1.000	0.000	0.387	0.832	0.008	0.379	0.79	0.01	0.370

[a] P-values before sequential Bonferroni correction are presented here
[b] Banks Peninsula $PI_{sib} = 7.9 \times 10^{-3}$
[c] Hororata $PI_{sib} = 3.4 \times 10^{-3}$
[d] Lewis Pass $PI_{sib} = 3.3 \times 10^{-3}$

Table 2.5: Pairwise comparison matrix of genetic structure (F_{st}) between localities sampled in Canterbury

	F_{st}		
Locality	Lewis Pass	Banks Peninsula	Hororata
Lewis Pass	0.000		
Banks Peninsula	0.051	0.000	
Hororata	0.026	0.024	0.000

across localities. Most notably, the magnitude of PI_{sib} observed in Lewis Pass was very similar to that observed in Hororata despite Hororata having a larger sample size.

Whilst none of the locus exclusions drove the magnitude of PI_{sib} to unacceptable levels (i.e. $\geqslant 0.01$, Waits et al. 2001; Mills et al. 2000), this analysis allowed detection of two important patterns (Figure 2.5). First, it was clear that locus Tv54 contributed very little information to drive PI_{sib} down, which was an expected consequence of the low allelic diversity detected for this marker in the present study. Second, contributions from loci Tv19, Tv27, TvM1 and Tv53 seemed to constitute the basis of the panel's ability to discriminate individuals, supplemented by the information provided by the remaining loci.

2.3.5 Genotyping error

The mean error rate per locus was 4.76% yielding an overall high observed error rate per multi-locus genotype (33.33%). Of these errors, two instances were identified as allelic drop-out, which yielded a weighted average of allele drop-out rate of 0.05. Two loci were affected by allelic drop-out: Tv58 ($p_l = 0.2$) and Tv27 ($p_l = 0.17$). Similarly, two instances of false alleles

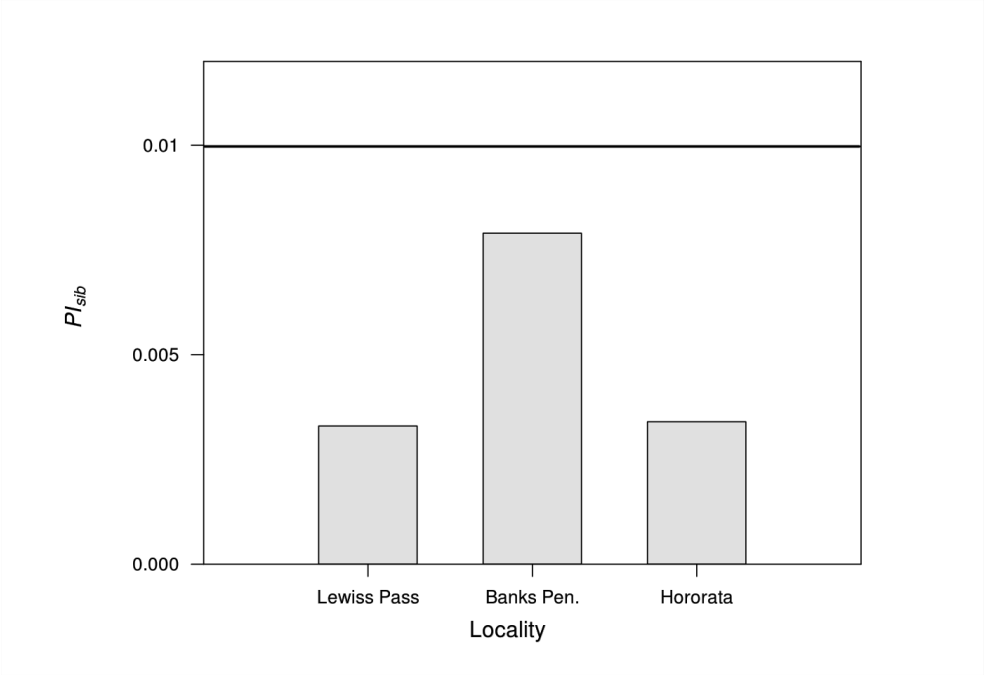

Figure 2.4: Probability of identity adjusted for siblings variation across localities in Canterbury. The horizontal line marks the minimum value of PI_{sib} required to avoid bias in genetic estimators of population size as suggested by Waits et al. (2001)

were detected at two different loci: Tv53 ($f_l = 0.06$ and Tv54 ($f_l = 0.06$), resulting in a weighted average of false allele rate (f_w) of 0.02.

Two instances of human error could be clearly detected. Most likely, these errors were due to sample confusion rather than contamination, as negative controls of each particular experiment came out clean. No scoring errors were detected.

No loci from Hororata or Lewis Pass showed showed evidence of the presence of null alleles according to MICROCHECKER tests.

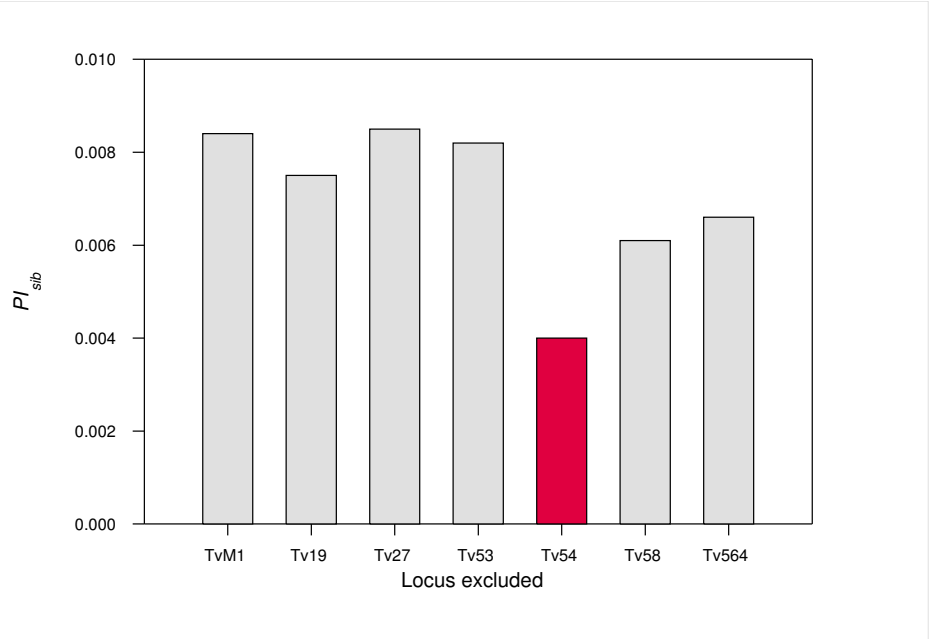

Figure 2.5: Contribution of each locus to the magnitude of PI_{sib}. Bars represent the magnitude of PI_{sib} estimated when the information of the locus nominating the bar is missing. The bar highlighted in red shows the estimated PI_{sib} when locus Tv54 is excluded from the calculations.

2.4 Discussion

Microsatellite diversity patterns observed in this experiment indicate a moderate level of genetic diversity in possums collected at Canterbury. Taylor & Cooper (1998) pioneered the evaluation of *T. vulpecula* microsatellite diversity when they described primers for most of the microsatellites included in the present study. Along with the description of primers, high levels of allelic diversity and expected heterozygosity for possum populations in New Zealand were reported (Taylor & Cooper, 1998). However, no mention was made of the specific geographic origin of the samples in question, thus preventing a direct comparison with the present study. In a later study focusing on the reproductive system of possums, Taylor et al. (2000) also reported high levels of heterozygosity and allelic diversity. In that occasion,

samples only came from the North Island. Building on those first efforts, Taylor et al. (2004) sampled a greater portion of the geographic range of *T. vulpecula*'s distribution in New Zealand, finding differences in the allelic diversity and heterozygosity of possums in the North Island compared to their counterparts in the South Island and off shore islands. Taylor et al. (2004) found that possums in the South Island show depleted diversity as opposed to those in the North Island, suggesting this pattern is due to a more recent introduction and reduced genetic diversity of the founding populations in the South Island (i.e. animals introduced were only from Tasmanian origin). The genetic diversity of localities sampled in this study seem consistent with the patterns reported by Taylor et al. (2004).

The studies discussed above cover a subset of the loci evaluated here (Tv19, Tv27, Tv53, and Tv58). Existing information on allelic diversity and heterozygozity patterns of the remaining loci included in this panel is scarce. Locus Tv54 was known to have low allelic diversity when first reported (three alleles Taylor et al., 2000) but has been rarely used since. However, in the few publications where it is reported, allelic diversity and heterozygosity levels are consistently low (Banks et al., 2010). With regards to locus TvM1, this thesis constitutes the first account of genetic diversity patterns for this locus in the South Island. Lam et al. (2000), reported data on diversity and heterozygosity for TvM1 in two localities in the North Island. Since then, TvM1 has been used to infer dispersal patterns of possums within their native range (Stow et al., 2006) and the relationship of microsatellite heterozygosity with survival in a closely related species (*T. cunninghami*) (Banks et al., 2008). It is encouraging to see that allele numbers and heterozygosity levels for TvM1 reported in this study are very similar to what has been already published, despite samples having come

from distinct geographic areas. Locus Tv5.64 was first used in a study of possum dispersal in Hawkes Bay (Cowan et al., 2007). Unfortunately at the time of write-up, it was not possible to access the details of that work. A later study conducted on populations in the South Island (Lewis Pass and Selwyn) was able to characterise the allelic diversity of Tv5.64 using polyacrilamide gel electrophoresis (Mahood et al., unpublished). Mahood et al.'s study was an early attempt to amplify DNA retrieved from saliva, in which the authors reported seven alleles and a high level of expected heterozygosity (0.75). While the findings of the later study are in close agreement to what is observed here, differences in the technological approaches used to score alleles across studies would potentially account for the differences in allele sizing and diversity observed.

The significant departures from HW equilibrium observed in samples from Hororata and Lewis Pass are difficult to interpret. Departures from equilibrium should indicate that one, or many, of several evolutionary processes assumed in the HW model are operating in a particular subpopulation from where the sample is taken (Allendorf & Luikart, 2007; Freeland et al., 2011; Hamilton, 2009; Hartl & Clark, 2007). Individuals collected in this study are but a sample of larger subpopulations, or a combination of them; and thus may or may not accurately represent the situation in the actual subpopulations these individuals were drawn from. Because of the restrictive nature of HW's assumptions, it is expected that eventually a significant departure from equilibrium will be observed if the effect of the evolutionary process of interest is strong enough to be detected by available analytical procedures (Hartl & Clark, 2007). The difficulty lies in interpreting which process — or combination of — is acting on each particular case.

Having laid down the pitfalls of interpreting HW test results, it is nec-

essary to discuss the assumptions that make up the HW model, in order gain insight into which processes might be responsible for the observed departures. Despite the fact that mutation rates of microsatellites are known to be higher than the rest of the genome (see section 1.3), it is reasonable to assume that mutation had no influence on the observed results given the relatively short time possums have inhabited in New Zealand ($\sim$ 150 years, Pracy, 1974 cited in Cowan 1998). Similarly, as microsatellites are considered neutral markers, it seems unlikely that natural selection is the cause of the disequilibria. The last statement is open to debate, given that recent studies have suggested an association between the increase in the number of repeats of microsatellite loci and some human genetic diseases (Oliveira et al., 2006), which will indicate they are affected by selective pressures. Quite notably, a study on the microsatellite diversity of a closely related species of possum (*Trichosurus cunninghami*) found an association between survival and elevated heterozygosity at the major histocompatibility complex linked locus TvM1 (Banks et al., 2010). Nonetheless, in the absence of similar evidence for *T. vulpecula*, the assumption that loci included in this study are not under selection appears valid. With regards to random mating, existing evidence on the mating system of possums is inconclusive. Previous research has found that mate fidelity is uncommon and there is little variation in reproductive success between male possums in New Zealand (Taylor et al., 2000). In contrast, evidence of substantial differences in male mating success has been found in two further separate studies from New Zealand and Australia (Sarre et al., 2000; Clinchy et al., 2004). The variation in results on possum mating systems may be a result of local or temporal conditions, and is expected, as it has been acknowledged that possums in New Zealand exhibit a high plasticity in their behaviour that enables

them to adapt and thrive under the most adverse circumstances (Day et al., 2000). Even if random mating is assumed to be the rule for New Zealand populations, there will be a limit to this behaviour imposed by distance or geographic barriers (e.g. isolation by distance, Hamilton 2009). Although the effect of isolation by distance may be attenuated by the tendency of juvenile possums to actively disperse away from their native home ranges (average of 5 km, Cowan & Clout 2000), mere migration is not necessarily an indication of gene flow. In the absence of evidence that clearly points in any direction, the only possible conclusion is that random mating might occur in some subpopulations, but not in all, and not consistently through time.

Based on the discussion above, it is clearly difficult to single out which evolutionary processes is responsible for the observed departures. What seems more likely in the context of this study is that the inclusion of related individuals, or of individuals of overlapping generations in some of the samples, might be having a disproportionate effect on the observed test results. Alternatively, the HW test could be picking up the signature of the Wahlund effect (Allendorf & Luikart, 2007; Hamilton, 2009; Freeland et al., 2011), which would indicate the samples contain a mixture of individuals from two separate subpopulations. This last possibility is supported by the partition of HW tests among loci, where an overall heterozygote deficit for the majority of loci in Hororata and Lewis Pass is apparent, while the remaining loci exhibit heterozygote excess. However, the hypothesis of Wahlund effect is incompatible with the lack of genetic structuring among localities in the South Island found in this study.

The evidence for a lack of genetic structuring among localities found here is consistent with previous studies. Taylor et al. (2004) reported a lack of

evidence for structuring across localities in the South Island, explained as a combination of founder effects and population bottlenecks. Given that: i) possum introductions are a relatively recent event in New Zealand; ii) there is historical evidence suggesting the founding population was likely to have been constituted by a few individuals brought from Tasmania; and iii) populations of this species are frequently controlled; it is reasonable to think that founder effects added to a population bottleneck caused by slow growth of the founder population is responsible for the patterns observed. This homogenenisation effect among South Island localities might have been exacerbated by frequent turnover of migrating individuals and random mating. Although these patterns should have changed since the adoption of population control, the relatively recent nature of control measures may mean that as yet the effect is not strong enough to detect.

The PI_{sib} values estimated in this experiment confirm that the assembled microsatellite panel is sufficiently robust to achieve reliable individual identification of possums by their multi-locus genotype. With this probability of identity adjusted for siblings, only three possums in a thousand are expected to have matching multi-locus genotypes by chance alone. According to the literature, this is an acceptable value to conduct population size estimates (Taberlet & Luikart, 1999; Waits et al., 2001; Mills et al., 2000). For example, in a study of red deer (*Cervus elaphus*), Valiere et al. (2007) estimated a PI_{sib} of 1.17 $\times 10^{-2}$ based on tissue of 40 individuals. This study had a final panel composed by 6 microsatellites. Similarly, Adams & Waits (2007) reported a PI_{sib} of 7.25 $\times 10^{-3}$ in the red wolf (*Canis rufus*), estimated for a panel of 7 loci using blood of 175 individuals.

Because PI_{sib} is a direct function of locus allelic frequencies, the robustness of the panel might vary from one subpopulation to the next, as

effectively corroborated here. As suggested by Frantz et al. (2003), it is also possible that the number of loci necessary to attain reliable individual identification may vary between regions or groups within species (e.g. males versus females). Regardless, it is encouraging to see that because of the reduced level of genetic structure observed among samples across Canterbury, the robustness of the present panel remained relatively constant across subpopulations. While it would have been desirable to obtain a multi-locus genotype with all eight loci originally selected, the exclusion of locus Tv16 was necessary given the time constraints imposed by a study of this nature. Even if it were possible to optimise PCR conditions for this particular locus, there is strong evidence that suggests this locus is physically linked to locus Tv27 (Taylor et al., 2004). It follows that its exclusion from the panel is warranted on these grounds alone.

The mean error rate per locus reported here is unexpectedly higher than what has been reported for similar pilot studies. Bonin et al. (2004) found an mean error rate per locus of 0.8% and a observed error rate per multi-locus genotype of 17.6% during a study of microsatellites of the brown bear (*Ursus arctos*). Whilst the overall error rates reported here are apparently high, there are several studies that have reported non-negligible rates of allele drop-out and false alleles when using good quantities and quality of DNA as template (Soulsbury et al., 2007; Hoffman & Amos, 2005; Bonin et al., 2004). It is clear that despite the widespread assumption that DNA of good quality should not suffer from the occurrence of stochastic genotyping error, the implementation of quality control systems proves this assumption false. The error identification approach adopted in this study also revealed that half of the errors detected can be attributed to systematic genotyping errors (i.e. sporadic contamination or sample confusion). This is consistent with

reports of other studies that detected a high prevalence of systematic genotyping errors (Hoffman & Amos, 2005; Paetkau, 2003; Fernando et al., 2003), and possibly reflects the level of training and lack of previous experience in genotyping techniques of the author. While the occurrence of systematic genotyping errors is undesired, it is reasonable to expect that human originated errors can be minimised as personnel experience increases. Finally, it is encouraging to report that despite HW equilibrium tests suggesting the potential occurrence of null alleles at locus Tv58, results of specialised tests do not support this conclusion (see section 2.3.5). van Oosterhout et al. (2004) developed a method to discriminate whether the signature of significant heterozygote deficit is due to null alleles, or other causes such as deviations from panmixia or the Wahlund effect. In the van Oosterhout et al. approach, if the frequency of homozygotes constituted by the different alleles detected at a locus is equivalent, it is considered as the signature of null alleles occurrence. Taylor et al. (2000) also dismissed the possibility of this locus suffering from null alleles. It is anticipated that with further experimentation, detection of stochastic errors will be increased, in particular allelic drop-out and false alleles. This issue will be re-addressed further in the following chapters.

2.5 Conclusions

Microsatellite diversity patterns observed in this study are consistent with previous findings. Loci included in the panel, with the exception of locus Tv54, are sufficiently polymorphic to warrant inclusion. In this respect, it is also important to conclude that the optimised microsatellite panel provides sufficient statistical power to identify individuals in Canterbury by their multi-locus genotype. However, this conclusion cannot be generalised to

other areas where microsatellite diversity patterns might differ. In such cases the inclusion of additional markers in the existing panel may be necessary.

Given the observed departures from HW equilibrium, possibly reflecting the inclusion of related individuals or individuals of overlapping generations in the samples, it is difficult to arrive at a meaningful conclusion in this regard. However, the partition of HW test per locus and locality seems to confirm that loci evaluated here can be assumed to be at equilibrium. Similarly, because of limited sample sizes, it is not possible to definitely conclude that there is no genetic structure in the sampled localities, despite evidence pointing in that direction.

The overall prevalence of genotyping error is moderate, not tending to affect any loci or individual in particular. Although routine assessment and reporting of the prevalence and nature of stochastic and systematic error rates is uncommon, results found in this study are in accordance with the literature. Careful inspection of profiles after automated scoring of alleles allowed association of particular errors to their possible causes. This revealed that half of the errors originated from human mistakes and thus are expected to be mitigated with increased experience and training. Specific experiments that test if this trend holds when low quantities and quality of DNA are used as template are required to fully validate loci in the panel.

Given the low polymorphism and heterozygosity of locus Tv54, its poor contribution to the overall PI_{sib}, and the high incidence of adjacent alleles in this locus, exclusion of locus Tv54 from future assays is recommended.

Chapter 3

Development of a robust multiplex assay

3.1 Introduction

In the last decade, numerous technical improvements targeted to increase the efficiency of non-invasive genotyping regimes have been developed (reviewed in Beja-Pereira et al. 2009). Yet these techniques still often need to be tailored to individual projects needs. Such a situation presents an opportunity for a pilot study to be conducted. Doing so contributes to providing evidence supporting the inclusion or exclusion of a particular technique in a given genotyping regime (Gervasi et al., 2010; Valiere et al., 2007; Adams & Waits, 2007). Once optimised, amplification of DNA and subsequent typing - both labour intensive and expensive tasks, can process a large number of samples with reduced resource input.

One technique expected to optimise the genotyping process is PCR multiplexing. PCR multiplexing was introduced for the first time more than two decades ago (Chamberlain et al., 1988). Unlike standard PCR (hereafter re-

ferred as singleplex), multiplexing allows for the simultaneous amplification of several loci on a single reaction, providing high throughput while requiring a lower quantity of DNA template. Increased amplification efficiency and economy of resources are desirable characteristics, since a successful combination of this amplification strategy with highly automated resolution techniques, such as capillary electrophoresis, can boost the genotyping process. As a result, multiplex PCR is often portrayed as a cost-effective way to increase the efficiency of sample processing and potentially overcome the difficulties of dealing with low quantity and/or low quality DNA (Arandjelovic et al., 2009; Beja-Pereira et al., 2009; Butler, 2005).

However, the development of a robust multiplex assay requires careful and demanding calibration to avoid incompatibilities among primers, and to ensure a balanced yield of product across loci (Butler, 2005), which imposes a limit to its application. When the development of a multiplex assay becomes impractical, an alternative approach, known as pseudo-multiplexing (or pool-plexing, Guichoux et al. 2011b), can also guarantee a reduction in the costs of obtaining a genotype. Pseudo-multiplex differs from multiplexing in that PCR products of different loci are obtained using singleplex reactions that are subsequently combined during capillary electrophoresis. While pseudo-multiplexing uses more resources than multiplexing, it circumvents the multiplex assay lengthy calibration process, and thus can prove more convenient in the long term.

Aside from the problems of optimising a multiplex protocol, any modification to the protocol used to generate alleles needs to be tested for precision and consistency. In some situations, allele size estimates of the same allele that have been generated with differing protocols might suffer discrepancies. These discrepancies could result from differences in the allele migration pat-

terns through capillary electrophoresis (i.e. mobility shifts, see Section 1.4 and Figure 3.1). While recent literature has recognised the issue of mobility shifts, in particular when conditions of capillary electrophoresis vary across laboratories (Ellis et al., 2011; Pasqualotto et al., 2007; Moran et al., 2006), there has been no formal comparison of the effects of implementing a multiplex amplification strategy on the electrophoretic mobility of fragments and the implications for result reproducibility.

This chapter explores the construction and implementation of a multiplex assay and its variations using six of the loci refined in the previous chapter. While the primary objective was to optimise the genotyping process, the opportunity was also taken to quantify the economic benefits of adopting multiplex PCR as a protocol for genotyping.

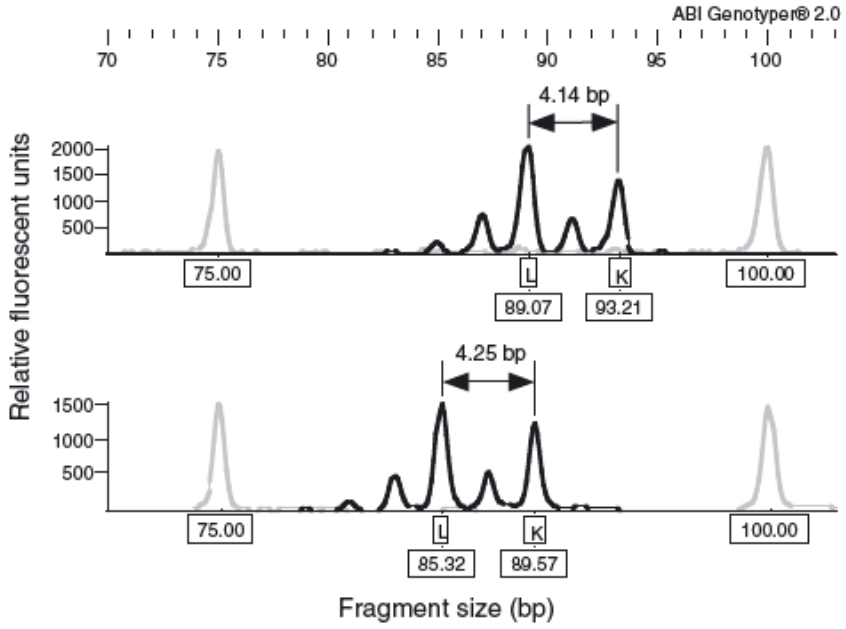

Figure 3.1: Illustration of a size shift published in Moran et al. (2006). The black peaks in the figure represent the same allele but the profiles were obtained with capillary electrophoresis run on different machines. The grey peaks represent the size standard. This figure is reproduced with permission of the author (Moran, pers. comm. 2012)

3.2 Materials and methods

3.2.1 Data preparation

DNA extracts

Fresh DNA was extracted from the tissue of 27 individuals collected for the previous experiment (see chapter 2) and used as template for multiplex validation. An aliquot of 100 μL of the original DNA extract was separated into an independent tube.

In order to standardise DNA concentration across samples, aliquots of DNA extracts were measured using a NanoDrop ND-1000 spectrophotometer. Based on these measurements, the extract with the minimum DNA concentration (27.5 ng/μL) was identified and the remaining extracts were diluted to that concentration by adding appropriate volumes of Buffer AE (QIAGEN).

3.2.2 Multiplex construction

Optimal primer combination

Information on the characteristics of six loci (Tv19, Tv27, Tv53, Tv58, Tv5.64 and TvM1, as generated in Section 2.2.3) was introduced into specialised software for multiplex design. MULTIPLEX MANAGER v.1.0 is open source software that uses an annealing algorithm (Kirkpatrick et al., 1983) to simulate thousands of virtual PCR reactions (Holleley & Geerts, 2009). On each simulated PCR, the algorithm trials different primer combinations. Finally, the software generates a suitability score of the top ranked primer combinations and a graphical output that enables the user to interpret the results.

Loci characteristics, such as optimal annealing temperature, allele size

range, forward and reverse primer sequence and fluorochrome dye class, were introduced into the software. The control parameters to generate primer pair combinations were: i) to minimise the number of reactions; ii) to maximise the space between markers; iii) to minimise annealing temperature differences among loci in the same reaction (Holleley & Geerts, 2009). In order to minimise cross-reactivity of primer pairs (e.g. intra primer binding or hairpin structure formation Vallone & Butler 2004; Shen et al. 2010), the complementarity threshold was set to 7. The complementary threshold is the maximum number of AT or CG matches minus the number of mismatches for any two primers within a multiplex reaction, and is set to seven because empirical work has demonstrated that is the maximum number required for multiplex PCR to work (Butler et al., 2001). Since the number of primer sets attempted to multiplex here was small, it was possible to run 1 $\times 10^9$ iterations of the algorithm.

The first ranking solution generated by the software was taken into account as a template for empirical validation.

Multiplex empirical validation

How comparable the output of MULTIPLEX MANAGER is with a real PCR depends on the accuracy of the information provided to the program. However, it does not guarantee a suitable multiplex assembly. Therefore, it is necessary to validate the output by reproducing the most suitable solutions under real laboratory conditions.

Taking into account the MULTIPLEX MANAGER results, primer pairs were combined in 100 μL aliquots. The optimal concentration in which a readable product was generated in singleplex reactions was used to build working primer solutions (see Table 2.2). Working solutions were created

according to the method described in section 2.2.3. Primer solutions were then used to amplify DNA extracts prepared in section 3.2.1.

A multiplex PCR kit (QIAGEN) was used for amplification, with slight modifications to the manufacturer's standard multiplex protocol. Whilst the standard QIAGEN protocol is set for a total reaction volume of 25 μL, ten μL reactions were optimised here. The reactions contained: 2.5 μL of template, 5 μL of 2x multiplex PCR kit master mix (final concentration 1x), 0,5 μL of RNase free water and 1 μL of forward and reverse primer solution. All PCRs were performed with a MultiGene TC9600-G thermal cycler. The thermal cycler profile for the first ranked solution was as follows: denaturation at 95 °C for 15 min, then 25 cycles of denaturation at 94 °C for 30 sec, annealing temperature (57.8 °C) for 90 sec and extension at 72 °C for 90 sec, followed by a final extension time at 72 °C for 10 min.

PCR products were visualised with capillary electrophoresis and alleles were scored as described in section 2.2.4.

Pseudo-multiplex

Pseudo-multiplexing was trialled in the previous chapter when generating profiles of singleplex reactions. A minimum of two and a maximum of six PCR products were loaded simultaneously — by adding 0.5 μL of each — in the same well. Subsequent scoring of alleles remained unmodified from what was described for the singleplex approach in chapter 2.

3.2.3 Data analysis

Reproducibility of results

To assess whether multiplex reactions affected the reproducibility of results, allele size estimates generated with multiplex assays were compared with

allele size estimates of the same individuals generated with the singleplex approach (and in some instances, the pseudo-multiplex approach). Observed allele size estimates were obtained with the "GenotypeTable" routine in MICROSATELIGHT.

Prior to comparing allele measurements, profiles were visually inspected to ensure correct detection of all the peaks present on a profile. Repeated allele measurements were subsequently cross-referenced. Instances of sample confusion or contamination, allele drop-out and false alleles were removed from the dataset. The remaining discrepancies, such as as mobility shifts, were tallied and included in the final dataset.

Inspection of the dataset allowed errors to be associated with their potential causes. The same error categories outlined in section 2.2.6 were considered and recorded. Mismatches associated with stochastic and scoring errors were taken into account to estimate mean error rate per locus and observed error rate per multi-locus genotype in the same fashion as described in Section 2.2.6. When new instances of allele drop-out or false alleles were detected, the correspondent rates were re-estimated.

Once the data set was inspected and cleaned, the correspondence of measurements between singleplex and multiplex approaches was assessed. Size differences were calculated and the mean and standard deviation of these differences was estimated per locus. Allele size estimates obtained from singleplex and mutliplex approaches were subsequently plotted against allele size categories (i.e. bins), designated in chapter 2 section 2.3. Allele categories assigned in chapter 2 were assumed to reflect the true size of each allele. Although the last assumption is unlikely to hold, this approach was taken for analytic purposes and by no means intends to determine the actual size of each allele. Finally, the correspondence of raw allele size to the called

alleles was fitted with linear regression.

To determine whether there were significant size estimate differences between products of single versus multiplex amplification, an analysis of covariance (ANCOVA) for each locus was ran as follows:

$$X_{var} \sim Y_{var} * factor \tag{3.1}$$

where factor is the amplification strategy with levels singleplex and multiplex (i.e. the categorical variable), Y_{var} is the allele size estimate and represents the response variable and X_{var} is the allele category, which is numerical discrete, and represents the covariate. If the interaction between factor and covariate was not significant, a second and more parsimonious model to test whether there were significant size estimate differences between singleplex versus multiplex amplification was formulated as follows:

$$X_{var} \sim Y_{var} + factor \tag{3.2}$$

A 95% confidence interval of linear regression parameters (i.e. slope and intercept) for each loci was estimated post-hoc to assess the change in proportions of the relationship between co-variables.

The effect of replacing the amplification strategy (e.g. singleplex replaced by multiplex) on the mean error rate per locus was analysed with a generalised linear model with a binomial error distribution and a logit link function.

All analyses were conducted in R (R Development Core Team, 2011) as implemented in the program R Commander (Fox, 2005).

Economic benefits of multiplex implementation

To quantify the economic advantages of implementing the outlined techniques, cost estimations of genotyping a specimen applying different techniques were contrasted. The cost of genotyping was estimated on the basis of current reagent prices in consideration to six loci. Labour costs were excluded from this analysis.

The genotyping cost per specimen was estimated with the following equation:

$$G = R_t(\sum \frac{r_v}{p} + e_w) \tag{3.3}$$

where R_t is the total number of reactions required for each technique, r_v is total number of possible 10 μL PCRs given the volume of the reagent provided by suppliers, p is the price of that particular reagent and e_w is the current cost per well of running capillary electrophoresis.

An example of the calculations above follows. Based on the total amount of primer supplied by the manufacturer (10,000 pmoles, see Section 2.2.3) the total volume of working strength solutions (assuming an average concentration of 5 μM) is 2000 μL. This volume, divided by the average volume of primer used in each reaction (2μL) gives a total of 1000 possible reactions (parameter r_t in equation 3.3). This product divided by the current price of a primer pair, 250 NZD (p in equation 3.3), gives a unit price of 0.25 NZD per reaction. Similar calculations were used to obtain the unit price of reagents used in each particular case.

Singleplex unit cost considered the price of the *i*-StarTaq kit, primers and capillary electrophoresis reagents. In contrast, the pseudo-multiplex unit price was estimated with a slight modification of equation 3.3 which

corrected for the price of parameter e_w as follows:

$$G_{pm} = [R_t(\sum \frac{r_v}{p})] + \frac{e_w}{6} \tag{3.4}$$

The equation above assumes it is possible to combine PCR products of all six loci into one well.

Finally, the cost of genotyping with the multiplex technique was estimated using equation 3.4, but also considered the cost of multiplex PCR kits, primers and capillary electrophoresis reagents.

3.3 Results

3.3.1 Multiplex optimisation

MULTIPLEX MANAGER results suggested the possibility of combining all six loci in one single reaction (Figure 3.2). The next ranked result indicated the possibility of constructing two multiplexes of three loci each. As the objective was to minimise the number of reactions, the best ranked solution was chosen for empirical testing.

When DNA extracts were amplified using the suggested multiplex assay, all specimens were amplified successfully. Profiles exhibited sharp peaks (i.e. narrow and well defined), and homogeneous amplification across loci (Figure 3.3).

Pseudo-multiplexing was successfully trialled while generating some of the genetic profiles in chapter 2. Genotypes generated with this approach also exhibited sharp peaks. Despite observation of some variation in the intensity of peaks across loci, the simultaneous addition of PCR products generated on different singleplex reactions did not appear to effect the resolution of capillary electrophoresis, nor to have generated negative interactions

Figure 3.2: MULTIPLEX MANAGER optimal primer set combination. Boxes represent each of the six loci included in the multiplex assay. The scale on top of the figure is in base pairs and the axis on the left displays the suggested annealing temperature for PCR

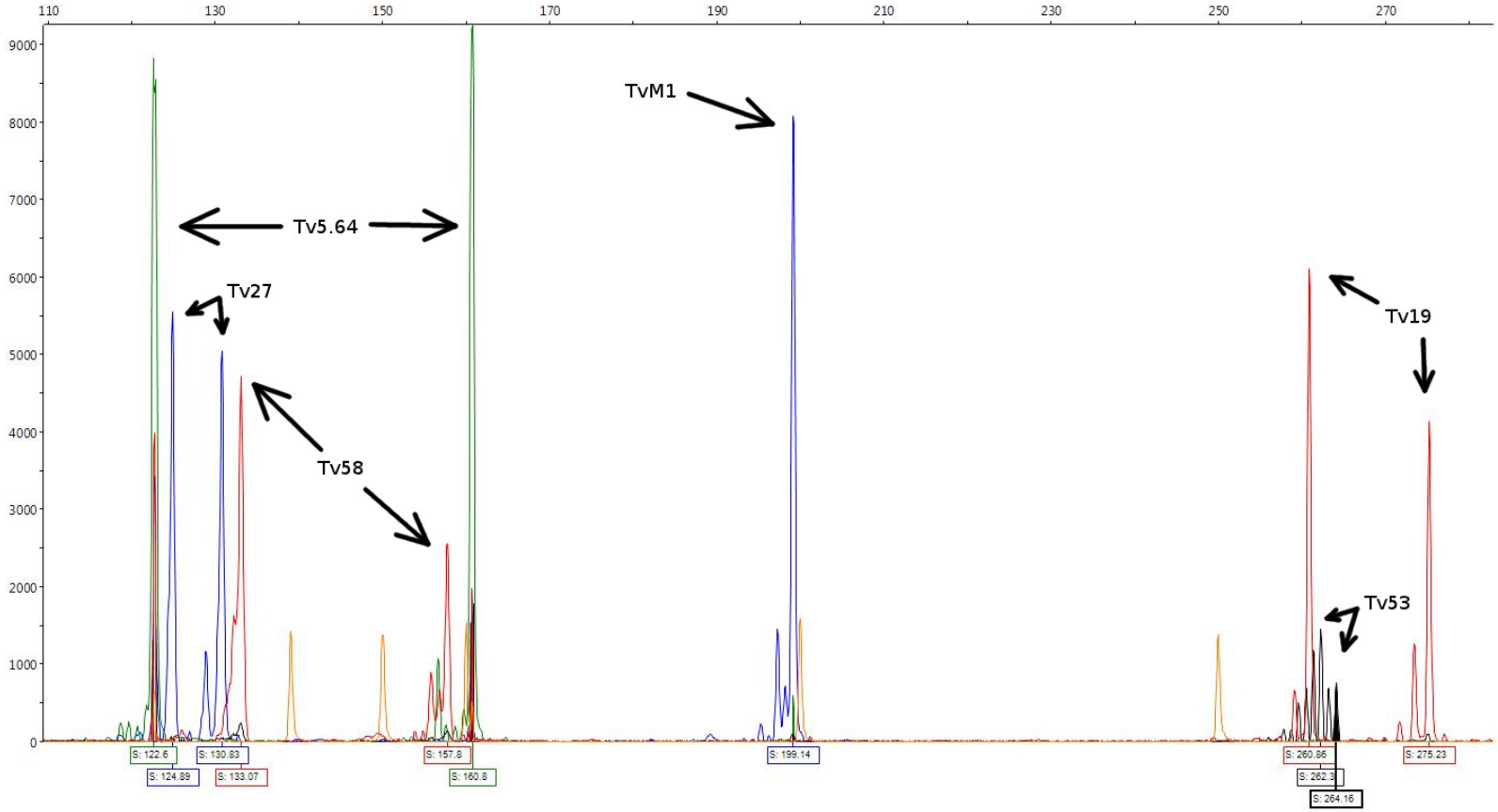

Figure 3.3: Profile of individual 3 as visualised in Peak Scanner v1.0. The scale on top of the graph is in base pairs and the x-axis is in relative fluorescent units. A legend with the estimated allele size is shown below each peak

between alleles of PCR products.

Multiplex evaluation

Capillary electrophoresis produced different allele size estimates between fragments amplified by singleplex and multiplex reactions. Locus Tv53 showed the largest difference in base pairs (mean, range: 0.95, -2.05–1.45) followed by locus Tv58 and TvM1 (0.68, -1.42–0.75; 0.11, -0.44–0.30 respectively). The largest difference between allele size estimates was observed in locus Tv19 (0.06, -5.75–5.58). Locus Tv27 and Tv5.64 showed the lowest differences (0.07, -0.37–0.35; 0.02, -0.98–0.96 respectively).

Size estimates of alleles generated with singleplex and multiplex approaches fitted well with designated allele categories. The first ANCOVA model showed the effects of the interaction between covariate and treatment was not significantly different from zero at any loci (data not shown), which indicates the slopes of the regression between allele size estimates and allele size categories are similar across amplification regimes. The second model (equation 3.2) showed significant covariate effects across all loci and a significant treatment effect at three loci: Tv53, Tv58 and TvM1 (Table 3.1). Significant covariate effects suggests there is one regression line regardless of the factor level. A significant covariate effect plus a significant treatment indicates that although regression lines have similar slopes their intercepts are different. The difference in intercepts suggests that size estimates obtained from different amplification regimes are different. Differences in intercept values can be confirmed in Table 3.1 and visualised in Figure 3.4. Despite these effects, intercept confidence intervals of regression lines at locus Tv53, Tv58 and TvM1 overlap, suggesting the relationship with the covariate is still approximately 1:1.

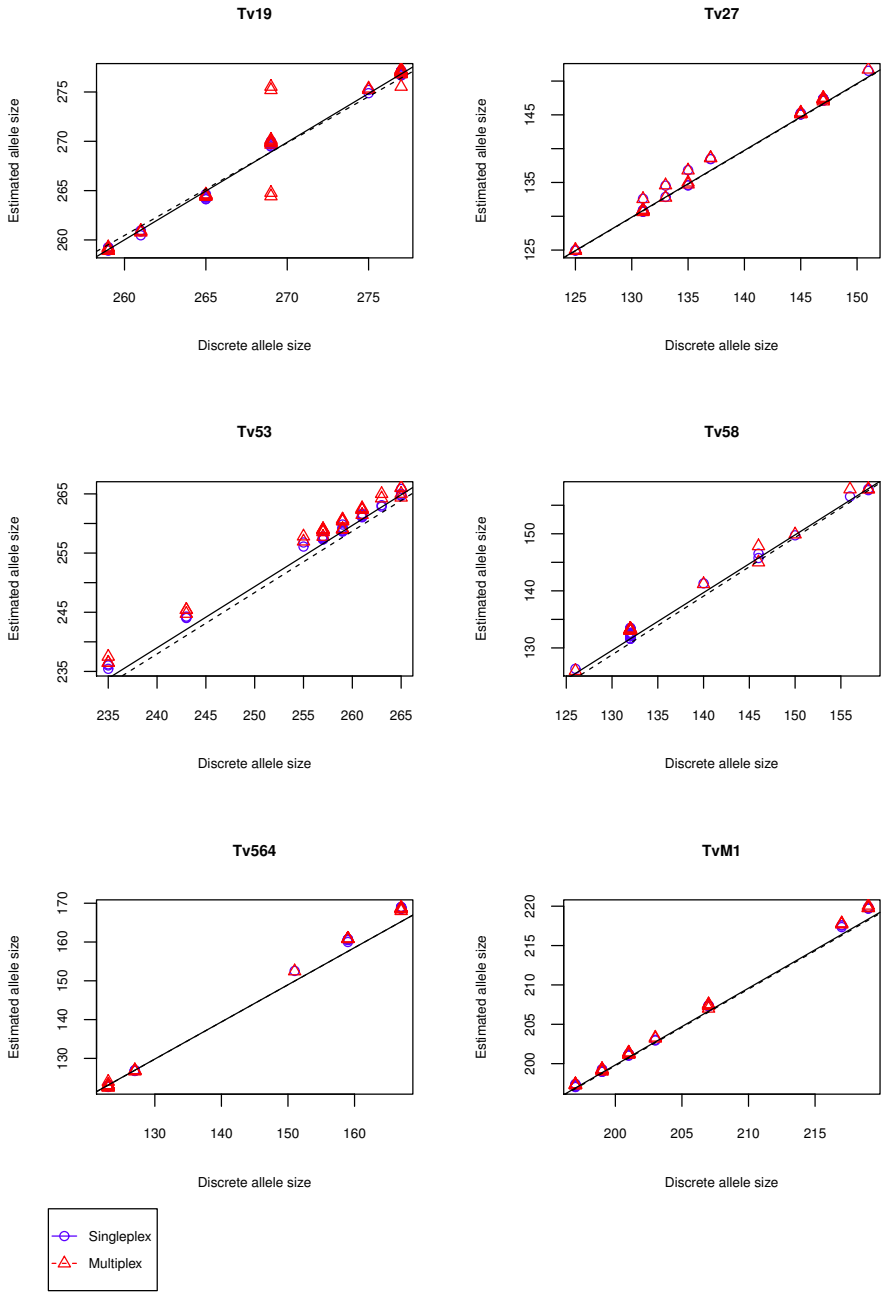

Figure 3.4: Linear regression models of allele sizes estimates of products generated by singleplex and multiplex PCR approaches versus discrete allele categories designated in chapter 3

Table 3.1: Summary of linear regression parameters between amplification approaches. *P*-values correspond to ANCOVA's results of treatment effects. A non-significant result suggests there is no significant difference between the intercept values - i.e. there is no treatment effect

	Singleplex		Multiplex		
Locus	Intercept	CI	Intercept	CI	*P*-value
Tv19	2.81	8.53, -2.92	16.34	35.10, -2.42	0.799
Tv27	1.41	4.67,-1.86	1.52	4.77, -1.73	0.595
Tv53	-9.21	-4.52, -13.90	-10.99	-2.79, -19.18	< 0.001
Tv58	-1.90	0.96,-4.75	-4.89	-2.05, -7.72	< 0.001
Tv564	5.38	5.76, 5.01	5.28	5.92, 4.64	0.745
TvM1	4.81	5.97, 3.66	5.12	6.21, 4.02	< 0.001

The mean error rate per locus was 7% yielding an overall observed error rate per multi-locus genotype of 25%. In a more detailed partition of error, locus Tv27, Tv53 and Tv58 were affected by allele drop-out yielding an overall weighted average of allele drop-out rate of 5%. All loci except locus Tv19 were affected by the scoring of false alleles, producing a weighted average of false allele rate of 7% (see Table 3.2 for per locus false allele rate). Visual inspection of the profiles confirmed that the majority of these discrepancies were due to a great intensity of stuttering peaks which were mistakenly scored as true alleles. MICROSATELIGHT peak detection routine failed to detect the presence of peaks in six occasions, primarily affecting locus TvM1. Similarly, the automated peak detection routine mistakenly recognised the off-scale fluorescent signal of peaks corresponding to alleles of locus Tv5.64 as alleles of locus Tv58 (termed here as fluorochrome confusion). The former two error classes were corrected in the final dataset. Allele mobility shifts

Table 3.2: Summary of genotyping error categories found by comparing singleplex replicates with multiplex replicates

	Tv19	Tv27	Tv53	Tv58	Tv564	TvM1
Per locus allele drop-out rate	0.00	0.09	0.11	0.13	0.00	0.00
Per locus false allele rate	0.00	0.04	0.6	0.02	0.04	0.04
Fluorophore confusion	0	0	0	3	0	0
Mobility shift	4	0	0	0	0	0
Undetected peaks	0	0	1	1	1	3

were detected but only found to affect locus Tv19. Finally, three instances of contamination were detected.

Implementing multiplex as the amplification strategy did not have a significant effect on the incidence of mean error rate per locus ($\chi^2 = 0.646$, $d.f. = 1$, P-value $= 0.421$).

3.3.2 Economical benefits of multiplex implementation

After estimating the cost of obtaining a genotype with the three different protocols outlined above, calculations showed that the implementation of a multiplex assay allowed for a substantial reduction in the total price incurred by running a six-locus genotype per specimen. While the processing price per unit was lower with the pseudo-multiplex approach, the reduction in the total number of PCRs required with the multiplex assay decreased the overall resource usage (Table 3.3).

A projection of the cost of processing large numbers of specimens confirms the economic advantages of implementing true multiplexing over other amplification strategies. Figure 3.5 projects a reduction of 92.28 % in the total cost of genotyping 1000 specimens when implementing a multiplex ap-

Table 3.3: PCR costs (NZD) per sample unit as incurred by different amplification approaches

Genotyping factors	Singleplex	Pseudo-Multiplex	Multiplex
PCR constituents	0.35	0.35	1.00
Primers	0.25	0.25	0.25
Capillary electrophoresis	7.50	1.25	1.25
Unit price per PCR	8.10	1.85	2.50
Number of PCRs required	6	6	1
Total genotyping cost (NZD)	48.57	5.42	3.75

proach. Pseudo-multiplex followed closely with a reduction of 88.85% of the total cost of genotyping the same number of individuals with singleplex approach.

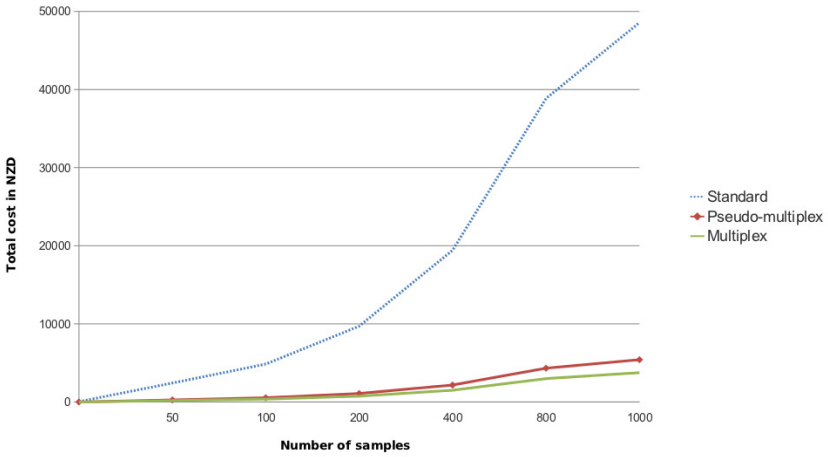

Figure 3.5: Projection of the cost of genotyping with different amplification approaches

3.4 Discussion

The evidence presented here demonstrates that it is possible to assemble a multiplex assay with six of the loci refined in Chapter 2, with little optimisation. This represents a substantial reduction in the resources required for sample processing without compromising the reproducibility and reliability of the results.

While multiplexing has still not been widely implemented in the context of NIGM, its potential to improve the efficiency of monitoring regimes has been acknowledged repeatedly (Guichoux et al., 2011a; Beja-Pereira et al., 2009). Consistent with what has been found here, in recent years the number of case studies that have adopted PCR multiplexing in the context of NIGM has increased (Luikart et al., 2008b; Martin-Galvez et al., 2011; Valiere et al., 2007; Skrbinsek et al., 2010). The uptake of multiplexing as a suitable amplification strategy for wildlife monitoring might be further facilitated in the medium term thanks to the continued development of ever more sophisticated software dedicated to constructing multiplex assays (Vallone & Butler, 2004; Shen et al., 2010), refined guidelines for multiplex development (Guichoux et al., 2011a), and the adoption of faster and more economic techniques to discover candidate microsatellite loci (e.g. 454 pyrrosequencing, Lepais & Bacles, 2011).

Under certain circumstances, the benefits of a multiplex amplification strategy targeted at amplifying non-invasively collected samples might be hampered by the costs of developing a robust protocol, particularly if existing primers need to be redesigned or a two stage multiplex strategy is being considered (Arandjelovic et al., 2009; De Barba & Waits, 2010). Here the use of amplification kits developed for multiplexing, the prior refinement of loci involved, and the use of highly concentrated fresh DNA as the

template might have contributed to the positive results observed. It follows that further refinement will be necessary when the protocol is used with low quantities and quality of DNA.

The implementation of a multiplex PCR assay appears not to have affected the consistency of allele size estimation across loci. While a significant treatment effect was detected for loci Tv53, Tv58 and TvM1, the fact that the intercepts confidence intervals of linear models at these loci overlap suggests that these differences could be considered part of the standard error. Standard error in the sizing of alleles between replicates of the same samples are seldom reported, but error is known to occur even when using the same protocols. For instance, in a study that demonstrated non-negligible effects in allele sizing when using different fluorochrome dyes on the same locus, Sutton et al. (2011) found a within-dye difference in sizing of 0 to 0.25 bp in replicates that were generated with the same protocols. Haberl & Tautz (1999) in turn found a small range of mean differences between replicates of selected alleles (0.07–0.14), and Pasqualotto et al. (2007) reported a mean difference of 0.2 and a range of 0.0–1.0 for alleles sized in the same machine.

Although small differences in the sizing of alleles between replicates is expected, cross-referencing confirmed the presence of a compelling source of discrepancies in the dataset: mobility shifts. While inspecting the possible cause of mobility shift cases, it was found that singleplex PCR products of the specimens that displayed this type of error were submitted to a different sequencing machine for capillary electrophoresis (located at Canterbury University School of Biological Sciences). It is possible that cross-platform differences in the chemical conditions of capillary electrophoresis are the cause of the large differences in allele size estimates observed at this locus, as has been already suggested in the literature (Pasqualotto et al., 2007;

Haberl & Tautz, 1999; Ellis et al., 2011). It is remarkable that despite this issue, the mean difference between replicates in these two loci is still one of the lowest registered, demonstrating a high level of consistency in allele sizing provided samples are measured under the same electrophoretic conditions.

The large mean difference in size estimates observed for locus Tv53 highlighted the issue of false allele occurrence. It has been suggested that highly polymorphic loci that amplify larger fragments are more prone to error and should be avoided or their primers redesigned to amplify smaller fragments (Hoffman & Amos, 2005). This might be the case for locus Tv53, which typically exhibited what has been termed "hedgehog" topography (Ellis et al., 2011). That is, high intensity of stuttering peaks and high prevalence of adjacent alleles (i.e. alleles separated by one repeat motif). Hedgehog profiles makes the task of recognising the true underlying allele difficult, even with the help of automated software. Despite a lengthy optimisation process, profiles of locus Tv53 still displayed a high incidence of stuttering bands, which potentially warrants primer re-design. Alternatively, Tv53 could be excluded from the multiplex assay and its amplification trialled in standard reactions using next generation polymerases (e.g. fuzion enzymes Fazekas et al., 2010) or by decreasing the denaturation temperature during the initial step of the PCR (Olejniczak & Krzyzosiak, 2006). Since these solutions require further time and resources beyond the scope of this thesis, the most practical solution is to consider the addition of new microsatellite loci to the panel in order to replace Tv53.

Given that a greater number of samples were replicated in this trial (n = 27), the observed error rate per multi-locus genotype decreased in relation to the rate reported in the previous chapter (33.33%). In addition, one

more locus was found to be affected by allele drop-out (Tv53), while the the remaining two showed a reduction in the per locus allele drop-out rate. This reduction in rates suggests a slight over-estimation of these parameters in the previous chapter, potentially due to the low sample size of replicated experiments ($n = 8$). In contrast, the number of loci affected by false allele occurrence increased noticeably in relation to the previous experiment, driving the overall increase in the f_w. It is possible that the change in amplification parameters implemented for multiplexing, in particular the change in annealing temperature, increased the intensity of stuttering peaks and thus the scoring of false alleles. Such non-specific amplification has been suggested as one of the causes preventing the widespread adoption of multiplexing for wildlife forensic applications (Karaiskou & Primmer, 2008). An alternative explanation for non-specific amplification is the use of suboptimal concentrations of template DNA, which has been deemed responsible for exacerbating stuttering issues. For instance, human forensic scientists have considered it necessary to establish a narrow but optimal range of template DNA concentrations for microsatellite amplification systems in order to mitigate the occurrence of stuttering peaks (Kline et al., 2005; Nicklas & Buel, 2003).

Three instances of contamination were detected, which forced the exclusion of those samples from the analysis. Since the negative controls of the experiments were clean, contamination or sample confusion probably occurred at the moment of transferring aliquots of DNA extracts into new tubes. Sample confusion could have also occurred while labelling tubes. Despite a higher number of samples being replicated, only one more case of contamination/sample confusion than the previous chapter is reported. While this finding is consistent with the conclusion that these errors reflect

the level of training of laboratory personnel, it highlights the need for a strict database management system in order to prevent such errors (Paetkau, 2003; Van Rossum et al., 2010; Truong et al., 2011).

3.5 Conclusions

This chapter provides evidence of successful multiplex amplification of six loci originally included in the panel. Multiplex PCR did not have an effect on the reproducibility of results and represents a substantial saving of resources. Alleles from locus Tv53 should be scored with caution. It is recommended that protocols for amplification of this particular locus be re-optimised. Alternatively, its replacement by the inclusion of a locus with equivalent polymorphism should be considered in the future.

Despite the higher incidence of stochastic genotyping error in this experiment than in the previous chapter, genetic profiles generated by multiplex protocols can be considered reliable when amplification and capillary electrophoresis is performed under standardised conditions. However, in the future it will be necessary to find the optimal concentration of template DNA, as high concentrations can possibly increase the incidence of amplification artefacts such as stuttering peaks. Since it is anticipated that differences in size estimates will be observed when slight modifications of the protocol are implemented in different laboratories, it is recommended that all alleles should be sequenced in order to determine their true length. The information generated will serve to build a customised size standard ladder, as recommended by LaHood et al. (2002), that will facilitate standardisation across laboratories.

Although it is encouraging to see that only one more case of contamination was reported in relation to the previous chapter, a call for greater

caution against clerical errors is reiterated, as this still represents a major cause for discrepancies among genotypes. Future studies that use the protocol developed here will require reinforced training of personnel involved and implementing tighter quality control measures. Parallel to the use of automated routines for scoring and binning, independent visual inspection and scoring of alleles should become standard practice.

Chapter 4

Protocol evaluation with low template DNA

4.1 Introduction

Despite a large number of studies focused on the more common use of hair and faeces as non-invasive sources of DNA, saliva is known as a suitable medium from which to collect DNA. DNA retrieved from saliva is used routinely for human clinical studies (Ng et al., 2004). Usage has also extended to the study of population genetics and ecology of several vertebrate groups including mammals (Blejwas et al., 2006), birds (Yannic et al., 2011), reptiles (Schulte et al., 2011), amphibians (Broquet et al., 2007a) and fish (Livia et al., 2006). In genetic studies of wildlife populations, the preferred method of saliva collection is by means of buccal swabs. While in some situations the collection of buccal epithelial cells can be achieved without locating the animal (e.g. by swabbing injuries inflicted by predators, in cases where the animal in question is the predator, Blejwas et al., 2006), mostly epithelial cell collection still requires the capture and physical handling of animals.

With obvious drawbacks to this method of collection, the use of interference devices to collect saliva samples is a sound alternative. A challenging limitation in using saliva from interference devices is that the DNA retrieved can often be of low quantity and/or quality.

In the NIGM literature, low quantity and quality DNA is frequently associated with the occurrence of stochastic genotyping error, affecting different loci to varying degrees (Broquet et al., 2007b; Hoffman & Amos, 2005). The extent to which stochastic genotyping errors are caused by DNA of low quality or low quantity, is not well understood. Investigating the incidence of genotyping error when either DNA quality or quantity is controlled could allows us to identify which variable is of greatest influence. While establishing the causes of error, it is also possible to identify loci that are particularly susceptible to the influence of these factors (i.e. DNA quality or quantity).

This chapter investigates the influence of decreasing template DNA concentration retrieved from a source assumed to yield DNA of optimal quality (e.g. tissue), on the optimised multiplex protocol performance. In a second instance, the effects on the same protocol were assessed using template DNA retrieved from Waxtags, assuming DNA obtained from saliva is of lesser quality, as well as quantity, than that retrieved from tissue. The effect of these variables on the amplification success and occurrence of genotyping error across loci was also investigated. Finally, recommendations on how to improve the use of salivary samples as a substrate to obtain DNA non-invasively are discussed.

4.2 Materials and methods

4.2.1 Data preparation

Dilution of DNA extracted from tissue

Equalised DNA extracts prepared in section 3.2.1 were subjected to a 1:6 dilution in buffer AE (QIAGEN). Dilutions were prepared using a 5–50 μL multi-channelled pipette and a 96 well plate. Diluted extracts were measured twice to provide an approximate range of their final concentration. Assuming the concentration of equalised samples was 27.5 ng/μL, a concentration of approximately 3.43 ng/μL per sample (equivalent to 12.5% of the original extract) was attained. Diluted extracts (n = 29) were separated into sterile 1.5 mL screw cap tubes (Axygen, Inc.), each fitted with a printed label with a unique identifier, and stored at $-20\,^{\circ}$C.

DNA collection with WaxTags

Waxtags were presented to captive possums kept in individual pens at the CWMC. Collectors were given the option to present two tags to the same possum at their discretion. A total of 24 tags were collected. Tags were left overnight and collected the following morning, when they were finally submitted to the molecular ecology laboratory at Lincoln University.

Extraction of DNA followed the "AL" extraction protocol of Vargas et al. (2009), with slight modifications. Waxtags were examined for teeth marks, which were subsequently excised away from the main wax block using a surgical blade. Blades were replaced for each tag to prevent contamination. The resulting pieces of wax containing clear dentition marks were stored in a 15 ml sterile screw-cap tubes (Axygen, Inc.). Fourteen mL of Phosphate buffered saline, pH 7.4 (PBS) was added to each tube and centrifuged at 1500

g for 10 min. The supernatant was extracted and disposed of, along with the pieces of wax. The remaining 200 μL were transferred to a 1.5 μL micro centrifuge tube. 200 μL of buffer AL (QIAGEN) and 20 μL of proteinase K (QIAGEN) were then added to each tube. The mixture was incubated for 24 hours in a dry oven at 56 °C. After incubation, 200 μL of 100% EtOH was added to the mix and vortexed thoroughly. The remainder of the extraction protocol followed the steps described in section 2.2.2, except that samples were eluted twice in 50 μL of buffer AE (QIAGEN), ultimately yielding 100 μL solutions.

Quantification and screening of DNA extracts from Waxtags

DNA concentration of samples obtained from Waxtags was not quantified, since these samples were assumed to contain exogenous DNA (e.g. DNA from bacteria) and possibly other contaminants (e.g. food compounds). Accordingly, samples obtained form Waxtags were only screened for the presence of possum DNA using the procedure described in section 2.2.2. When positive bands were detected, samples were used as a template for microsatellite amplification.

Microsatellite amplification

In order to investigate whether a reduction in the concentration of template DNA affects the quality of genotyping results, the optimised multiplex protocol from chapter 3 was used to amplify extracts diluted in section 4.2.1.

A second trial was also set up to assess whether quantity and quality of template DNA retrieved from Waxtags affected the precision and sensitivity of the protocol. Extracts from Waxtags were amplified in 12 μL reactions containing 4 μL of template DNA, 5 μL of 2x multiplex PCR kit master mix

(final concentration 1x, QIAGEN), 0,5 μL of RNAse free water and 1 μL of forward and reverse primer solution. In order to trial the amplification of very low concentrations of template DNA, the number of PCR cycles was increased as follows: denaturation at 95 °C for 15 min, then 40 cycles of denaturation at 94 °C for 30 sec, annealing temperature (57.8 °C) for 90 sec and extension at 72 °C for 90 sec, followed by a final extension time at 72 °C for 10 min. Amplification of DNA extracts obtained from Waxtags was repeated twice.

4.2.2 Data analysis

Precision of DNA quantification

To account for variability when measuring DNA concentration, the mean, median and the range of all measurements in each trial were estimated. Differences between values were compared using a paired t-test. The null hypothesis tested was that the mean difference between the measurements was not significantly different from zero.

Amplification success and genotyping error of profiles obtained from DNA of reduced concentration

Protocol sensitivity to changes in the concentration of template DNA was estimated by determining amplification success and genotyping error rates. Amplification success, defined as the ratio of the number of positive reactions to the total number of reactions attempted, was estimated per locus. The visualisation of at least one allele at four or more loci was considered a positive amplification. The occurrence of stochastic and systematic genotyping error was quantified using reference genotypes generated in chapter 2 and chapter 3. Reference genotypes were compared to genotypes generated from

diluted samples, allowing the inference of a final consensus genotype. Alleles of consensus genotypes were considered only when observed at least twice among replicates. Since discrepancies between the two reference genotypes were already reported in chapter 3, only discrepancies between genotypes obtained from diluted extracts and reference genotypes were accounted for to maintain a consistent assessment of genotyping error causes, the same categories of genotyping error outlined in section 2.2.6 were taken into account. Mean error rate per locus and mean error rate per multi-locus genotype were also estimated as described in section 2.2.6.

Amplification success and genotyping error of profiles obtained from DNA retrieved from Waxtags

The per locus amplification success was estimated for the profiles obtained from Waxtags as described in the previous section. Genotypes were further categorised as "partially" positive if they exhibited alleles for at least one of four loci. Positive, as well as partially positive, samples were included in the amplification success quantification.

The genotypes of samples obtained from Waxtags were replicated twice only. At this level of replication it was not possible to associate each error with a possible cause. Hence, only the mean error rate per locus and observed error rate per multi-locus genotype were quantified based on the comparison between positive profiles ($n = 10$ pairwise comparisons between positive replicates).

Effects of reduced DNA quantity or quality on amplification success and frequency of error

The per locus ratios of amplification success obtained from both template DNA classes (i.e. diluted DNA and DNA retrieved from Waxtags) were ranked by the length of locus amplicon. Two categories were considered: loci producing amplicons < 200 base pairs and loci producing amplicons ≥ 200 base pairs. The interaction of amplicon length and template DNA class, and its effect on locus amplification success was analysed with a generalised linear model with a binomial error distribution and logit-link function. Binomial error distribution was considered appropriate as the sample size from the different classes of template DNA was unbalanced and non-normally distributed.

Results of mean error rate per locus were categorised according to the quantity and quality of DNA sources. The dataset included the e_l between profiles estimated on the experiments in this chapter and also those from chapter 3. Three classes of template DNA were considered: DNA of good quantity and quality (i.e. DNA extracts as prepared in section 3.2.1), DNA of low quantity and good quality (i.e. diluted DNA as prepared in section 4.2.1), and DNA of low quantity and quality (i.e. DNA retrieved from Waxtags). The effect of template DNA quantity or quality on the e_l was also analysed using a generalised linear model with a binomial error distribution and logit link function.

When generalised linear models indicated significant differences post-hoc, pairwise comparisons of mean values were undertaken using Fisher's restricted LSD test at α=0.05.

All analyses were conducted using the GENSTAT package (v. 14).

4.3 Results

4.3.1 Precision of DNA quantification

The mean difference between measurements of diluted DNA concentrations was 0.18 ng/μL (t = -0.69, P-value = 0.49), thus the H_0 was not rejected. The median and the range differed slightly on two occasions (NS) (median, range: 2.61 ng/μL, 0.36–5.2 ng/μL; 2.62 ng/μL, 0.28–5.7 ng/μL).

4.3.2 Amplification of diluted DNA

Of the 29 diluted extracts, 28 (96,55%) consistently showed positive peaks for 4 or more loci.The only sample failing to show positive peaks for all loci had a DNA concentration of $\approx$ 1.93 ng/μL. Amplification success varied among loci. Locus Tv53 and locus Tv19 were the most affected by reaction failure (Table 4.1).

Thirty one mismatches were observed in 150 pairwise comparisons of single locus genotypes, yielding a mean error rate per locus of 10%. However, as the observed discrepancies were concentrated in just 15 of the samples this increased the observed error rate per multi-locus genotype to 30%. Samples were more affected by allele drop-out (weighted average of allele drop-out rate (p_w) = 10%) than by false alleles (f_w = 1%). Four instances of contamination were detected.

4.3.3 Amplification of DNA retrieved from Waxtags

Of 24 samples, 18 exhibited positive bands for the CO-I barcoding region (74.9%). Ten of the eighteen samples generated positive profiles for microsatellite amplification (55.5%). The remaining eight generated only partially positive genotypes (44.4%).

Table 4.1: Summary of amplification success (*success*), per locus allele dropout rate (p_l) and per locus false allele rate (f_l) for replicates obtained from diluted DNA template. Only error rate per locus (e_l) is presented for replicates obtained from Waxtags

	Diluted DNA				DNA retrieved from Waxtags	
Locus	success (%)	p_l	f_l	e_l	success (%)	e_l
Tv19	86.21	0.09	0.00	0.12	77.78	0.25
Tv27	93.10	0.13	0.00	0.10	88.89	0.25
Tv53	82.76	0.16	0.03	0.20	44.44	0.25
Tv58	96.55	0.38	0.00	0.08	94.44	0.05
Tv5.64	93.10	0.00	0.00	0.04	83.33	0.2
TvM1	93.10	0.11	0.00	0.08	50.00	0.25

Low amplification success affected loci Tv19, Tv53 and TvM1 while loci Tv27, Tv58 and Tv5.64 showed higher and more consistent amplification (Table 4.1). The per locus error rate was high for all loci except locus Tv58 and Tv5.64 (Table 4.1), yielding a mean error rate per locus of 21% and a observed error rate per multi-locus genotype of 45%.

4.3.4 Effects of locus and template DNA characteristics on amplification success and frequency of error

For each of the two template DNA classes, amplification success of alleles $\geq$ 200 base pairs was lower than alleles $<$ 200 base pairs (Figure 2.1). However, the *post-hoc* comparison of means indicated that difference between amplicon length and template DNA class was only significant for samples retrieved from Waxtags and amplifying fragments $\geq$ 200 base pairs ($\chi^2 = 31.9$, $d.f. = 3$, P-value < 0.001) (Figure 2.1). Amplification success of am-

plicons extracted from Waxtag DNA was statistically lower than that of diluted DNA.

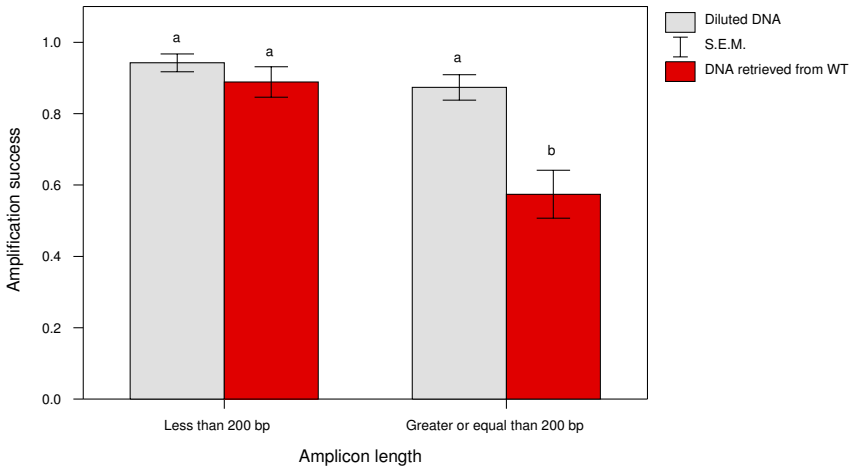

Figure 4.1: Decrease in amplification success due to the interaction between amplicon size and quantity and quality of template DNA used to generate profiles for individual possums. Letter 'b' indicates significant differences between means according to Fisher's LSD test ($\alpha = 0.05$)

DNA retrieved from Waxtags also demonstrated a significant increase in the proportion of mean error rate per locus compared to diluted DNA ($\chi^2 = 16.26$, $d.f. = 2$, P-value < 0.001) (Figure 4.2).

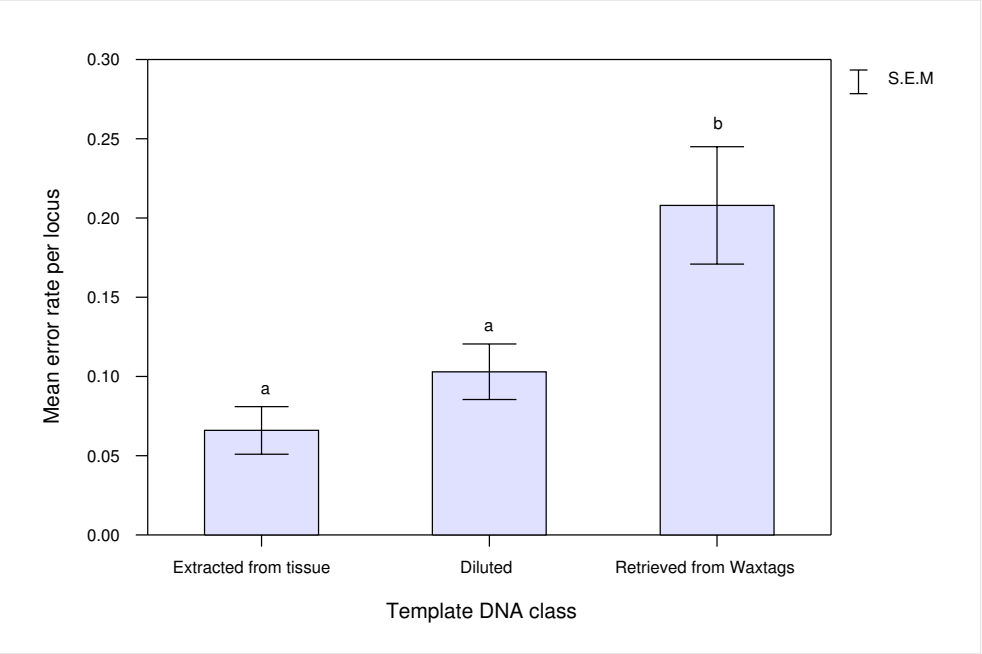

Figure 4.2: Increase in mean error rate per locus according to the quantity and quality of template DNA used to generate profiles of individual possums. Letter 'b' indicates significant differences between means according to Fisher's LSD test ($\alpha = 0.05$)

4.4 Discussion

Despite substantial optimisation of the microsatellite amplification protocol, it was not possible to achieve individual identification of possums using DNA retrieved from Waxtags. Results presented in this chapter suggest that, rather than a technical limitation of the microsatellite amplification system, there is great variability in the ability of Waxtags to provide DNA of sufficient quantity and/or quality to achieve positive and reliable amplification. Whilst the concentration of template DNA retrieved from Waxtags was not quantified (see section 4.2.1), measurements of DNA concentration from a previous batch of extracts — excluded for the present trials due to evidence of contagion — indicated that Waxtags typically retrieve low quantities of DNA (median, range: 0.93 ng/μL, 0.1–9.22 ng/μL). While it is clearly possible to recover DNA from the saliva of possums, new strategies to maximise DNA yield from this source material, as well as to reduce its exposure to the environment, need to be explored when deciding upon the adoption of the microsatellite amplification system developed throughout this thesis.

The high frequency of error and the low amplification success of larger amplicons associated with DNA from Waxtags (Figure 4.1 and 4.2) suggests that quality of DNA (i.e. DNA degraded by exposure to the environment or contaminated with PCR inhibitors), rather than quantity, is the limiting factor in the sensitivity and reliability of the protocol developed in this thesis. Given that most of the microsatellite loci used in this study are dinucleotides, these results are partially consistent with the conclusions of Broquet et al. (2007b). The authors of this study demonstrated that microsatellite loci with large amplicons and longer repeat motifs exacerbate the already deleterious effects of low quantity and quality DNA (e.g. retrieved

from faeces or hair) on amplification success and the occurrence of stochastic genotyping error. Broquet et al. (2007b) did not discard, however, the occurrence of additional case specific factors interfering with amplification and genotyping error. Results found here suggest that alongside locus characteristics, the presence of PCR inhibitors or DNA degradation of salivary samples might be affecting the sensitivity and reliability of the microsatellite amplification system developed throughout this work.

DNA degradation is a common occurrence of non-invasively collected samples, such as that collected from Waxtags. Exogenous endonuclease activities, as well as DNA hydrolitic and oxidative decomposition, start rapidly as cellular metabolic processes decline (Lindahl, 1993). Because there are many pathways in which DNA can be degraded, the rate of degradation appears to be sample specific, although a possible pattern can be detected: copy number decreases as fragment size increases (Deagle et al., 2006). In that respect, salivary samples collected with Waxtags might contain a higher proportion of smaller fragments of DNA which explains why the rate of amplification success of loci with smaller amplicons is higher. As Waxtags are left in the environment for less than 26 hours, it could be argued that DNA degradation by environmental exposure is not the primary factor in this case. It is possible that other factors, such as the presence of PCR inhibitors in the samples, are interfering with the amplification of DNA retrieved from Waxtags.

A variety of PCR inhibitors have been reported in the literature, ranging from endogenous substances, such as contaminants found in forensic samples (e.g. haemoglobin), to exogenous agents including food constituents (e.g. organic phenolic compounds) and contaminants found in the environment (e.g. heavy metals or bacterial cells) (Wilson, 1997; Alaeddini, 2012). Most

relevant to salivary samples is the presence of bacterial and food traces. For instance, the amount of bacterial DNA present in salivary samples has been associated with decreased reliability of human genotyping assays (Herraez & Stoneking, 2008). The high abundance of bacteria and presence of inhibitory compounds on dental surfaces (Parrish & Greenberg, 1995) indicates the possibility that possum salivary samples retrieved from Waxtags could also have a high amount of bacteria and PCR inhibitors.

While the rate of genotyping error in profiles obtained from diluted samples (Table 4.1) was not significantly higher in relation to the genotyping error found in the previous chapter (i.e. samples of good quality), the increase of the per locus allele drop-out rate is apparent. Locus Tv58 in particular, suffers from a high rate of this type of error apparently caused by the preferential amplification of allele 132 over allele 124. The higher rate of allele drop-out is intriguing because it is contrary to the hypothesis that smaller fragments have higher amplification success (which is largely supported by this chapter's own results). One of the suggested mechanisms for allele drop-out is that a lower content of CG at the flanking sequence favours the denaturation of one of the alleles over the other, and as a result is preferentially amplified (Walsh et al., 1992).

The difficulty in associating cases of allele drop-out or false alleles to possible causes that fit all cases precludes the formulation of solid conclusions on how to avoid such errors. However, in practical terms, it seems there is a threshold of optimal template DNA concentration above which stuttering is exacerbated thus increasing the rate of false allele occurrence, and below which allele drop-out becomes more frequent. The existence of an optimal template DNA concentration range for microsatellite amplification systems is consistent with what has been observed in model organisms such

as humans, a fact that was highlighted earlier in this thesis (see section 3.4).

The inability to quantify the rates of false alleles and allele drop-out in profiles generated from salivary samples is admittedly a weak point of this chapter. Time constraints and small sample sizes precluded the estimation of these important parameters, which are critical for a complete assessment of the suitability of salivary samples for genotyping assays. Such parameters allow simulation of the impact of stochastic genotyping error on intended downstream analyses (e.g. population abundance estimation). The classic multi-tube approach developed by Taberlet et al. (1996) is the most widely adopted procedure to estimate the rates of allele drop-out and false alleles in non-invasive samples. The multi-tube method requires the replication of PCRs up to seven times for every sample in order to obtain a consensus genotype. Despite its widespread adoption, this approach is time consuming and expensive. Such drawbacks have lead to the proposal of slightly more efficient variations of this method such as the comparative multi-tube approach (Frantz et al., 2003) and reference genotypes approach (Adams & Waits, 2007). While these methods have contributed to reducing the number of replicate PCRs, a considerable number of replicates are still needed under this framework. As an alternative, recently developed methods incorporate statistical approaches to estimate the magnitude of genotyping error (Johnson & Haydon, 2007; Miller et al., 2002). Statistical estimates could easily be tested as these require only a duplicated set of PCRs.

Currently there are several methods to quantify DNA damage and the presence of PCR inhibitors (Deagle et al., 2006; King et al., 2009). Once inhibitors are detected, the use of appropriate additives that facilitate amplification of contaminated samples (e.g. BSA) could be tested. Similarly, the use of whole genome amplification can be tested to increase the copy

number of low quantity and degraded template DNA (Konakandla et al., 2006). It must be noted that the cost of implementing these techniques needs to be judged against the cost of developing a better strategy to collect DNA, as discussed below.

4.4.1 Recommendations on the collection of salivary samples

The evidence gathered in this chapter indicates that DNA retrieved from Waxtags is of variable low quality, yet saliva remains an attractive medium from which to isolate possum DNA. The majority of studies reporting the use of saliva as a non-invasive DNA source have collected samples by buccal swabbing. The surface of the collection material (e.g. the cotton swab) on which saliva is gathered can play a critical role in preserving DNA quantity and integrity. Novel methods developed for collecting DNA samples for human clinical studies should also be considered for enhancing collection of possum DNA. The Oragene technique (OraSure Technologies, Inc.) is based on the prolonged soaking of a sponge in saliva where the sponge is placed within the cheek pouch (i.e. the space between the cheek and the gums) of the subject for approximately 15 minutes. The Oragene collection method has been subsequently trialled in the domestic dog and demonstrated a 28-fold increase in DNA yield compared to cotton swabbing (Mitsouras & Faulhaber, 2009). Ng et al. (2004) report that a minimum of 2 mL of human saliva is necessary to retain DNA of the required quality to amplify a fragment of up to 581 bp. Furthermore, it is possible to delay the amplification of samples for up to one month in sub-optimal conditions and still get a positive result. While the time required for an optimal collection of saliva with Oragene method seems impractical to achieve with free ranging animals, the sponge collection methods could be considered and adapted to

collect salivary samples from possums. Another alternative is the use of FTA cards, which have been successfully trialled as low cost substrates to store salivary samples containing DNA (e.g. Livia et al., 2006). It is clear from this discussion that the way interference devices collect DNA can be improved.

4.5 Conclusions

A reduction in the concentration of template DNA isolated from tissue did not cause a reduction in the performance of the optimised microsatellite amplification system developed here. While statistical tests did not detect a significant reduction in the amplification success or an increase in the incidence of genotyping error, there was an increase in the per locus incidence of allele drop-out and a reduction in the occurrence of false alleles. This suggests the existence of an optimal template DNA concentration threshold that balances the occurrence of these classes of error. In future studies, it would be desirable to explore the existence of this threshold in order to mitigate the occurrence of stochastic genotyping error in non-invasively collected samples.

Waxtags allow non-invasive collection and isolation of DNA from possums. Although this DNA is suitable for microstellite amplification in some cases, low quantity retrieved, DNA degradation due to exposure to the environment and the possible presence of PCR inhibitors make DNA isolated from Waxtags of overall unsuitable quality for genotyping essays. New strategies to maximise DNA yield from saliva collected with interference devices, as well as to reduce its exposure to the environment, need to be explored in order to fully exploit the capabilities of the microsatellite amplification system developed in this thesis (see section 4.4.1).

It is necessary to quantify the occurrence of the major classes of stochastic genotyping error (i.e. allele drop-out and false alleles) in profiles obtained from DNA isolated from saliva. Quantification will help to further assess suitability of saliva as a substrate for non-invasive collection of possum DNA. Parallel to this effort, it is necessary to quantify and characterise the rate of degradation and presence of PCR inhibitors in order to adopt adequate techniques to overcome these problems.

Chapter 5

General discussion

The development of novel methodologies that allow accurate monitoring of shifts in population size provides new tools to manage and eradicate pests. The combination of DNA genotyping technology with the ability to retrieve DNA non-invasively promises to overcome the limitations of traditional monitoring methods, in particular when targeted species are elusive. Results of this thesis indicate that it is possible to implement an efficient, reproducible, and reliable microsatellite amplification system capable of identifying individual possums. However, the sensitivity and reliability of the methodology are compromised when template DNA is not of sufficient quantity and quality. The reduced amplification success and increased incidence of genotyping errors observed here could be related to factors such as the presence of exogenous DNA, PCR inhibitors, DNA degradation due to exposure to the environment, or the inability of interference devices to collect enough DNA to override the deleterious effects of the previous factors.

The patterns of genetic diversity observed are in agreement with Taylor et al. 2004, who concluded that the introduction of possums to the South Island was undertaken repeatedly and supplemented by subsequent intro-

ductions of mixed stocks. While the same study established that possums in the South Island might descend predominantly from Tasmanian lineages, there is anecdotal evidence of the introduction of stocks from mainland Australia, at least in the Otago region. Genetic admixture would account for the inconsistency of heterozygote deficits observed here, as well as the observations of bimodal distribution of alleles at locus Tv5.64. Population admixture caused by behavioural changes after frequent control operations is also likely. Possums in their native range and in captivity exhibit territorial and social behaviour that is compatible with assortative mating (Stow et al., 2006; Clinchy et al., 2004). If this is assumed to be the default behaviour, possum populations would eventually show signs of structuring and inbreeding. However, previous studies on the mating system of possums in New Zealand suggest reproduction occurs randomly (Taylor et al., 2000; Sarre et al., 2000), and there is home range and mobility expansion as a result of population control (Blackie et al., 2011; Ji et al., 2001; Sweetapple & Nugent, 2009). While these studies indicate that introduced possums have a substantial level of behavioural plasticity, it is unlikely that behaviour alone counters the effects of isolation by distance. Therefore, the patterns observed here must pre-date the adoption of control operations and might be the result of the manner in which current populations were founded.

Moderate levels of genetic diversity confirm that a panel of six microsatellite loci has sufficient statistical power to allow identification of individual possums. The PI_{sib} estimated in this study guarantees that at least 250 possums can be identified by their multi-locus genotype even if some of these are full siblings. Assuming that typical possum density can be reduced to 1 possum/km^2 after aerial control (Sweetapple & Nugent, 2009), the present method is sufficient for monitoring an extensive area (approximately 25000

ha). The present protocol also allows monitoring of possum populations across New Zealand's main islands, considering the greater genetic diversity known in populations of the North Island (Taylor et al., 2004). However, despite demonstrating the statistical reliability of the present panel, it is anticipated that the inclusion of additional loci will be required in order to allow gender identification and extend the monitoring capacity to areas with very high possum densities.

Researchers of non-invasive genetic monitoring have acknowledged the importance of detecting and monitoring the occurrence of stochastic genotyping error (Pompanon et al., 2005). The monitoring scheme developed here was expanded to include categories of error that seldom receive attention, such as mobility shifts, scoring errors originated by automated detection routines, contamination, and human error (Amos et al., 2007; Paetkau, 2003; Pasqualotto et al., 2007; Haberl & Tautz, 1999; Ellis et al., 2011). It was also considered important to include DNA extracted from tissue in the analysis, supporting the idea that errors can occur regardless of the perceived quality of the samples. The occurrence of non-negligible rates of systematic error caused by changes in protocol (mobility shifts), software (failure to recognise peaks, size shifts), and human error (subjectivity, contamination, sample swapping) illustrates the potentially high impact these errors can have in the quality of genotyping process. Reporting this data is important to facilitate comparison with future studies, and to stress the need for caution, thorough training, and experience in reliable procedural implementation.

The incidence of stochastic and systematic genotyping error remained relatively constant despite changes during protocol development. Significant increases in e_l and e_{obs} were observed only when the concentration and

quality of template DNA was reduced. These observations were expected and suggest that the incidence of genotyping error cannot be attributed to modifications in the amplification protocol (e.g. multiplexing). Although the rates of genotyping error observed for tissue samples were unexpectedly high, the rates fell into the range reported by similar studies (Soulsbury et al., 2007; Hoffman & Amos, 2005; Bonin et al., 2004). Similarly, while an increase in the occurrence of false alleles with the multiplex protocol was observed, it is suspected the increase was due to the high concentration of template DNA used. The low incidence of false alleles when the template DNA concentration was reduced supports this interpretation. Nonetheless, since the panel developed here was constituted mainly of microsatellites with dinucleotide repeats, it is possible that the occurrence of false alleles is related to inherent characteristics of these loci (Broquet et al., 2007b). Altogether these observations suggest an optimal concentration of template DNA is required to mitigate the occurrence of stochastic genotyping error. We observed that high template DNA concentrations can increase the occurrence of false alleles, while a decrease in DNA quantity can lead to the increase in the incidence of allelic drop-out. The existence of an optimal DNA concentration threshold for microsatellite amplification systems has been confirmed in studies of both wildlife and humans (MacDonald et al., 2011; Nicklas & Buel, 2003; Kline et al., 2005).

Waxtags were incapable of providing DNA of sufficient quantity and quality to allow reliable genotyping. Interference devices such as Waxtags have a proven capacity to collect DNA (Vargas et al., 2009) but the less than expected amplification success suggests Waxtags are unsuitable for collecting DNA of sufficient quality possibly due to the deleterious effects of environmental exposure. Studies that have used saliva obtained from buccal

swabbing of free ranging animals have obtained concentrations of DNA comparable to those of tissue samples (Yannic et al., 2011; Prunier et al., 2012; Martin-Galvez et al., 2011; Broquet et al., 2007a). Our amplification success from Waxtags saliva was similar to what has been observed for stool and hair samples (e.g. Frantz et al. 2003; Valiere et al. 2007; Gleeson et al. 2003), possibly explained by our sampling methodology being limited to one season (autumn) and one interference device per individual. Suboptimal preservation and extraction methods, individual variation, and high environmental moisture could have exacerbated the degradation of DNA present in the samples (Piggott, 2004).

DNA quantity is also an issue. From a separate batch of samples collected in the field (data not included) it was estimated Waxtags retrieve very low quantities of DNA (median, range: 0.93 ng/μL, 0.1–9.22 ng/μL). Despite this concentration being similar to what is typically recovered from hair or stool samples, the quantification of DNA by ultraviolet spectroscopy at these concentrations can be inaccurate (Yannic et al., 2011). Therefore the implementation of a more sensitive screening method (e.g. qPCR) is recommended to establish the actual concentration of DNA that Waxtags are capable of retrieving (Morin et al., 2010). Regardless of the quantification method, it is clear that the low concentration of DNA added to the presence of PCR inhibitors, and/or high rates of exogenous DNA (e.g. bacterial DNA) may have inhibited amplification reactions.

Increasing the sampling frequency of each individual might overcome the reduced capacity of interference devices to collect sufficient DNA. Non-invasive genetic studies focus on collecting several specimens per individual in order to increase capture probabilities and amplification rates (Gleeson et al., 2003; Piggott et al., 2008; Marks et al., 2009). In spite of an enhanced

sampling strategy, interference devices could be affected by contagion or negative interactions with non-target species (see Introduction). Consequently, the design of an improved saliva collection device that allows interaction with only a single individual and maximises the collection of buccal epithelial cells is recommended.

5.1 General conclusion

The robustness, efficiency, and reproducibility of the protocol developed here is encouraging, yet this pilot study is still incomplete. Different collection and preservation strategies, the number of replicate PCRs required to obtain a reliable multi locus genotype, or the impacts of the estimated error rates on demographic estimation have not yet been addressed. Since implementation of the multiplex protocol was estimated to reduce costs by 92.3%, replicating PCRs would not impose a substantial increase in genotyping costs. However, analytical approaches such as the maximum likelihood method for error estimation (Johnson & Haydon, 2007) can avoid the need for extensive replication and should be taken into account for future studies. Despite the outlined issues demand attention, it is first necessary to improve the current salivary DNA collection method to allow the protocol to become a suitable alternative or supplement to field based monitoring methodologies. It is also clear from this discussion that a lack of understanding of the geographic structure of possum populations hinders the ability to control and eradicate them. In response to this need, the present method, achieved by non-invasive or invasive sampling, can contribute to elucidate aspects of the reproductive and movement behaviour of possums in New Zealand.

Bibliography

Adams, J. R., & Waits, L. P. (2007). An efficient method for screening faecal DNA genotypes and detecting new individuals and hybrids in the red wolf (*Canis rufus*) experimental population area. *Conservation Genetics*, *8*(1), 123–131.

Alaeddini, R. (2012). Forensic implications of PCR inhibition- a review. *Forensic Science International Genetics*, *6*(3), 297–305.

Allendorf, F. W., & Luikart, G. (2007). *Conservation and the genetics of populations*, vol. 8. Malden, MA: Blackwell Publishing.

Amos, W., Hoffman, J. I., Frodsham, A., Zhang, L., Best, S., & Hill, A. V. S. (2007). Automated binning of microsatellite alleles: problems and solutions. *Molecular Ecology Notes*, *7*(1), 10–14.

Applied Biosystems (2005). *GenneMapper ® software version 4.0, reference and troubleshooting guide*. Applied Biosystems.

Arandjelovic, M., Guschanski, K., Schubert, G., Harris, T. R., Thalmann, O., Siedel, H., & Vigilant, L. (2009). Two-step multiplex polymerase chain reaction improves the speed and accuracy of genotyping using DNA from noninvasive and museum samples. *Molecular Ecology Resources*, *9*(1), 28–36.

Balloux, F., & Lugon-Moulin, N. (2002). The estimation of population differentiation with microsatellite markers. *Molecular Ecology*, *11*(2), 155–165.

Banks, S., Hoyle, S., Horsup, A., Sunnucks, P., & Taylor, A. (2003). Demographic monitoring of an entire species (the northern hairy-nosed wombat, *Lasiorhinus krefftii*) by genetic analysis of non-invasively collected material. *Animal Conservation*, *6*(Part 2), 101–107.

Banks, S. C., Dubach, J., Viggers, K. L., & Lindenmayer, D. B. (2010). Adult survival and microsatellite diversity in possums: effects of major histocompatibility complex-linked microsatellite diversity but not multilocus inbreeding estimators. *Oecologia*, *162*(2), 359–370.

Banks, S. C., Knight, E. J., Dubach, J. E., & Lindenmayer, D. B. (2008). Microhabitat heterogeneity influences offspring sex allocation and spatial kin structure in possums. *Journal of Animal Ecology*, *77*(6), 1250–1256.

Beebee, T., & Rowe, G. (2008). *An introduction to molecular ecology*. Oxford University Press, second ed.

Beja-Pereira, A., Oliveira, R., Alves, P. C., Schwartz, M. K., & Luikart, G. (2009). Advancing ecological understandings through technological transformations in non-invasive genetics. *Molecular Ecology Resources*, *9*(5), 1279–1301.

Benschop, C. C., van der Beek, C. P., Meiland, H. C., van Gorp, A. G., Westen, A. A., & Sijen, T. (2011). Low template STR typing: effect of replicate number and consensus method on genotyping reliability and DNA database search results. *Forensic Science International Genetics*, *5*(4), 316 – 328.

Blackie, H. M., Russell, J. C., & Clout, M. N. (2011). Maternal influence on philopatry and space use by juvenile brushtail possums (*Trichosurus vulpecula*). *Journal of Animal Ecology*, *80*(2), 477–483.

Blejwas, K. M., Williams, C. L., Shin, G. T., McCullough, D. R., & Jaeger, M. M. (2006). Salivary DNA evidence convicts breeding male coyotes of killing sheep. *Journal of Wildlife Management*, *70*(4), 1087–1093.

Bonin, A., Bellemain, E., Eidesen, P. B., Pompanon, F., Brochmann, C., & Taberlet, P. (2004). How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, *13*(11), 3261–3273.

Borchers, D. L., Buckland, S. T., & Zucchini, W. (2002). *Estimating animal abundance: closed populations*. Springer Press.

Broquet, T., Berset-Braendli, L., Emaresi, G., & Fumagalli, L. (2007a). Buccal swabs allow efficient and reliable microsatellite genotyping in amphibians. *Conservation Genetics*, *8*(2), 509–511.

Broquet, T., Menard, N., & Petit, E. (2007b). Non-invasive population genetics: a review of sample source, diet, fragment length and microsatellite motif effects on amplification success and genotyping error rates. *Conservation Genetics*, *8*(1), 249–260.

Broquet, T., & Petit, E. (2004). Quantifying genotyping errors in noninvasive population genetics. *Molecular Ecology*, *13*(11), 3601–3608.

Brown, J. A., McAuliffe, R. J., & Hicling, G. J. (2004). A new trap-catch index of patchiness in animal distribution. *Wildlife Research*, *31*(4), 433–436.

Buschiazzo, E., & Gemmell, N. J. (2006). The rise, fall and renaissance of microsatellites in eukaryotic genomes. *Bioessays*, *28*(10), 1040–1050.

Butler, J., Ruitberg, C., & Vallone, P. (2001). Capillary electrophoresis as a tool for optimization of multiplex PCR reactions. *Fresenius Journal of Analytical Chemistry*, *369*(3-4), 200–205.

Butler, J. M. (2005). Constructing STR multiplex assays. *Methods in Molecular Biology*, *297*, 53–66.

Chamberlain, J., Gibbs, R., Ranier, J., Nguyen, P., & Caskey, C. (1988). Deletion screening of the Duchenne muscular-dystrophy locus via multiplex DNA amplification. *Nucleic Acids Research*, *16*(23), 11141–11156.

Clinchy, M., Taylor, A., Zanette, L., Krebs, C., & Jarman, P. (2004). Body size, age and paternity in common brushtail possums (*Trichosurus vulpecula*). *Molecular Ecology*, *13*(1), 195–202.

Cowan, P., & Clout, M. (2000). *The brushtail possum: biology, impact and management of an introduced marsupial*, chap. Possums on the move: activity patterns, home ranges and dispersal, (pp. 24–34). Manaaki Whenua Press.

Cowan, P., Sarre, S., & Aitken, N. (2007). The genetics of possum dispersal. Kararehe Kino Vertebrate Pest Research 10, Landcare Research.

Cowan, P., Sarre, S., & Chapman, J. (2002). Mapping possum dispersal across buffers with genetic markers. Tech. rep., Landcare Research, Palmerston North, NZ.

Cowan, P. E. (1998). *The handbook of New Zealand mammals*, chap. Species accounts: brushtail possum, (pp. 68–98). Oxford University Press.

Curtis, C., Stewart, B. S., & Karl, S. A. (2009). Pleistocene population expansions of Antarctic seals. *Molecular Ecology*, *18*(10), 2112–2121.

Day, T., O'Connor, C., & Mattews, L. (2000). *The brushtail possum: biology, impact and management of an introduced marsupial*, chap. Possum social behaviour, (pp. 35–46). Manaaki Whenua Press.

De Barba, M., & Waits, L. P. (2010). Multiplex pre-amplification for non-invasive genetic sampling: is the extra effort worth it? *Molecular Ecology Resources*, *10*(4), 659–665.

Deagle, B. E., Eveson, J. P., & Jarman, S. N. (2006). Quantification of damage in DNA recovered from highly degraded samples–a case study on DNA in faeces. *Frontiers in Zoology*, *3*(11), 10.

DeWoody, J., Nason, J. D., & Hipkins, V. D. (2006). Mitigating scoring errors in microsatellite data from wild populations. *Molecular Ecology Notes*, *6*(4), 951–957.

DeYoung, R. W., & Honeycutt, R. L. (2005). The molecular toolbox: genetic techniques in wildlife ecology and management. *Journal of Wildlife Management*, *69*(4), 1362–1384.

Ellis, J. S., Gilbey, J., Armstrong, A., Balstad, T., Cauwelier, E., Cherbonnel, C., Consuegra, S., Coughlan, J., Cross, T. F., Crozier, W., Dillane, E., Ensing, D., Garcia de Leaniz, C., Garcia-Vazquez, E., Griffiths, A. M., Hindar, K., Hjorleifsdottir, S., Knox, D., Machado-Schiaffino, G., McGinnity, P., Meldrup, D., Nielsen, E. E., Olafsson, K., Primmer, C. R., Prodohl, P., Stradmeyer, L., Vaha, J. P., Verspoor, E., Wennevik, V., & Stevens, J. R. (2011). Microsatellite standardization and evaluation of genotyping error in a large multi-partner research programme for conservation of Atlantic salmon (*Salmo salar L.*). *Genetica*, *139*(3), 353–367.

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567.

Fazekas, A. J., Steeves, R., & Newmaster, S. G. (2010). Improving sequencing quality from PCR products containing long mononucleotide repeats. *Biotechniques*, *48*(4), 277–281.

Fernando, P., Vidya, T., Rajapakse, C., Dangolla, A., & Melnick, D. (2003). Reliable non-invasive genotyping: fantasy or reality? *Journal of Heredity*, *94*(2), 115–123.

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol*, *3*(5), 294–299.

Forsyth, D. M., Link, W. A., Webster, R., Nugent, G., & Warburton, B. (2005). Nonlinearity and seasonal bias in an index of brushtail possum abundance. *Journal of Wildlife Management*, *69*(3), 976–984.

Fox, J. (2005). The R commander: A basic-statistics graphical user interface to R. *Journal of Statistical Software*, *14*(9), 1–42.

Frantz, A. C., Pope, L. C., Carpenter, P. J., Roper, T. J., Wilson, G. J., Delahay, R. J., & Burke, T. (2003). Reliable microsatellite genotyping of the Eurasian badger (*Meles meles*) using faecal DNA. *Molecular Ecology*, *12*(6), 1649–1661.

Freeland, J., Kirk, H., & Petersen, S. D. (2011). *Molecular ecology*. Wiley-Blackwell, second edition ed.

Garvin, M. R., Saitoh, K., & Gharrett, A. J. (2010). Application of single nucleotide polymorphisms to non-model species: a technical review. *Molecular Ecology Resources*, *10*(6), 915–934.

Gervasi, V., Ciucci, P., Davoli, F., Boulanger, J., Boitani, L., & Randi, E. (2010). Addressing challenges in non-invasive capture-recapture based estimates of small populations: a pilot study on the Apennine brown bear. *Conservation Genetics*, *11*(6), 2299–2310.

Gleeson, D., Byrom, A., Howitt, R., & Nugent, G. (2003). Monitoring vertebrate pests by using their DNA. Tech. Rep. 2, Kararehe Kino Vertebrate Pest Research.

Gleeson, D. M., Byrom, A. E., & Howitt, R. L. J. (2010). Non-invasive methods for genotyping of stoats (*Mustela erminea*) in New Zealand: potential for field applications. *New Zealand Journal of Ecology*, *34*(3), 356–359.

Goudet, J., Raymond, M., deMeeus, T., & Rousset, F. (1996). Testing differentiation in diploid populations. *Genetics*, *144*(4), 1933–1940.

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Leger, P., Lepais, O., Lepoittevin, C., Malausa, T., Revardel, E., Salin, F., & Petit, R. J. (2011a). Current trends in microsatellite genotyping. *Molecular Ecology Resources*, *11*(4), 591–611.

Guichoux, E., Lagache, L., Wagner, S., Leger, P., & Petit, R. J. (2011b). Two highly validated multiplexes (12-plex and 8-plex) for species delimitation and parentage analysis in oaks (*Quercus* spp.). *Molecular Ecology Resources*, *11*(3), 578–585.

Guo, S., & Thompson, E. (1992). Performing the exact test of Hardy-Weinberg proportion for multiples alleles. *Biometrics*, *48*(2), 361–372.

Haberl, M., & Tautz, D. (1999). Comparative allele sizing can produce inaccurate allele size differences for microsatellites. *Molecular Ecology*, *8*(8), 1347–1349.

Hajibabaei, M., Janzen, D. H., Burns, J. M., Hallwachs, W., & Hebert, P. D. N. (2006). DNA barcodes distinguish species of tropical Lepidoptera. *Proc Natl Acad Sci U S A*, *103*(4), 968–971.
URL `http://dx.doi.org/10.1073/pnas.0510466103`

Hamilton, M. B. (2009). *Population genetics*. Chichester: Wiley-Blackwell.

Hartl, D. L., & Clark, A. G. (2007). *Principles of population genetics*. Sunderland, Mass. : Sinauer Associates, 4th edition ed.

Herraez, D. L., & Stoneking, M. (2008). High fractions of exogenous DNA in human buccal samples reduce the quality of large-scale genotyping. *Analytical Biochemistry*, *383*(2), 329–331.

Hoffman, J. I., & Amos, W. (2005). Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology*, *14*(2), 599–612.

Holleley, C. E., & Geerts, P. G. (2009). Multiplex Manager 1.0: a cross-platform computer program that plans and optimizes multiplex PCR. *Biotechniques*, *46*(7), 511.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.

Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F(ST). *Nature Reviews Genetics*, *10*(9), 639–650.

Idury, R., & Cardon, L. (1997). A simple method for automated allele binning in microsatellite markers. *Genome Research*, *7*(11), 1104–1109.

Ji, W., Sarre, S., Aitken, N., Hankin, R., & Clout, M. (2001). Sex-biased dispersal and a density-independent mating system in the Australian brushtail possum, as revealed by minisatellite DNA profiling. *MOLECULAR ECOLOGY*, *10*(6), 1527–1537.

Johnson, P. C. D., & Haydon, D. T. (2007). Maximum-likelihood estimation of allelic dropout and false allele error rates from microsatellite genotypes in the absence of reference data. *Genetics*, *175*(2), 827–842.

Karaiskou, N., & Primmer, C. (2008). PCR multiplexing for maximising genetic analyses with limited DNA samples: an example in the collared flycatcher, *Ficedula albicollis*. *Annales Zoologici Fennici*, *45*(6), 478–482.

King, C. E., Debruyne, R., Kuch, M., Schwarz, C., & Poinar, H. N. (2009). A quantitative approach to detect and overcome PCR inhibition in ancient DNA extracts. *Biotechniques*, *47*(5), 941–949.

Kirkpatrick, S., Gelatt, C., & Vecchi, M. (1983). Optimisation by simulated annealing. *Science*, *220*(4598), 671–680.

Kline, M., Duewer, D., Redman, J., & Butler, J. (2005). Results from the NIST 2004 DNA quantitation study. *Journal of Forensic Sciences*, *50*(3), 571–578. NIJ/DNA Grantees Meeting, Washington, DC, JUN 30, 2004.

Konakandla, B., Park, Y., & Margolies, D. (2006). Whole genome amplification of chelex-extracted DNA from a single mite: a method for studying

genetics of the predatory mite *Phytoseiulus persimilis*. *Experimental and Applied Acarology*, *40*(3-4), 241–247.

LaHood, E., Moran, P., Olsen, J., Grant, W., & Park, L. (2002). Microsatellite allele ladders in two species of Pacific salmon: preparation and field-test results. *Molecular Ecology Notes*, *2*(2), 187–190.

Lam, M. K. P., Hickson, R. E., Cowan, P. E., & Cooper, D. W. (2000). A major histocompatibility (MHC) microsatellite locus in brushtail possums (*Trichosurus vulpecula*). *Online Journal of Veterinary Research*, *4*, 139–141.

Lepais, O., & Bacles, C. F. E. (2011). Comparison of random and SSR-enriched shotgun pyrosequencing for microsatellite discovery and single multiplex PCR optimization in *Acacia harpophylla F. Muell. Ex Benth*. *Molecular Ecology Resources*, *11*(4), 711–724.

Lindahl, T. (1993). Instability and decay of the primar structure of DNA. *Nature*, *362*(6422), 709–715.

Livia, L., Antonella, P., Hovirag, L., Mauro, N., & Panara, F. (2006). A non-destructive, rapid, reliable and inexpensive method to sample, store and extract high-quality DNA from fish body mucus and buccal cells. *Molecular Ecology Notes*, *6*(1), 257–260.

Luikart, G., Pilgrim, K., Visty, J., Ezenwa, V. O., & Schwartz, M. K. (2008a). Candidate gene microsatellite variation is associated with parasitism in wild bighorn sheep. *Biology Letters*, *4*(2), 228–231.

Luikart, G., Ryman, N., Tallmon, D. A., Schwartz, M. K., & Allendorf, F. W. (2010). Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. *Conservation Genetics*, *11*(2), 355–373.

Luikart, G., Zundel, S., Rioux, D., Miquel, C., Keating, K. A., Hogg, J. T., Steele, B., Foresman, K., & Taberlet, P. (2008b). Low genotyping error rates and non-invasive sampling in bighorn sheep. *Journal of Wildlife Management*, *72*(1), 299–304.

MacDonald, A. J., Sarre, S. D., FitzSimmons, N. N., & Aitken, N. (2011). Determining microsatellite genotyping reliability and mutation detection ability: an approach using small-pool PCR from sperm DNA. *Molecular Genetics and Genomics*, *285*(1), 1–18.

MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L., & Hines, J. E. (2006). *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence*. Academic Press.

Mahood, M. L., Cruickshank, R. H., Ross, J. G., Holyoake, A. J., Ogilvie, S. C., & Paterson, A. M. (unpublished). Genotyping of brush-tail possum (*Trichosurus vulpecula*) using non-invasive recovered DNA from bait interference devices.

Marks, C. A., Gigliotti, F., McPhee, S., Piggott, M. P., Taylor, A., & Glen, A. S. (2009). DNA genotypes reveal red fox (*Vulpes vulpes*) abundance, response to lethal control and limitations of contemporary survey techniques. *Wildlife Research*, *36*(8), 647–658.

Martin-Galvez, D., Peralta-Sanchez, J. M., Dawson, D. A., Martin-Platero, A. M., Martinez-Bueno, M., Burke, T., & Soler, J. J. (2011). DNA sampling from eggshell swabbing is widely applicable in wild bird populations as demonstrated in 23 species. *Molecular Ecology Resources*, *11*(3), 481–493.

Miller, C. R., Joyce, P., & Waits, L. P. (2002). Assessing allelic dropout and genotype reliability using maximum likelihood. *Genetics*, *160*(1), 357–366.

Mills, L. S. (2007). *Conservation of wildlife populations: demography, genetics, and management*. Malden, MA: Blackwell Publishing.

Mills, L. S., Citta, J. J., Lair, K. P., Schwartz, M. K., & Tallmon, D. A. (2000). Estimating animal abundance using non-invasive DNA sampling: promises and pitfalls. *Ecological Applications*, *10*(1), 283–294.

Mitsouras, K., & Faulhaber, E. A. (2009). Saliva as an alternative source of high yield canine genomic DNA for genotyping studies. *BMC Research Notes*, *2*, 219.

Moran, P., Teel, D. J., LaHood, E. S., Drake, J., & Kalinowski, S. (2006). Standardising multi-lab oratory microsatellite data in Pacific salmon: an historical view of the future. *Ecology of Freshwater Fish*, *15*(4), 597–605.

Morgan, D., Nugent, G., Gleeson, D., & Howitt, R. (2007). Are some possums untrappable, unpoisonable, and therefore unmonitorable? Tech. rep., Landcare Research, Christchurch, NZ.

Morin, P. A., Martien, K. K., Archer, F. I., Cipriano, F., Steel, D., Jackson, J., & Taylor, B. L. (2010). Applied conservation genetics and the need for quality control and reporting of genetic data used in fisheries and wildlife management. *Journal of Heredity*, *101*(1), 1–10.

Morrison, M. (2009). *Restoring wildlife: ecological concepts and practical applications*. Science and practice of ecological restoration. Island Press.

National Possum Control Agencies (2008a). *Possum population monitoring using the trap-catch method*. National Possum Control Agencies, Wellington, New Zealand.

National Possum Control Agencies (2008b). *Possum population monitoring using the WaxTag® method*. National Possum Control Agencies, Wellington, New Zealand.

Ng, D., Koh, D., Choo, S., Ng, V., & Fu, Q. (2004). Effect of storage conditions on the extraction of PCR-quality genomic DNA from saliva. *Clinica Chimica ACTA*, *343*(1-2), 191–194.

Nicklas, J., & Buel, E. (2003). Quantification of DNA in forensic samples. *Analytical and Bioanalytical Chemistry*, *376*(8), 1160–1167.

Ogilvie, S. C., Paterson, A. M., Ross, J. G., & Thomas, M. D. (2006). Improving techniques for the WaxTag® possum (*Trichosurus vulpecula*) monitoring index. *New Zealand Plant Protection*, *59*, 28–33.

Olejniczak, M., & Krzyzosiak, W. J. (2006). Genotyping of simple sequence repeats - factors implicated in shadow band generation revisited. *Electrophoresis*, *27*(19), 3724–3734.

Oliveira, E., Padua, J., Zucchi, M., Vencovsky, R., & Vieira, M. (2006). Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology*, *29*(2), 294–307.

O'Reilly-Wapstra, J. M., & Cowan, P. (2010). Native plant/herbivore interactions as determinants of the ecological and evolutionary effects of invasive mammalian herbivores: the case of the common brushtail possum. *Biological Invasions*, *12*(2), 373–387.

Paetkau, D. (2003). An empirical exploration of data quality in DNA-based population inventories. *Molecular Ecology*, *12*(6), 1375–1387.

Paetkau, D. (2004). The optimal number of markers in genetic capture-mark-recapture studies. *Journal of Wildlife Management*, *68*(3), 449–452.

Palero, F., Gonzalez-Candelas, F., & Pascual, M. (2011). MICROSATELIGHT- pipeline to expedite microsatellite analysis. *Journal of Heredity*, *102*(2), 247–249.

Palstra, F. P., & Ruzzante, D. E. (2008). Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? *Molecular Ecology*, *17*(15), 3428–3447.

Parrish, K., & Greenberg, E. (1995). A rapid method for extraction and purification of DNA from dental plaque. *Applied and Environmental Microbiology*, *61*(11), 4120–4123.

Pasqualotto, A. C., Denning, D. W., & Anderson, M. J. (2007). A cautionary tale: lack of consistency in allele sizes between two laboratories for a published multilocus microsatellite typing system. *Journal of Clinical Microbiology*, *45*(2), 522–528.

Peakall, R., & Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, *6*(1), 288–295.

Pech, R., Byrom, A., Anderson, D., Thomson, C., & Coleman, M. (2010). The effect of poisoned and notional vaccinated buffers on possum (*Trichosurus vulpecula*) movements: minimising the risk of bovine tuberculosis spread from forest to farmland. *Wildlife Research*, *37*(4), 283–292.

Petricevic, S., Whitaker, J., Buckleton, J., Vintiner, S., Patel, J., Simon, P., Ferraby, H., Hermiz, W., & Russell, A. (2010). Validation and development of interpretation guidelines for low copy number (LCN) DNA profiling in New Zealand using the AmpFlSTR SGM Plus multiplex. *Forensic Science International Genetics*, *4*(5), 305 – 310.

Piggott, M. (2004). Effect of sample age and season of collection on the reliability of microsatellite genotyping of faecal DNA. *WILDLIFE RESEARCH*, *31*(5), 485–493.

Piggott, M. P., Wilson, R., Banks, S. C., Marks, C. A., Gigliotti, F., & Taylor, A. C. (2008). Evaluating exotic predator control programs using non-invasive genetic tagging. *WILDLIFE RESEARCH*, *35*(7), 617–624.

Pompanon, F., Bonin, A., Bellemain, E., & Taberlet, P. (2005). Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, *6*(11), 847–859.

Prasanth, V., Chandra, S., Hoisington, D., & Jayashree, B. (2006). AlleloBin - a program for allele binning of microsatellite markers based on the algorithm of Idury and Cardon, 1997. Tech. rep., International Crops Research Institute for the Semi-Arid Tropics, Delhi.

Prunier, J., Kaufmann, B., Grolet, O., Picard, D., Pompanon, F., & Joly, P. (2012). Skin swabbing as a new efficient DNA sampling technique in amphibians, and 14 new microsatellite markers in the alpine newt *Ichthyosaura alpestris*. *MOLECULAR ECOLOGY RESOURCES*, *12*(3), 524–531.

R Development Core Team (2011). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Raymond, M., & Rousset, F. (1995). An exact test for population differentiation. *Evolution*, *49*(6), 1280–1283.

Rieseberg, L., Vines, T., & Kane, N. (2010). Editorial and retrospective 2010. *Molecular Ecology*, *19*(1), 1–22.

Rousset, F. (2008). GENEPOP'007: a complete re-implementation of the GENPOP software for Windows and Linux. *Molecular Ecology Resources*, *8*(1), 103–106.

Sargeant, G., Johnson, D., & Berg, W. (1998). Interpreting carnivore scent-station surveys. *Journal of Wildlife Management*, *62*(4), 1235–1245.

Sarre, S., Aitken, N., Clout, M., Ji, W., Robins, J., & Lambert, D. (2000). Molecular ecology and biological control: the mating system of a marsupial pest. *Molecular Ecology*, *9*(6), 723–733.

Sarre, S., Aitken, N., Quinn, A., MacDonald, A., & Cowan, P. (2010). *Trichosurus vulpecula* microsatellite Tv5.64 sequence, direct submission to GenBank. Genetic analyses for limiting the spread of bovine tuberculosis by brushtail possums in New Zealand.
URL `http://www.ncbi.nlm.nih.gov/nuccore/300488325`

Sarre, S. D., & Georges, A. (2009). Genetics in conservation and wildlife management: a revolution since Caughley. *Wildlife Research*, *36*(1), 70–80.

Schulte, U., Gebhard, F., Heinz, L., Veith, M., & Hochkirch, A. (2011). Buccal swabs as a reliable non-invasive tissue sampling method for DNA analysis in the lacertid lizard *Podarcis muralis*. *North-Western Journal of Zoology*, *7*(2), 325–328.

Schwartz, M. K., Luikart, G., & Waples, R. S. (2007). Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution*, *22*(1), 25–33.

Shen, Z. Y., Qu, W. B., Wang, W., Lu, Y. M., Wu, Y. H., Li, Z. F., Hang, X. Y., Wang, X. L., Zhao, D. S., & Zhang, C. G. (2010). MPprimer: a program for reliable multiplex PCR primer design. *BMC Bioinformatics*, *11*, 143.

Sinclair, J. M. C. G., A. R. E. Fryxell (2006). *Wildlife ecology, conservation, and management*. Blackwell Publishing, second ed.

Skrbinsek, T., Jelencic, M., Waits, L., Kos, I., & Trontelj, P. (2010). Highly efficient multiplex PCR of non-invasive DNA does not require pre-amplification. *Molecular Ecology Resources*, *10*(3), 495–501.

Soulsbury, C. D., Iossa, G., & Edwards, K. J. (2009). The influence of evolutionary distance between cross-species microsatellites and primer base-pair composition on allelic dropout rates. *Conservation Genetics*, *10*(3), 797–802.

Soulsbury, C. D., Iossa, G., Edwards, K. J., Baker, P. J., & Harris, S. (2007). Allelic drop-out from a high-quality DNA source. *Conservation Genetics*, *8*(3), 733–738.

Stow, A. J., Minarovic, N., Eymann, J., Cooper, D. W., & Webley, L. S. (2006). Genetic structure infers generally high philopatry and male-biased dispersal of brushtail possums (*Trichosurus vulpecula*) in urban Australia. *Wildlife Research*, *33*(5), 409–415.

Sutton, J. T., Robertson, B. C., & Jamieson, I. G. (2011). Dye shift: a neglected source of genotyping error in molecular ecology. *Molecular Ecology Resources*, *11*(3), 514–520.

Sweetapple, P., & Nugent, G. (2009). Possum demographics and distribution after reduction to near-zero density. *New Zealand Journal of Zoology*, *36*(4), 461–471.

Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., Waits, L. P., & Bouvet, J. (1996). Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, *24*(16), 3189–3194.

Taberlet, P., & Luikart, G. (1999). Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society*, *68*(1-2), 41–55.

Taylor, A., & Cooper, D. W. (1998). Microsatellite markers for the phalangerid marsupial, the common brushtail possum (*Trichosurus vulpecula*). *Molecular Ecology*, *7*, 1771–1788.

Taylor, A., Cowan, P., Fricke, B., & Cooper, D. (2000). Genetic analysis of the mating system of the common brushtail possum (*Trichosurus vulpecula*) in New Zealand farmland. *Molecular Ecology*, *9*(7), 869–879.

Taylor, A. C., Cowan, P. E., Fricke, B. L., Geddes, S., Hansen, B. D., Lam, M., & Cooper, D. W. (2004). High microsatellite diversity and differential structuring among populations of the introduced common brushtail possum, *Trichosurus vulpecula*, in New Zealand. *Genetical Research*, *83*(2), 101–111.

Thomas, M., Brown, J., Maddigan, F., & Sessions, L. (2003). Comparison of trap-catch and bait interference methods for estimating possum densities. *New Zealand Plant Protection*, *56*, 81–85.

Thomas, M., Morgan, D., & Maddigan, F. (2007). Accuracy of possum monitoring using Waxtags. Tech. rep., Pest Control Research Ltd, Christchurch, NZ.

Truong, C. V. C., Groeneveld, L. F., Morgenstern, B., & Groeneveld, E. (2011). MolabIS - An integrated information system for storing and managing molecular genetics data. *BMC Bioinformatics*, *12*, 425.

Valiere, N., Bonenfant, C., Toigo, C., Luikart, G., Gaillard, J. M., & Klein, F. (2007). Importance of a pilot study for non-invasive genetic sampling: genotyping errors and population size estimation in red deer. *Conservation Genetics*, *8*(1), 69–78.

Vallone, P., & Butler, J. (2004). AutoDimer: a screening tool for primer-dimer and hairpin structures. *Biotechniques*, *37*(2), 226–231.

van Oorschot, R. A., Ballantyne, K. N., & Mitchell, R. J. (2010). Forensic trace DNA: a review. *Investigative Genetics*, *1*(1), 14.

van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., & Shipley, P. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, *4*(3), 535–538.

Van Rossum, T., Tripp, B., & Daley, D. (2010). SLIMS - a user-friendly sample operations and inventory management system for genotyping labs. *Bioinformatics*, *26*(14), 1808–1810.

Vargas, M. L., Cruickshank, R. H., Ross, J. G., Holyoake, A. J., Ogilvie, S. C., & Paterson, A. M. (2009). Non-invasive recovery and detection of possum *Trichosurus vulpecula* DNA from bitten bait interference devices (WaxTags). *Molecular Ecology Resources*, *9*(2), 505–515.

Waits, J., & Leberg, P. (2000). Biases associated with population estimation using molecular tagging. *Animal Conservation*, *3*(Part 3), 191–199.

Waits, L., Luikart, G., & Taberlet, P. (2001). Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Molecular Ecology*, *10*(1), 249–256.

Waits, L. P., & Paetkau, D. (2005). Non-invasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. *Journal of Wildlife Management*, *69*(4), 1419–1433.

Walsh, P. S., Erlich, H. A., & Higuchi, R. (1992). Preferential PCR amplification of alleles: mechanisms and solutions. *PCR Methods and Applications*, *1*(4), 241–250.

Warburton, B. (2000). *The brushtail possum: biology, impact and management of an introduced marsupial*, chap. Monitoring possum populations, (pp. 132–138). Manaaki Whenua Press.

Warburton, B., Barker, R., & Coleman, M. (2004). Evaluation of two relative-abundance indices to monitor brushtail possums in New Zealand. *Wildlife Research*, *31*(4), 397–401.

Weir, B. S. (1996). *Genetic data analysis II: methods for discrete population genetic data*. Sunderland, Mass.: Sinauer Associates.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), pp. 1358–1370.

Williams, B., Nichols, J., & Conroy, M. (2002). *Analysis and management of animal populations: modeling, estimation, and decision making*. Academic Press.

Wilson, I. (1997). Inhibition and facilitation of nucleic acid amplification. *Applied and Environmental Microbiology*, *63*(10), 3741–3751.

Yannic, G., Sermier, R., Aebischer, A., Gavrilo, M. V., Gilg, O., Miljeteig, C., Sabard, B., Strom, H., Pouive, E., & Broquet, T. (2011). Description of microsatellite markers and genotyping performances using feathers and buccal swabs for the Ivory gull *Pagophila eburnea*. *Molecular Ecology Resources*, *11*(5), 877–889.