DNA-based Methods for Studying the Diet of Marine Predators

Bruce E. Deagle

BSc, University of Western Ontario

MSc, University of Victoria

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy.

University of Tasmania (March, 2006)

Declaration of originality

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma in any tertiary institute, and that, to the best of my knowledge and belief, the thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Bruce E. Deagle March 30th, 2006

Statement of authority of access

This thesis may be made available for loan and limited copying in accordance with the Copyright Act 1968.

Bruce E. Deagle March 30th, 2006

Thesis Abstract

Diets of large marine predators have been extensively studied to assess interactions with fisheries, monitor links between diet and reproductive success, and understand trophic interactions in marine ecosystems. Since marine species can rarely be observed foraging directly, most studies rely on the identification of prey remains in stomach contents or faeces to determine the prey items being consumed. While this approach has provided a wealth of information, it has several limitations resulting primarily from difficulties identifying digested prey and from biased recovery of remains due to differential digestion. My thesis explores the use of molecular genetic methods in dietary studies of large marine predators. DNA-based identification techniques have been used in several diet studies, but the methods and applications are still in the early stages of development. Through a number of studies, I investigated the ability to recover genetic data from various dietary samples using a range of genetic techniques.

A) *Genetic screening for prey in the gut contents from a giant squid* – I assessed the use of polymerase chain reaction (PCR)-based methods for isolation of prey DNA from an *Architeuthis* gut content sample. A taxonomically informative molecular marker was selected and a screening method developed using denaturing gradient gel electrophoresis. The methodology was used to identify prey from otherwise unidentifiable hard-part remains and the amorphous slurry component of the squid gut sample. The techniques developed here provided a framework for later chapters.

B) *Analysis of prey DNA in faeces of captive sea lions*

Part I: DNA detection, distribution and signal persistence – A feeding trial with captive Steller sea lions (*Eumetopias jubatus*) was carried out to investigate the use of genetic faecal analysis as a tool to study diet. I used group-specific PCR detection to determine: (i) the reliability of prey DNA recovery, (ii) the distribution of prey DNA within faeces and (iii) the persistence of the genetic signal after a prey item was removed from the diet. The proportions of prey DNA in several samples were also determined using a clone library approach to determine if DNA quantification could provide semi-quantitative diet composition data. Results show that the prey DNA could be reliably detected in sea lion faeces and the genetic signal could persist in samples up to 48 hours after ingestion. Proportions of prey DNA isolated from faeces were roughly proportional to the mass of the prey items consumed.

Part II: DNA quantification – Quantitative real-time PCR was used to further investigate if quantitative diet composition data could be obtained through quantification of the DNA present in faeces. I quantified the relative amounts of DNA in three fish species being fed to captive sea lions, then determined the amount of DNA recovered from these prey items in the sea lions' faeces. The results indicate that diet composition estimates based on the relative amounts of DNA in faeces can be biased due to the differential survival of DNA from different fish species; however, these biases may be less than those commonly observed in the conventional analysis of prey hard remains.

C) *Quantification of damage in DNA recovered from faecal samples* – I developed a general method to quantify the frequency of DNA damage present in specific gene regions. The technique was applied to assess the amount of DNA damage in predator and prey DNA recovered from sea lion faeces. The estimated frequency of DNA damage was always higher for the prey DNA than for the predator DNA within a faecal sample. The findings have implications for marker development and comparison of results obtained in future DNA-based diet studies.

D) *Studying seabird diet through genetic analysis of faeces* – I investigated the diet of macaroni penguins **(***Eudyptes chrysolophus***)** through conventional analysis of stomach contents and through the analysis of prey DNA extracted from faeces. Genetic data was obtained from faecal samples using PCR tests to determine the presence or absence of DNA from potential diet items and also using a clone library approach. Approximately half of the faecal samples tested positive for one or more of the prey groups targeted with PCR tests. Euphausiid DNA was most commonly detected in early stages of chick rearing and DNA from a myctophid fish was prevalent in faeces collected later; this trend mirrored the data obtained from the stomach contents. Analysis of prey sequences in "universal" clone libraries revealed a highly biased recovery of sequences from fish prey; this bias is most likely caused by the use of degenerate primers with a higher binding affinity for fish DNA template compared to DNA from other prey groups. Results obtained from the genetic and traditional approaches are compared, and potential future applications of the genetic techniques to studying seabird diet are discussed.

This series of studies has contributed significantly to our understanding of the strengths and the limitations of DNA-based diet analysis. The work identifies situations where genetic methods can be successfully applied to study the diet of marine predators and provides guidance for future studies in this emerging field.

Acknowledgements

The work in this thesis stemmed from research initiated by two of my supervisors, Nick Gales and Simon Jarman from the Australian Antarctic Division (AAD). Their enthusiasm, valuable advice and unfettered support made my project possible. Mark Hindell, my university supervisor, welcomed me into his group and provided a home for me at the Antarctic Wildlife Research Unit (AWRU). An excellent cohort of Research Fellows, PhD students and Honours students were at the AWRU during my tenure, making this a good place to be – thanks to you all.

The Steller sea lion feeding trial carried out at the Vancouver Aquarium could not have been done without the support from Andrew Trites and Dominic Tollit. Andrew agreed to the make room for my project in the busy research schedule at the aquarium. Dom generously lent his time, experience and enthusiasm in order to make sure the feeding trial happened, and that I got all the samples that I required. My time at the aquarium was also made enjoyable and productive due to help from members of the UBC Marine Mammal Research Unit (particularly Susan Heaslip, Rebecca Barrick, Chad Nordstrom and David Rosen) and the marine mammal trainers (Troy Neale, Nigel Waller and Billy Lasby). A special thanks to the sea lions (Hazy and Nuka) for their vital contributions.

The macaroni penguin diet study was part of a large research expedition to Heard Island undertaken by the AAD in the summer of 2003–04. Karen Evans and Rowan Trebilco carried out the field work with me at Capsize Beach. Karen's meticulous planning and stomach flushing expertise were very much appreciated, as were Rowan's bad jokes and enthusiasm. Thanks to all the expeditioners for their time and friendship during the trip, and of course thanks to the penguins for putting up with us. Back in Hobart, Sarah Robinson helped with the tedious sorting of the macaroni penguin stomach samples and identified the otoliths that we recovered. John Kitchener helped me with identification of amphipods and euphausiids.

Most of the laboratory work was carried out in the Molecular Genetics Laboratory at the University of Tasmania – Adam Smolenski deserves credit for keeping everything running smoothly in this busy facility. The laboratory costs of the project were covered by the AAD, without this funding the project simply would not have happened. Abe Passmore was a good mate to share so many hours in the lab with. Conversations with Abe over coffee provided many welcome breaks and valuable insight into the project.

The writing of this thesis has benefited from input from a large number of people. Regular discussions with Simon Jarman and his "goblin army" provided me with a constant source of ideas (special thanks to Abe, Ruth, Glenn, Kevin and Megan). All of my supervisors and various co-authors helped by editing and commenting on each of the chapters. Abe Passmore and Glenn Dunshea also provided useful comments on some of the chapters. Anonymous reviewers made useful contributions to the published work (Chapters 2, 3 and 4) and comments from my thesis reviewers (Bill Symondson and Scott Baker) were also very constructive.

My fondest thanks go to Paige who provided an endless supply of encouragement, support and love. If that wasn't enough, she also endured almost daily reports of my successes and failures, provided helpful advice and therapeutic discussion, carefully edited drafts of each chapter and provided much needed statistical guidance. My family also provided strong support over the project – even from the other end of the globe.

Table of Contents

List of Figures

List of Tables

Chapter 1

General Introduction

“I suppose I ought to eat or drink something or other; but the great question is, what?”

— LEWIS CARROLL,
Alice`s Adventures In Wonderland

1.1 Introduction

The determination of trophic relationships within biological communities has been the central goal of innumerable ecological studies (Pimm 2002). In the marine environment, concerns over the effects of increased global fishing pressures, climate change and other human induced ecosystem modifications have resulted in increasingly detailed studies of oceanic food webs (Trites 2003). Data from these studies are critical for implementation of current conservation strategies, such as ecosystem-based approaches to fisheries management (Garcia & Cochrane 2005) and the development of marine protected areas (Sale *et al.* 2005). Research on the diet of upper trophic level animals is especially extensive due to their long history of exploitation (Myers & Worm 2005), fluctuations in their population sizes (Springer *et al.* 2003) and their sensitivity to changes in ecosystem structure (Reid *et al.* 2005). Methods of determining diets of these species are varied. Numerous studies have employed gut content analysis (e.g. Ealey 1954; Croxall *et al.* 1985), identification of hard remains in faecal samples (e.g. Tollit & Thompson 1996; Orr *et al.* 2004), analysis of fatty acid signatures (e.g. Phillips *et al.* 2001; Iverson *et al.* 2004), analysis of tissue stable isotope ratios (e.g. Hobson *et al.* 1997; Cherel *et al.* 2000) or observations of feeding behaviour (e.g. Bowen *et al.* 2002). The diversity of methods used to study diet reflects the fact that none are universally adequate, and the amount of interest in diet studies indicates further refinements would be valuable.

In this thesis, I investigate the applicability of recently developed DNA-based methods for studying the diet of top-level marine predators. The focus is primarily on obtaining diet information through genetic analysis of faeces collected from pinnipeds and penguins, but the methods can be applied more generally to identify prey remains in gut contents or faeces from a range of organisms. In this general introduction, I will begin with a brief review of the non-genetic methods used to study the diet of marine mammals and seabirds, outline the current state of knowledge in the field of DNA-based diet analysis and highlight some relevant literature in allied fields. I will finish by outlining the studies presented in later chapters of the thesis.

1.2 Conventional dietary analysis methods

1.2.1 Stomach content analysis

Traditionally, the diet of marine mammals and seabirds was determined through analysis of stomachs contents (Pitcher 1980; Croxall *et al.* 1985). For a diet study carried out in the Gulf of Alaska, Pitcher (1980) collected 548 harbour seals "by rifle"; almost all early diet studies used lethal sampling directly or obtained stomach samples from commercial hunts. In most situations, destructive sampling is now ethically unacceptable and commercial hunts of marine mammals and seabirds are very limited. Procedures have been developed to allow collection of stomach contents without using lethal sampling (Hyslop 1980; Wilson 1984; Harvey & Antonelis 1994). While these methods are a huge improvement ethically, they still can be quite invasive. For example, the stress of handling and loss of fluid during the stomach flushing procedure has been reported to cause hyperthermia in penguins, which can be lethal to the sampled bird in rare cases (e.g. Chiaradia *et al.* 2003). In addition, the collection of stomach contents from seabirds typically occurs as the adults return to feed chicks; interception of this food has the potential to impact on chick development (Chiaradia *et al.* 2003). In pinnipeds, sample collection usually involves the use of anaesthetic agents for animal immobilization (van den Hoff *et al.* 2003), a procedure that is not without risk. Due to these potential impacts, the number of stomach samples that can be collected in diet studies is often restricted.

Ethical concerns aside, there are several other problems with the approach. One drawback is that prey remains present in stomachs represent only very recently consumed prey. This means that a large number of samples need to be analysed to obtain a representative picture of the diet of a population. In addition, often a considerable proportion of stomachs are empty. Food was present in less than half of the 548 harbour seal stomachs collected by Pitcher (1980). Stomachs that do contain food usually contain a mixture of partially digested soft tissue and various hard parts (fish otoliths and other bones, crustacean exoskeletons and cephalopod beaks). Identification of these prey items is time consuming and requires considerable taxonomic expertise. Often a significant proportion of the sample cannot be positively identified based on morphology (Croxall *et al.* 1985; Scribner & Bowman 1998). This

introduces significant bias into the analysis because some prey species will rapidly become unrecognizable, while others are resistant to digestion or possess easily identifiable hard parts (Hartley 1948; Hyslop 1980; Gales 1988). This bias is made worse by differential passage rates of remains from various prey species – such as the accumulation of large cephalopod beaks in seal stomachs (Harvey & Antonelis, 1994). Another problem is that different enumeration methods (frequency of occurrence, percentage by number, percentage by mass, or reconstitution of biomass) can lead to considerably different pictures of diet composition (Hyslop 1980; Lescroel *et al.* 2004). A final difficulty with the approach is that the collection of stomach samples requires animal capture. Depending on the predator species being studied, this can be quite time consuming and operationally difficult (van den Hoff *et al.* 2003).

Even with the weaknesses outlined above, stomach content analysis can clearly provide valuable dietary information. Provided a large number of samples are collected and analysed, and various enumeration methods are sensibly employed, a reasonable assessment of diet composition can be expected. The approach can also provide data that is difficult to obtain using alternate methods, such as direct information on the size of prey being consumed and data on meal size (Barlow & Croxall 2002). For seabirds, analysis of stomach contents obtained by stomach flushing remains the standard method of diet analysis (Green *et al.* 1998; Hull 1999).

1.2.2 Faecal analysis

Over the last 20 years, the study of pinniped diet has been primarily carried out through analysis of prey remains in faeces (Hall *et al.* 1998; Harcourt *et al.* 2002; Lake *et al.* 2003). The advantages of this method over stomach content analysis are that sample collection is easier, the disturbance to animals is minimal and often a high proportion of samples contain prey remains. Despite these benefits, many of the same drawbacks remain: faecal samples provide only a snapshot of prey consumed over a short time period (Hammond & Rothery 1996); the identification of prey hard parts is a slow process requiring development of appropriate taxonomic keys or extensive reference collections (Olesiuk 1993); and the data can be biased due to species-

specific differences in hard part survival during digestion (Jobling 1987; Harvey 1989).

The biases caused by differential digestion have been well documented (Jobling 1987). Some prey species have no hard parts that survive digestion and are missed entirely in faecal-based estimates of diet composition (Da Silva & Neilson 1985). For those prey species that are represented by hard parts in faeces, only a fraction of the hard parts ingested remains intact (Harvey 1989; Tollit *et al.* 1997; Bowen 2000). This results in the dietary importance of prey species with small, fragile bones being underestimated and prey species with big, robust bones being overestimated (Bowen 2000; Tollit *et al.* 2003). In addition to determining prey numbers, many studies calculate the original prey biomass (Laake *et al.* 2002). This is accomplished by using the lengths of otoliths and beaks to estimate the length and masses of the ingested fish and cephalopods. This procedure can introduce further bias in diet estimates because hard parts erode at different rates (depending on species and size of prey). Captive feeding trials have been employed to assess the changes in number and size of hard parts that occur as they pass through the digestive tract of various pinnipeds (Harvey 1989; Tollit *et al.* 2003) and results from these studies can be used to develop correction factors in order to compensate for digestion biases. While correction factors are certainly useful, it has been demonstrated that digestion rates are affected by many different factors (e.g. species of predator, activity level of predator and size of prey) making their development and application challenging (Bowen 2000; Tollit *et al*. 2003).

An additional limitation of dietary studies based on faecal analysis is that there is usually no link between a sample and a particular animal. This means that dietary differences between categories of the predator (such as sex or age class) cannot be explored (for exception see Reed *et al*. 1997).

Nevertheless, as with stomach content analysis, faecal analysis can provide useful dietary data, especially in situations where the system has been relatively well studied. This allows major biases to be accounted for and meaningful interpretations can be made when large numbers of samples are analysed. Even where biases associated with the approach exist, they should remain relatively consistent within predator species; this allows temporal and spatial changes in prey composition to be monitored. It should be noted that conventional faecal analysis is not a useful method

for dietary analysis in most seabirds and some pinnipeds due to the very low recovery of prey hard parts in their faeces (Hartley 1948; Gales & Cheal 1992).

1.2.3 Tissue biomarker methods

In order to overcome problems associated with stomach content and faecal analysis, indirect biochemical approaches have been developed to study diet. Two methods are commonly applied to marine predators and will be briefly considered here: fatty acid signature analysis and stable isotope techniques.

Fatty acid analysis of predator tissue can provide information on diet since many prey species differ in their fatty acid composition (due to differences in lipid synthesis pathways) and, upon consumption, these fatty acid signatures are incorporated into predators' fat stores (Iverson *et al.* 2004). Due to the large variety of fatty acids present in phytoplankton, fatty acids are particularly good biomarkers in marine ecosystems (Raclot *et al.* 1998; Iverson *et al.* 2004). The potential of this approach has been recognised for well over 20 years (Bishop *et al.* 1983), but only recently has the method been widely applied and assessed in detailed diet studies (Bradshaw *et al.* 2003; Grahl-Nielsen *et al.* 2003; Iverson *et al.* 2004). While numerous studies have inferred differences in diet based on fatty acid signatures, many factors complicate interpretation of results. These factors include: geographic and temporal variation in prey fatty acid profiles (Iverson *et al.* 1997); metabolic changes to fatty acids within the predator (Iverson *et al.* 2004); selective use/deposition of fatty acids (Hooker *et al.* 2001); and questions over the appropriate statistical treatment of data (Grahl-Nielsen 1999). The taxonomic resolution of the technique (i.e. ability to classify prey groups present in the diet) varies considerably. Some studies obtain species level discrimination (Iverson *et al.* 2004), whereas others only differentiate between broad prey groups (e.g. fish- or squid-dominated diets; Bradshaw *et al*. 2003)

Like fatty acid analysis, stable isotope techniques rely on a chemical signature from prey items being incorporated into a predator's tissues. Specifically, ratios of stable isotopes of nitrogen ($^{15}N/^{14}N$) and carbon ($^{13}C/^{12}C$) have been used in dietary studies (Hodum & Hobson 2000). The rare heavy isotope of nitrogen is enriched 3-4‰ per trophic level relative to the lighter isotope; this means the ratio of $^{15}N/^{14}N$ in

an animal's tissues can be used to determine its trophic position (Hobson *et al.* 1997). Carbon only experiences minor enrichment, but the carbon isotope ratio provides information on contributions from different sources of primary production (e.g. freshwater versus marine, inshore versus offshore) (Cherel *et al.* 2000). The taxonomic resolution of the technique has been improved with the development of mixing models (Phillips & Gregg 2001; Phillips *et al.* 2005), but stable isotope techniques are generally most appropriate for assessment of broad dietary shifts and changes in foraging location (Bocher *et al.* 2000).

A major advantage of using one of the tissue biomarker methods to study diet is that they provide data over longer time scales. This means that fewer samples need to be analysed to obtain a community level view of diet. It also means information about prey consumed during the non-breeding period can be obtained. Most of the traditional diet data available for sea birds and pinnipeds has been collected during the breeding season, when animals are easily accessible and stomach/faecal samples can be obtained, even though this may only be a small proportion of the annual cycle.

1.3 Application of DNA-based techniques to diet analysis [1]

The ability to identify prey remains at a molecular level promises to help overcome several of the inherent limitations of existing techniques (Symondson 2002). The application of molecular species detection techniques to dietary samples can allow for more accurate taxonomic identification of prey remains (Purcell *et al.* 2000) and could speed up sample processing (Walter *et al.* 1986). In addition, a molecular approach can allow dietary information to be retrieved from samples which traditionally have been considered uninformative (e.g. bird faeces; Sutherland 2000). Finally, detection of prey by molecular means could reduce the biases caused by reliance on identification of digestion-resistant morphologically recognisable hard parts (Symondson 2002). A number of studies have delivered on some of these promises (Table 1.1 and 1.2), but this field of research is still in the early stages of development. Most published studies have focused on technical aspects of the research, or present results from preliminary studies.

[1] For this section (1.3) I have reviewed only literature published before 2004 since this is the work that was available when I starting writing the first chapters of my thesis. Relevant papers that have been published since then will be discussed in the context of my work in later chapters of the thesis.

Early diet studies using molecular identification of prey relied on the use of monoclonal antibody techniques to detect prey-specific proteins in dietary samples. This immunological approach has been used extensively to detect species of prey being consumed by insect predators (see recent review by Symondson 2002), and has also been applied to detect prey in the stomach samples of seabirds (Walter *et al.* 1986) and squid (Kear 1992). Large numbers of gut samples can potentially be screened rapidly using monoclonal antibodies, but development of each specific assay is time-consuming. In addition, it is unlikely predator faecal samples could be tested using the approach because only minute quantities of undigested prey protein survive digestion. The development of polymerase chain reaction (PCR) (Mullis *et al.* 1986) has allowed prey DNA to be detected in both stomach samples and faeces, and this approach has almost completely supplanted the use of immunological techniques. In a recent review of molecular identification of prey in predator diets, Symondson (2002) concluded that: "PCR-based techniques have proven to be highly effective and versatile in recent laboratory trials and are likely to rapidly displace all other approaches".

The PCR technique allows exponential amplification of a specific DNA region; the amplified region is defined by short, user-specified DNA primers (Mullis & Faloona 1987). Features of PCR that make it particularly well suited for use in DNA-based diet studies are: (1) it is extremely sensitive, requiring only very small amounts of substrate DNA; and (2) it is very specific, so only DNA fragments that match the PCR primers will be amplified. Virtually all DNA-based diet studies have used PCR to amplify DNA present in dietary samples in order to obtain enough material for subsequent analyses. Two general approaches can be taken to identify prey DNA from dietary sources using PCR. In the first approach, PCR primers are designed to amplify DNA only from a specific target prey species (or group of species). Successful PCR amplification of DNA extracted from diet samples using these primers indicates the presence of DNA from the particular target (e.g. Agustí *et al*. 2003). The second approach utilizes primers which bind to DNA regions conserved in a broad range of prey items. The origin of the DNA molecules present in PCR products amplified with conserved primers are subsequently determined either by sequencing (e.g. Poinar *et al.* 2001), restriction enzyme analysis (e.g. Asahida *et al.* 1997) or through hybridization techniques (e.g. Rosel & Kocher 2002).

The use of DNA-based methods in diet studies can focus on either material recovered from stomach contents or faeces. In the following review of the work that has been done to date, I will address each separately.

1.3.1 DNA-based stomach content analysis

In invertebrate diet studies, few non-molecular approaches exist to study diet (Symondson 2002); this may explain why research on these animals has featured so prominently in the development of DNA-based diet analysis (Table 1.1). The initial investigation of PCR-based methods to identify prey present in stomach contents was in a study on the predation by sand shrimp (*Crangon affinis*) on larval stone flounder (*Kareius bicoloratus*) (Asahida *et al.* 1997). This laboratory based study determined the ability of PCR to detect fish mitochondrial DNA (mtDNA) in the stomach contents of sand shrimp at various time points after predation (0-5 hours). The researchers used both fish-specific primers and *K. bicoloratus*-specific primers that targeted relatively large fragments of the mtDNA control region (~ 2600 bp and 1400 bp respectively). The results showed that the larger fish DNA fragments could only be detected in 50% of the stomach samples after three hours, and this DNA fragment was undetectable five hours after predation. Detection was better for the 1400 bp fragment; it could be detected in 50% of the samples collected after five hours. Besides highlighting the potential of this approach, this study also discussed the relative merits of using species-specific versus more general PCR primers.

Shortly after this study, a number of PCR-based diet experiments were published focusing on the application of the approach in studies of predator-prey relationships among terrestrial arthropods (reviewed by Symondson 2002). In one of the early arthropod studies, Zaidi *et al*. (1999) carried out an experimental feeding trial with carabid beetles as predators and mosquitoes as prey. Prey DNA was detected by PCR using primers targeting a mosquito-specific region of a multiple-copy nuclear esterase gene. Two fragment sizes were amplified (146 and 263 bp). The shortest fragment of mosquito DNA could be detected for up to 28 hours (the maximum time tested for in the feeding trial); whereas detection of the larger fragment was less consistent. This study clarified the advantages of targeting shorter DNA fragments and, along with other experimental feeding trials (Agustí *et al.* 1999;

Chen *et al.* 2000; Hoogendoorn & Heimpel 2001), opened the way for field-based applications. The first field-based study with terrestrial arthropods examined predation by spiders on three species of Collembola in an agricultural ecosystem (Agustí *et al*. 2003). These researchers targeted the mtDNA cytochrome oxidase I gene (211-276 bp), and in preliminary laboratory trials found prey DNA was detectable in 100% of spiders for 24 hours after being fed a single collembolan. In the field, they found collembolan DNA in almost half of the collected spiders. The Collembola species detected most frequently in the spider guts (*Isotoma anglicana*) was the least numerous in the field (based on visual surveys), indicating this prey was preferred by the spiders.

There have been only a few studies that have focused on the genetic identification of prey remains in vertebrate gut contents (Table 1.2). Unlike the situation for invertebrates, where DNA is extracted from the entire stomach contents (Asahida *et al.* 1997), or the entire predator (Zaidi *et al.* 1999), vertebrate studies can focus analysis on individually isolated prey remnants which are morphologically unidentifiable. This was the case in a study carried out by Scribner & Bowman (1998). These researchers isolated the remains of juvenile ducks and geese present in the stomachs of glaucous gulls (*Larus hyperboreus*) and then identified these remains using PCR amplification of microsatellite loci. Their results showed a significant proportion of gull stomachs contained goslings of emperor geese, *Chen canagica* (26.3%), Canada geese, *Branta canadensis* (41.4%) and greater white fronted geese, *Anser albifrons* (23.2%). Extrapolation of their data to the population of gulls and these three geese species in the region indicated a mortality rate of up to 40% of the hatched goslings due to gull predation. Another study on a vertebrate predator developed a PCR-based assay for the detection of larval cod (*Gadus morhua*) in homogenised stomach contents of predatory fish (Rosel & Kocher 2002). The assay was validated in an experimental feeding trial and tested on some field collected samples. Two additional rather specialised studies have used DNA-based gut content analysis to study aspects of vertebrate diet: (DeWoody *et al.* 2001) used microsatellite markers to documented filial cannibalism (i.e. consumption of one's own offspring) in nest-tending fish; and (Rollo *et al.* 2002) used the approach to analyse the intestinal contents of the Tyrolean Iceman (a naturally mummified corpse roughly 5000 years old).

Table 1.1 *Dietary studies*[a] *of invertebrate predators investigating the use of DNA-based methods for prey detection in gut contents.*

Predator	Diet Items Detected	Sample Tested	Gene	Marker Size	Sample Size[b]	Field-based	Reference
Sand shrimp (*Crangon affinis*)	Stone flounder (*Kareius bicoloratus*)	Stomach contents	mtDNA D-loop	1400 – 2600	24	No	Asahida *et al.* (1997)
Predatory bug (*Dicyphus tamanini*)	Moth eggs (*Helicoverpa armigera*)	Whole predator	Nuclear Various	250 – 1100	20	No	Agustí *et al.* (1999)
Carabid beetle (*Pterostichus cupreus*)	Mosquito (*Culex quinquefasciatus*)	Whole predator	Nuclear Esterase	150 – 260	70	No	Zaidi *et al.* (1999)
Predatory bug (*Dicyphus tamaninii*)	Whitefly (*Trialeurodes vaporariorum*)	Whole predator	Nuclear Various	310 – 2100	20	No	Agustí *et al.* (2000)
Lady beetle and lacewing	Cereal aphids (*Rhopalosiphum* spp.)	Whole predator	mtDNA COII	80 – 390	~ 100	No	Chen *et al.* (2000)
Carabid beetle (*Coleomegilla maculate*)	Corn borer moth eggs (*Ostrinia nubilalis*)	Whole predator	Nuclear rDNA	150 – 490	94	No	Hoogendoorn & Heimpel (2001)
Linyphiid spiders	Various Collembola species	Whole predator	mtDNA COI	210 – 280	82	Yes	Agustí *et al.* (2003a)
Predatory bug (*Anthocoris tomentosus*)	Pear psylla (*Cacopsylla pyricola*)	Whole predator	mtDNA COI	190 – 270	NA	No	Agustí *et al.* (2003b)

[a] Includes primary literature published before 2004 and listed in the Web of Science© database. Diet studies on parasites are excluded (e.g. studies analysing the source of insects' bloodmeals).

[b] Indicates the approximate number of separate samples tested for prey DNA. Many studies analysed each sample with several different PCR tests.

Table 1.2 *Dietary studies*[a] *of vertebrate predators investigating the use of DNA-based methods for prey detection in stomach contents or faeces.*

Predator	Diet Items Detected	Sample Tested	Gene	Marker Size	Sample Size[b]	Field-based	Reference
Brown bear (*Ursus arctos*)	Plants	Faeces soft matrix	Chloroplast *rbc*L	360	3	Yes	Höss *et al.* (1992)
Glaucous gull (*Larus hyperboreus*)	Various ducks and geese (Anatidae and Anserinae)	Stomach prey items	Nuclear Microsatellite	60 – 120	99	Yes	Scribner & Bowman (1998)
Harbour seal *(Phoca vitulina*)	Salmonid species (Onchorhynchus spp.)	Faeces hard-parts	mtDNA COIII	370	39 (116 bones)	Yes	Purcell *et al.* (2000)
Various passerine birds	Various insect species	Faeces soft matrix	mtDNA rDNA	125	> 50	Yes	Sutherland (2000)
Species of nest-tending fish	Conspecific cannibalized embryos	Stomach prey items	Nuclear Microsatellite	NA	38	Yes	DeWoody *et al.* (2001)
Penguin and whale	Krill (Malacostraca)	Faeces/stomach	Nuclear rDNA	220 – 330	7	Yes	Jarman *et al.* (2002)
Various predatory fish	Larval cod (*Gadus morhua*)	Stomach contents	mtDNA rDNA	130	220	Yes	Rosel & Kocher (2002)
Various marine predators	Japanese flounder (*Paralchthys olivaceus*)	Stomach contents	mtDNA D-loop	150	22	Yes	Saitoh *et al.* (2003)

[a] Includes primary literature published before 2004 and listed in the Web of Science© database, with the exception of Purcell *et al.* (2000) and Sutherland (2000) which were included due to their relevance to the thesis topic. Studies based on the analysis of ancient remains were not included.

[b] Indicates the approximate number of separate samples tested for prey DNA. Many studies analysed each sample with several different PCR tests.

1.3.2 DNA-based faecal analysis

The first study to demonstrate that dietary information could be obtained from DNA present in faeces was published by Höss *et al.* (1992). These researchers amplified a 356 bp fragment of DNA from European brown bear faeces with plant-specific PCR primers targeting the chloroplast ribulose-bisphosphate carboxylase (*rbc*L) gene. A positive PCR result indicated the presence of plant DNA, and direct sequencing followed by comparison with known sequences revealed the recovered plant DNA was from *Photinia villosa*. Berries from this plant form a dominant component of these bears' diet during late summer, so the dietary data itself was not surprising. However, the fact that plant DNA could pass through the bear's digestive system indicated that a wealth of previously unattainable dietary data existed in faeces. Little progress followed this initial success, probably because the lack of sequence data for most species meant a monumental sequencing effort would have been required to identify any unknown DNA sequences recovered from faeces. The use of species-specific primers was also limited by the lack of sequence data available at this time – since the development of these primers relies on comparative sequence data. Dietary analysis based on DNA in faeces was eventually continued by ancient DNA researchers who documented the diet of an extinct ground sloth through analysis of ancient faecal material (coprolites) (Poinar *et al.* 1998). They amplified a 183 base pair fragment of the chloroplast *rbc*L gene, cloned the PCR products and sequenced 72 clones. The recovered plant sequences were compared to sequences in GenBank and seven groups of plants were identified to family or order (Poinar *et al*. 1998). Further studies on plants in ground sloth diet and the dietary diversity of archaic humans (animal and plant components) have since been carried out by analysis of DNA obtained from coprolites (Hofreiter *et al.* 2000; Poinar *et al.* 2001).

Further diet work relying on analysis of DNA in the faeces of contemporary species was published by Purcell *et al.* (2000). This study used DNA-based methods to identify salmonid bones recovered from the faeces of harbour seals. Salmonid otoliths can be identified to species, but they are fragile and not well represented in seal faecal samples (Purcell *et al.* 2000). Other skeletal remains from salmon that are more commonly found in seal faeces can only be identified to family level. In order to determine predation rates of the harbour seals on specific species of salmon (including

some species from endangered populations), Purcell *et al.* (2000) extracted DNA from 146 bones obtained from seal faeces. They were able to amplify PCR products from 102 of the samples and the DNA was identified to species level by sequencing or restriction enzyme analysis. In their study, identification of prey still relied on some hard parts surviving digestion, and the bones had to be individually isolated and identified to family level before genetic analysis. Nevertheless, the DNA-based approach provided important data that could not have been obtained by other means.

A detailed assessment of the use of DNA-based faecal analysis in diet studies of insectivorous birds was carried out by Sutherland (2000). In this unpublished thesis, PCR was used to amplify insect DNA from faeces of captive starlings (*Sturnus vulgaris*) and from faeces of wild nestlings of blue tits (*Parus caeruleus*), great tits (*P. major*) and swallows (*Hirundo rustica*). Sutherland (2000) used PCR primers targeting relatively conserved regions of the 12S mtDNA gene in all of his studies and identified recovered prey DNA by restriction enzyme analysis or sequencing. The captive feeding trial with starlings demonstrated that insect DNA could be reliably recovered from passerine bird faeces. It also indicated that it might be possible to quantify the relative amounts of different prey species being consumed by determining the proportions of DNA present in faecal samples. The quantitative conclusions were based on results from DNA clone libraries created using mixtures of faeces obtained from starlings fed either locusts (*Locusta migratoria*) or grasshoppers (*Schistocerca gregaria*). The faeces from starlings on the two different diets were mixed in ratios of 9:1, 3:1, 1:1, 1:3 or 1:9; these ratios corresponded well to the proportion of clones from the prey species present in the respective clone libraries. In the analysis of wild nestling diets, results from the two species of tits showed the faeces from both species contained a similar range of mainly lepidopteran prey. The diet diversity of swallow nestlings was compared between sites on organic farms and conventional farms. Due to a relatively small sample size and some difficulty identifying recovered insect DNA, no firm conclusions could be made concerning differences between the sample sites, but once again DNA from a variety of insect prey species was recoverable.

One of the driving forces behind the development of DNA-based faecal analytical methods for studying diet has been that the approach potentially provides a non-invasive alternative to stomach content analysis (Symondson 2002). This was the motivation for a study of prey DNA in faeces of whales and penguins (Jarman *et al.*

2002). The faeces of these groups contain very few morphologically identifiable prey remains, and their diet is generally studied through stomach content analysis. The lethal sampling of whales for scientific research (including diet studies) has been particularly controversial (Ichii & Kato 1991; Aron *et al.* 2000). Jarman *et al.* (2002) collected faeces from pygmy blue whales (*Balaenoptera musculus*) and Adelie penguins (*Pygoscelis adeliae*). The species of krill being consumed by these predators was determined using krill-specific primers to amplify DNA extracted from their faeces. The main conclusion of this study was that recoverable prey DNA was present in amorphous faecal material obtained from these marine predators, and this material could be used in diet studies. This finding provided the impetus for the initiation of my PhD project.

1.4 Relevant research in allied fields

In the review of DNA-based diet analysis presented so far, I have only discussed genetic studies that have focused directly on obtaining dietary information. There is also a large amount of highly relevant work that has been done in related fields. This includes research on non-invasive genetic sampling, ancient DNA, microbial biodiversity and DNA barcoding. I will briefly outline some of the important contributions from these fields.

Non-invasive genetic sampling generally refers to research that obtains DNA samples from free-ranging animals without having to capture or even observe the animals of interest (Taberlet & Waits 1998). The DNA can be obtained from shed hair or feathers, eggshells, urine or faeces. DNA from these sources is usually present in low quantities and the DNA is often degraded. Prey DNA extracted from faeces or gut contents share these properties, therefore many of the methods that have been developed in this field are relevant for DNA-based diet studies (Taberlet *et al.* 1999; Morin *et al.* 2001; Broquet & Petit 2004). Particularly relevant are methods designed specifically for use with faecal DNA, such as preservation of DNA in faeces, (Frantzen *et al.* 1998) and faecal DNA extraction methods (Wehausen *et al.* 2004). A number of studies in this field have also assessed the ability to detect various genetic markers in degraded DNA samples (Kohn *et al.* 1995; Taberlet *et al.* 1996).

Scientists from the field of ancient DNA research were the first to characterize dietary DNA in faeces (Höss *et al.* 1992); this is not surprising since these researchers pioneered the recovery of degraded DNA using PCR (Pääbo 1989; Pääbo 1990; Höss *et al.* 1996). Much of the ancient DNA literature (recently reviewed by Pääbo *et al*. 2004) is relevant to DNA-based diet studies, especially the detailed studies that have been carried out on DNA damage (Hofreiter *et al.* 2001; Gilbert *et al.* 2003; Mitchell *et al.* 2005). The persistent problem of PCR contamination in ancient DNA studies is also highly pertinent to DNA-based diet studies. The extreme sensitivity of PCR amplification means that great care has to be taken to prevent false positive results being obtained. Numerous precautionary measures have been outlined by ancient DNA researchers to prevent the occurrence of this problem when working with small amounts of DNA template (Cooper & Poinar 2000; Gilbert *et al.* 2005).

Researchers interested in microbial biodiversity have adopted a genetic approach to characterize species diversity in environmental samples due to the difficulty in culturing and/or microscopically identifying the majority of microbes present in the environment (Delong *et al.* 1994). The methodology often used in these studies is to isolate total DNA obtained from a microbial community, amplify a region of DNA from the sample using PCR and then identify and quantify members of the microbial community using sequence polymorphisms (e.g. Delong *et al.* 1994; Orphan *et al.* 2000; Braker *et al.* 2001). This is the same approach that has been adopted by many DNA-based diet studies, a reflection that the objective of dietary work is essentially the same as that of microbial biodiversity studies (i.e. to identify and quantify the "community" of prey DNA in a sample). Several of the methods microbiologists have developed could be applied in DNA-based diet studies. For example, microbiologists commonly use genetic fingerprinting techniques (e.g. denaturing gradient gel electrophoresis) to allow physical separation of the DNA sequence polymorphisms in their samples (Muyzer 1999); this approach could be used to determine DNA diversity in dietary samples. The techniques used for quantitative analysis of microbe species present in complex mixtures of DNA are also highly relevant. These analyses have been done through cloning (Ward *et al.* 1990) and more recently using real-time PCR (Fierer *et al.* 2005); both approaches could be applied to quantify DNA in dietary samples. Finally, a large amount of important literature has been published by microbiologists on the pitfalls of using PCR-based

methods to obtain an accurate representation of the endogenous DNA present in a sample (reviewed by Forney *et al*. 2004).

DNA barcoding refers to the development of large-scale standardized sequence databases in order to allow taxonomic classification of unknown biological samples. This type of database has been in use by microbiologists for many years (Wuyts *et al.* 2001), and a comprehensive database of animal DNA sequences from the mitochondrial cytochrome oxidase I gene (COI) is in the process of being developed (Hebert *et al*. 2003). Information from the COI database could be directly applied to identify prey DNA isolated in diet studies if the PCR assays used target the COI gene. Even without making direct use of this DNA barcoding database, much of the literature in this field is useful since it provides a conceptual framework for the application of DNA-based taxonomy (Hebert *et al*. 2004; Blaxter *et al*. 2005).

1.5 Thesis outline

The data chapters of this thesis (Chapter 2 – Chapter 6) were written as separate scientific papers that have either been published, submitted, or are in the process of being published. These papers include research on a diverse group of marine predators and each focuses on distinct questions; the thesis is tied together by the common theme of developing a DNA-based approach for studying diet.

Chapter 2 details the analysis of the gut contents from a giant squid (*Architeuthis sp.*). This chapter differs from the rest of the thesis by focusing on DNA-based analysis of material from stomach contents, rather than on faecal samples. Considering only one sample was available, the study is quite detailed. However, very little is known about the diet of giant squids and previous dietary studies on squids commonly report significant numbers of unidentifiable remains, indicating that the development of a DNA-based approach for identification of their prey would be useful. In addition, the prey species likely to be consumed by the giant squid (fish, squid and crustaceans) were the same groups that I was interested in detecting in dietary samples from other predators. Thus, this study provided an opportunity to design suitable PCR primers to look at these prey groups and allowed me to trial the use of denaturing gradient gel electrophoresis. These methods are applied further in studies carried out later in the thesis.

Chapter 3 outlines a feeding trial carried out with captive Steller sea lions (*Eumetopias jubatus*) to investigate the feasibility of using DNA-based analysis of the soft matrix of faeces to study pinniped diet. Based on previous diet work (both molecular and conventional) it was apparent that controlled feeding experiments with captive animals would be required to validate the DNA-based approach. No previous controlled feeding trials investigating DNA-based diet methods had been carried out with mammals. In field-collected pinniped faeces, prey DNA had been isolated from hard parts, but not from the soft-matrix of faeces (Purcell *et al.* 2000). Since it is known that digestion bias effects recovery of hard-parts, I wanted to focus on prey DNA extracted from soft matrix of faeces. The feeding trial was designed to addresses a number of fundamental questions:

- Can prey DNA be reliably detected in the soft matrix of sea lion faecal samples?
- Can DNA from prey items fed as a small proportion of the diet be detected?
- How is prey DNA distributed in faeces?
- How long after ingestion does a signal from prey DNA persist in samples?
- How long does prey DNA survive in unpreserved samples?
- Are the relative amounts of DNA recovered from prey species proportional to their mass in the diet?[2]
- What is the quality of the prey DNA recovered?[3]

Chapter 4 contains a detailed analysis of the amount of prey DNA present in the sea lion faecal samples. Using quantitative real-time PCR (qPCR), I further investigated results from the previous chapter that showed (based on analysis of PCR clone libraries) that the proportions of fish DNA in faeces from captive sea lions were roughly proportional to the mass of the prey items consumed. The use of qPCR allowed more samples to be analysed and provided an independent method of measuring the amount of prey DNA present in the samples.

Chapter 5 focuses on measuring the quality of DNA recovered in faecal samples. The concept of DNA quality is often discussed, and usually refers to the level of DNA degradation in a sample – but there is no way of measuring this in many situations. I present a general method to quantify DNA damage present in specific DNA sequences. The approach is applied to determine the quality of DNA originating

[2] This question is further evaluated in the next data chapter.
[3] This question is considered in a separate chapter.

from predator (sea lion) and prey (herring) in the faecal samples collected during the captive feeding trial.

Chapter 6 presents results from a field-based study on the diet of macaroni penguins (*Eudyptes chrysolophus*). Dietary information was collected using DNA-based analysis of faecal material as well as through conventional stomach content analysis. Comparisons are made between the traditional and genetic data, and future prospects for using a DNA-based approach to study seabird diet are discussed.

The thesis concludes with a review of some concurrently published DNA-based diet studies on pinnipeds, a general discussion and suggestions for future directions of research in this field. The appendices include: (I) an evaluation of the strengths and weaknesses of nested-PCR; (II) details of PCR primers used in the thesis; (III) summary of qPCR data collected in Chapter 4; (IV) detailed results from the macaroni penguin faecal DNA clone library analysis (Chapter 6).

For chapters that have been published (or submitted) citation and co-authorship details are included at the end of this section. In all cases, I was the senior author, and conducted the laboratory work, analysis of data and writing of the papers. My co-authors contributed significantly to the initiation of the projects, sample collection, advice on analysis, and/or by critically reviewing the papers in preparation for publication. The published papers have been modified to integrate the chapters into this thesis; however, each chapter is essentially self-contained and there may be some repetition in content throughout the thesis.

1.6 Details of publications resulting from thesis

Chapter 2:

Deagle BE, Jarman SN, Pemberton D and Gales NJ (2005) Genetic screening for prey in the gut contents from a giant squid (*Architeuthis* sp.). *Journal of Heredity*, **96**, 417-423.

Chapter 3:

Deagle BE, Tollit DJ, Jarman SN, Hindell MA, Trites AW and Gales NJ (2005) Molecular scatology as a tool to study diet: analysis of prey DNA in scats from captive Steller sea lions. *Molecular Ecology*, **14**, 1831-1842.

Chapter 4:

Deagle BE and Tollit DJ (2006) Quantitative analysis of prey DNA in pinniped faeces: potential to estimate diet composition? *Conservation Genetics*, In press

Chapter 5:

Deagle BE, Eveson JP and Jarman SN (submitted) Quantification of damage in DNA recovered from highly degraded samples – a case study on DNA in faeces.

Chapter 6:

This chapter has not yet been submitted as a paper.

Additional publications containing results from thesis**:**

Jarman SN, Deagle BE and Gales NJ (2004) Group-specific polymerase chain reaction for DNA-based analysis of species diversity and identity in dietary samples. *Molecular Ecology*, **13**, 1313-1322. (Chapter 2)

Tollit DJ, Heaslip S, Deagle BE, Iverson SJ, Joy R, Rosen DAS and Trites AW (2006) Estimating Diet Composition in Sea Lions: What Technique to Choose? In: *Sea lions of the world* (eds. Trites AW, Atkinson S, DeMaster D, et al.). Alaska Sea Grant College Program, University of Alaska Fairbanks. (Chapter 3)

Chapter 2

Genetic screening for prey in the gut contents from a Giant Squid

An ocean without its unnamed monsters would be like a completely dreamless sleep.

— JOHN STEINBECK & EDWARD RICKETTS,
Sea of Cortez

Abstract

Information on the diet of giant squids (*Architeuthis* spp.) is scarce because these animals are rarely taken from their deep-sea habitat and very few specimens have morphologically recognisable remains in their digestive tracts. In this chapter, I explored the use of PCR-based methods for detection of DNA in the prey remains and amorphous slurry from an *Architeuthis* gut sample. Primers with conserved binding sites were designed to amplify a region of the 16S mtDNA from potential prey species. The amplified sequence varied in size allowing separation of fragments from fish (~ 255 bp in size) and squid (~ 180 bp in size). Prey DNA could be amplified from prey remains as well as from the amorphous stomach content slurry. Sequence comparisons identified fish prey as blue grenadier (*Macruronus novaezelandiae*). Isolation of *Architeuthis* DNA from an ingested tentacle and the presence of chitin fragments indicate cannibalism may occur in giant squid. Denaturing gradient gel electrophoresis was used to screen for less common DNA types, revealing a relatively high frequency of PCR generated false alleles but no additional prey species. A limitation of using universal primers to screen the gut sample was the prevalence of DNA from the predator (78 out of 80 clones initially screened came from *Architeuthis*). The application of a chordate-specific primer set allowed blue grenadier DNA to be detected without screening large numbers of clones. This finding suggests group-specific primers that exclude predator DNA will be useful in future DNA-based diet studies.

2.1 Introduction

The giant squids (*Architeuthis* spp.) have long captured the public's imagination because of the rarity of specimens, their enormous size and their existence in an alien habitat. *Architeuthis* squid have been found over a huge geographic range in the Pacific, Atlantic, Indian and Southern Oceans (Clarke 1966; Ellis 1995) and are common enough to form a significant part of sperm whales caloric intake in many areas (Clarke & MacLeod 1982; Clarke & Young 1998). Despite their apparent ubiquity very little scientific data has been collected on the species and only recently the first live giant squid was observed in the wild (Kubodera & Mori 2005). The taxonomy of the group is in a state of confusion due to the sporadic nature (both temporally and spatially) of collection and the poor quality of most specimens (Förch 1998). Ecological data are non-existent with the exception of some inferences which have been drawn from physiological and morphological observations (Brix 1983; Norman & Lu 1997; Lordan *et al.* 1998).

One of the most fundamental pieces of information needed to understand a species' biology and role in an ecosystem is knowledge of its diet. Because so little is known about giant squid, diet data is particularly valuable, and can shed light on topics as diverse as this species habitat preference and swimming ability (Lordan *et al.* 1998). Of the common methods for determination of diet (direct observation, gut content analysis and fecal analysis), only identification of prey remains in the digestive tract is feasible for these squid. Some data has been collected by morphological gut content analysis of specimens collected through occasional landing or chance stranding. However, these squid are usually in poor condition and the gut is often "empty" with no morphologically recognisable content (Förch 1998). Even if material is present, classification of the remains based on morphological features is notoriously difficult due to the squid tendency to tear apart and finely macerate prey items (Kear 1992). This has resulted in many scales, bones and lumps of flesh found in giant squid digestive tracts being reported as unidentifiable (Förch 1998; Lordan *et al.* 1998; Bolstad & O'Shea 2004). The limited information published on *Architeuthis* diet indicate fish and cephalopods are their most important prey, with crustacean remains occasionally being observed (Förch 1998; Lordan *et al.* 1998). This is consistent with findings from studies of other large squid (Phillips *et al.* 2003).

The difficulties associated with diet determination through visual identification of squid gut contents has led to the use of other methods such as fatty acid analysis (e.g. Phillips *et al.* 2001) and the identification of prey remains using immunological approaches (e.g. Kear 1992). The use of DNA identification techniques seems like a logical approach to help identify decomposed species in squid gut contents. PCR-based methods have been used to detect specific species of larval fish from predatory fish stomachs (Rosel & Kocher 2002), and to identify species of krill flushed from stomachs of Adelie penguins (Jarman *et al.* 2002). Both the slurry of digested material in the stomach and isolated prey remains may contain DNA which could be extracted and amplified using PCR. Three general approaches could be taken to amplify DNA from these sources: (i) PCR primers could be used to detect specific species (e.g. Rosel & Kocher 2002) (ii) PCR primers could amplify a specific group of prey items (e.g. Jarman *et al*. 2002) or (iii) "universal" primers could be used to amplify DNA from a broad range of unspecified prey items (e.g. Rollo *et al.* 2002). In the latter two approaches the amplified DNA would need to be identified through a phylogenetic-based approach. Since the data available on the diet of giant squid is so incomplete, targeting specific species is not appropriate. A group-based approach could be useful, especially as a method of excluding the predator's DNA from downstream analysis. A limitation in this particular case is that giant squid are known to feed on other squid, so the predator falls within a group of potential prey species. Amplification of DNA from a broad range of species could potentially identify all prey DNA present (including unexpected prey items). However, DNA from the predator, parasites and other non-prey species may be present in the gut, therefore relatively large numbers of DNA molecules may need to be characterised to isolate prey sequences.

In this chapter, I use PCR-based methods to determine the prey species present in the gut of a giant squid collected in Tasmania, Australia. The DNA was extracted from both amorphous gut material and isolated prey remains. I used primers which target conserved primer binding sites as well as group-specific primers which amplify only chordate DNA. DNA sequences obtained were compared with those available in the public data base and a sequence similarity approach was applied to identify prey species. Denaturing gradient gel electrophoresis (DGGE) was used to screen amplified DNA fragments in order to check for DNA molecules present at low level in the amplified mixture. The methodology presented will provide a framework for

future studies, considerably increasing the potential for diet data collection from scarce specimens of giant squid and other rare marine animals such as beaked whales.

2.2 Methods

2.2.1 Architeuthis Sample

The *Architeuthis* specimen was caught on June 14 1999 by a trawler fishing for blue grenadier at a depth of 500-700 meters off the west coast of Tasmania. The squid was a male weighing 190 kg, it was frozen on board the boat and kept frozen in storage until dissection in September 2002. It was opened along the ventral surface by cutting the mantle cavity via a longitudinal incision from the funnel to the rear. The posterior viscera were exposed and the oesophagus traced through to the caecum, stomach, intestine and rectum region. The gut contents were removed from the base of the oesophagus, stomach and caecum then stored in 95% ethanol. Only the knowledge that the sample was from the gut of a local *Architeuthis* specimen was provided to me before the end of the genetic analysis (no information on location/method of capture or potential prey species).

2.2.2 DNA extraction

DNA was extracted from amorphous particles and recognizable prey remains which were isolated under a dissection microscope. The settled volume of the amorphous particles was approximately 1 L. The isolated prey remains included: (a) > 50 scales, all of similar size and shape (b) three small (5-10 mm) tentacle fragments[4] (c) one bone fragment (d) 12 chitinous squid beak fragments (< 5 mm). For DNA extraction, the amorphous particles were resuspended and samples of 2 mL were centrifuged for five min – yielding approximately 200 mg of pelleted material. I extracted DNA from two pellets in independent procedures. In a third set of extractions DNA was extracted from individual scales (n = 10), tentacle fragments (n = 2) and the bone fragment. All extractions were done using the

[4] The third tentacle sample is in storage at the Tasmanian Museum and Art Gallery.

hexadecyltrimethylammonium bromide (CTAB) method (Doyle & Doyle 1987). In this procedure all steps were carried out in 1.5 mL tubes and centrifugations were at 14000 rpm in a microcentrifuge. The samples were homogenised in 175 μL CTAB buffer (2% CTAB; 100 mM Tris HCl, pH 8.0; 1.4 M NaCl; 20 mM EDTA; 0.2% 2-mercaptoethanol), 2 μL Proteinase K (20 mg/mL) was added and samples were incubated at 65ºC for one h. Sequential extractions were done with 150 μL of chloroform:isoamyl alcohol (24:1), 150 μL of phenol:chloroform:isoamyl alcohol (25:24:1) and 150 μL of chloroform:isoamyl alcohol (24:1), with five min centrifugations between steps. The DNA was precipitated by addition of 150 μL isopropanol (-20ºC) and pelleted by a 20 min centrifugation. The DNA was washed with 400 μL of 70% ethanol, centrifuged for 5 min, air dried and resuspended in 30 μL of distilled water. The concentration of purified DNA was determined using a PicoFluor fluorometer (Turner Designs). Near the end of the study, DNA was also extracted from fish tissue obtained from a local fish market using the method outlined above.

2.2.3 Primer design

Conserved PCR primers have previously been developed which amplify segments of mtDNA from a broad range of animal taxa (Palumbi 1996). These primers are often used in phylogenetic studies and in order to provide a suitable amount of sequence data the size of the products are generally >500 bp. Since DNA from the squid gut was likely to be degraded, I wanted to amplify a shorter fragment (~200 bp) and therefore designed a new primer pair. The 3' end of the mitochondrial ribosomal 16S gene (flanked by conserved primers 16Sar-5' and 16Sbr-3', Palumbi 1996)[5] was chosen as a potential target. This region has been widely characterized, providing a large dataset to help identify unknown sequences. Sequences were obtained from GenBank for a taxonomically diverse group of 30 fish (Osteichthyes and Chondrichthyes), 30 cephalopods and several crustaceans. These were aligned and suitable primers selected (see Table 2.1 for primer sequences and representative alignments). The primer binding region is highly conserved in cephalopods and fish with slightly more variation found in crustaceans. It is possible to design degenerate

[5] The 16Sbr-3' primer was inadvertently refered to as 16Sa-3' in Deagle *et al*. 2005a.

primers which would have incorporated the small amounts of variation that was observed. However, potential incorporation of mismatches in PCR products may have confused interpretation of the DGGE analysis and the mismatches seemed unlikely to significantly affect primer binding, so I chose non-degenerate primers based on the squid sequence. The amplified region is short and variable, maximizing the likelihood of amplification and of obtaining informative sequence data. An additional useful feature of the amplified region is that it varies in length within and between the major taxa targeted. Based on the complete set of species used in primer design, the size of the this fragment is 258 bp ± 8.4 SD in fish and 180 bp ± 25.9 SD in cephalopods, with no overlap identified between these groups (Figure 2.1). The size in crustaceans overlaps that of the squid (Table 2.1).

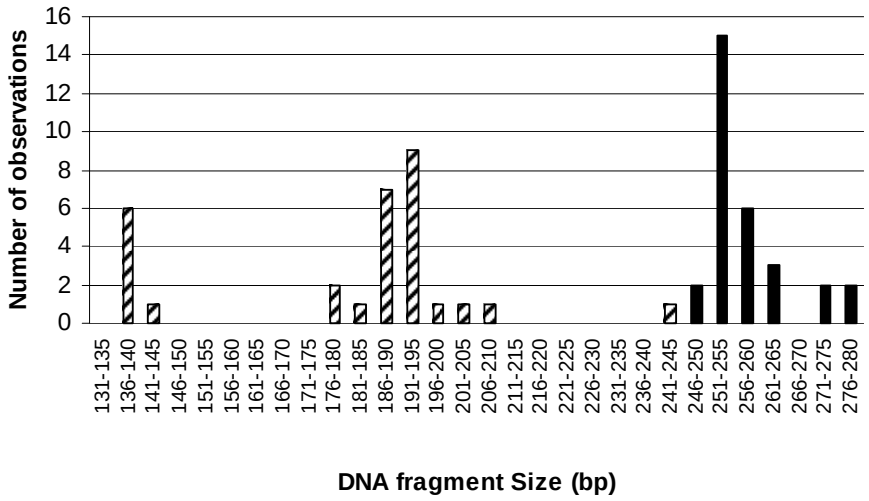


Figure 2.1 *Distribution of fragment sizes of the PCR products which would be amplified from a taxonomically diverse group of 30 fish (Osteichthyes and Chondrichthyes) (black bars) and 30 cephalopods (grey bars) using 16S mtDNA primers (16S1F and 16S2R). Sizes are based on sequence data taken from GenBank.*

In addition to the conserved primers, I used primers which specifically amplify DNA from chordates to obtain sequences from the squid gut. These primers take advantage of the unique arrangement of mitochondrial genes in the mitochondrial genome of Chordata and amplify parts of the two mitochondrial rDNA genes and the intervening valine tRNA gene (Jarman *et al.* 2004).

Table 2.1 *Primers used in the current chapter aligned with homologous sequences from representative target taxa.*

Phylum	Class	Family	Genus + species	Accession #	16S1F (5' → 3') GGACGAGAAGACCCT	16S2R (5' → 3') CGCTGTTATCCCTATGGTAACT	Product Size (bp)
Chordata	Mammalia	Physeteridae (sperm whales)	*Physeter macrocephalus*	AJ277029	A----------	----------G-------	228
	Elasmobranchii	Squalidae (dogfish sharks)	*Squalus acanthia*	Y18134	A----------	----------G-------	261
		Rajidae (skates)	*Raja radiata*	AF106038	A----------	----------G-------	261
	Actinopterygii	Amiidae (bowfins)	*Amia calva*	AY442347	A----------	----------G-------	280
		Congridae (conger eels)	*Conger myriaster*	AB038381	A----------	----------G-------	271
		Clupeidae (herrings, sardines)	*Sardinops melanostictus*	AB032554	A----------	----------G-------	249
		Cyprinidae (minnows, carps)	*Cyprinus carpio*	X61010	A----------	----------G-------	255
		Salmonidae (salmonids)	*Oncorhynchus mykiss*	L29771	A----------	----------G-------	259
		Neoscopelidae (lanternfishes)	*Neoscopelus microchir*	AP002921	A----------	----------G-------	254
		Gadidae (cods, haddocks)	*Gadus morhua*	X99772	A----------	----------G-------	254
		Berycidae (alfonsinos)	*Beryx splendens*	AP002939	A----------	----------G-------	255
		Zeidae (dories)	*Zenopsis nebulosus*	AP002942	A----------	----------G-------	251
		Gasterosteidae (sticklebacks)	*Gasterosteus aculeatus*	AP002944	A----------	----------G-------	251
		Sparidae (porgies)	*Pagrus major*	AP002949	A----------	----------G-------	254
Mollusca	Cephalopoda	Architeuthidae (giant squids)	*Architeuthis dux*	AY377629	----------	----------	190
		Onychoteuthidae (hook squids)	*Moroteuthis ingens*	X79580	----------	----------	190
		Loliginidae (squids)	*Loligo bleekeri*	AB009838	----------	----------	183
		Loliginidae (squids)	*Sepioteuthis lessoniana*	AY131035	----------	----------	186
		Ommastrephidae (squids)	*Nototodarus gouldi*	AY380810	----------	----------	191
		Ommastrephidae (squids)	*Todarodes pacificus*	AB158364	----------	----------	191
		Octopodidae (octopus)	*Octopus vulgaris*	AJ390312	------A-----	----------	203
		Octopodidae (octopus)	*Hapalochlaena maculosa*	AY545107	----------	----------	189
		Sepiidae (cuttlefishes)	*Sepia pharaonis*	AF369117	----------	----------	184
		Sepiidae (cuttlefishes)	*Sepiella maindroni*	AF369959	----------	----------	191
Arthropoda	Malacostraca	Euphausiidae (krills)	*Euphausia superba*	AB084378	T-----T-------	----------AA------	200
		Euphausiidae (krills)	*Nyctiphanes australis*	AF177181	T-----T-------	----------AA------	202
		Penaeidae (penaeid shrimps)	*Penaeus monodon*	AF217843	------T-------	----------AA------	207
		Penaeidae (penaeid shrimps)	*Xiphopenaeus kroyeri*	AF192093	------T-------	----------AA------	207
		Palinuridae (spiny lobsters)	*Jasus edwardsii*	AF337979	------T-------	----------AA------	210

2.2.4 PCR amplification and cloning

Standard PCR reactions were performed on 2 μL of template in a 25 μL volume containing 0.4 μM of each primer, 0.125 mM dNTPs, 2.0 mM $MgCl_2$, 1x AmpliTaq Gold® buffer and 0.625 units AmpliTaq Gold® thermostable DNA polymerase (Applied Biosystems). Thermal cycling conditions for both primer sets were as follows: 94°C for 10 min then 35 cycles (94°C, 30s / 55°C, 30s / 72°C, 45s) followed by 72°C for 2 min. Samples were separated on a 2.0% agarose gel. PCR products were cloned into the pCR®2.1-TOPO vector (Invitrogen). Positive transformants were identified using blue/white colour selection, insert size was checked by digestion with *Eco*RI and separation of bands on a 2.0% agarose gel.

2.2.5 Sequence analysis

Sequencing reactions were carried out with the Big Dye prism dideoxy sequencing dye terminator kit (Applied Biosystems). Electrophoresis was performed on an Applied Biosystems 377 automated DNA sequencer. DNA sequences where compared with publicly available sequences in GeneBank using BLAST search (Altschul *et al.* 1990). Sequence data were aligned using CLUSTAL_X (Thompson *et al.* 1997). To determine sequence relatedness I used the neighbor-joining algorithm (Saitou & Nei 1987) in MEGA version 2.01 (Kumar *et al.* 2001) based on distances calculated using Kimura's two-parameter model (Kimura 1980). All nucleotides transitions and transversions were included in the analysis; nucleotide positions containing insertions/deletions in the alignment were excluded.

2.2.6 DGGE analysis

In order to identify additional sequence variants in the 16S PCR products I used DGGE, a technique which can separate variable DNA sequences (Middleton *et al.* 2004). Separation is accomplished by electrophoresis of the DNA fragments in a polyacrylamide gel containing a gradient with an increasing concentration of denaturants. The mobility of the fragments is determined by their melting behaviour

as they denature and this is highly sequence dependent (Myers *et al.* 1987). Theoretical melting profiles were constructed by using the program MELT94, which can be found on the Internet at http://web.mit.edu/osp/www/melt.html (Figure 2.2). DGGE was performed using the DCode™ system (Bio-Rad, Hercules CA). Acrylamide gels (7.5%) were poured using a Model 475 Gradient Delivery System (Bio-Rad, Hercules CA) and run at 56ºC. For samples separated by DGGE, the16S1R primer was redesigned to incorporate a GC clamp (gggcgggggcggcgggacgggcgcggggcgcggc gggcg-CGCTGTTATCCCTATGGTAACT, Sheffield *et al.* 1989), the annealing temperature was lowered to 50 ºC and other conditions were the same as in the standard PCR. Template was 25 ng genomic DNA or 1µL of a 1:100 dilution of unclamped PCR product or plasmid DNA. Electrophoretic conditions (gradient range, voltage and length of run) which resulted in clear band separation were determined by experimenting with several different species of fish and squid.

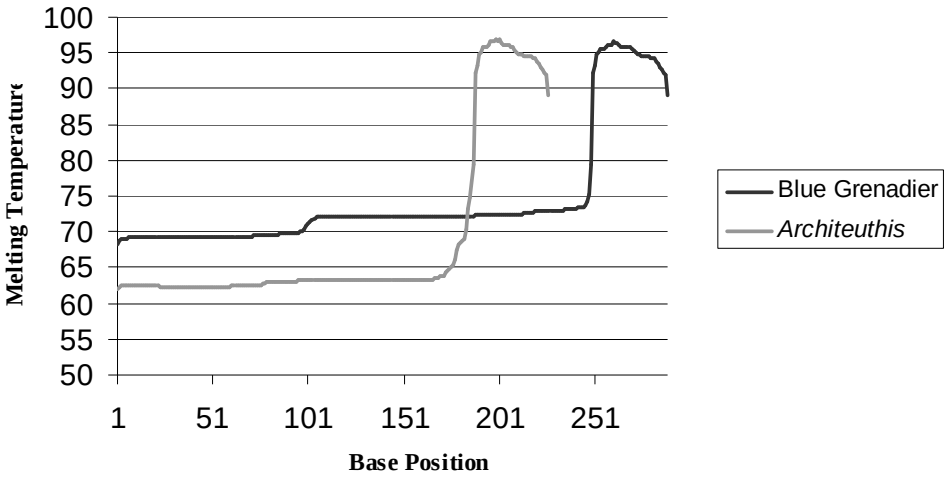


Figure 2.2 *Theoretical melt maps for 16S mtDNA fragments amplified from Blue Grenadier (291 bp) and Architeuthis (229 bp) template with a 3' GC clamped primer (constructed by using the program MELT94). The lower melting temperature of the Architeuthis fragment is typical for squid due to lower GC content; this is expected to result in lower mobility of squid fragments in a DGGE gel.*

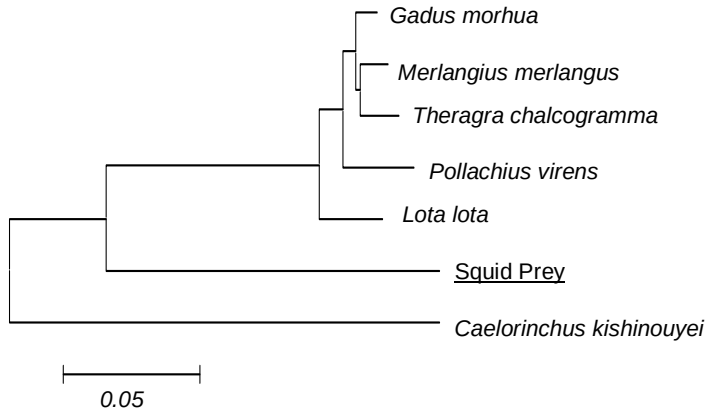
2.3 Results

2.3.1 Conserved 16S primers

The concentration of DNA purified from the two amorphous squid gut samples were 12 ng/μL and 8 ng/μL. PCR amplification of DNA extracted from both gave a strong "squid sized" band (~190 bp) and a much weaker "fish sized" band (~250 bp). The clones obtained from these amplifications contained two insert sizes corresponding to the two PCR bands observed (78 small and 2 large). In order to increase the number of clones with the larger insert a secondary band-stab PCR (Bjourson & Cooper 1992) was carried using the faint large band from the first PCR as template. Clones produced from the band-stab PCR gave almost equal proportions of the two inserts (26 small and 24 large).

Sequencing was initially carried out on seven clones containing short inserts and six clones with the longer insert. The short insert clones gave sequences 189 bp in length, six were identical and the seventh differed from these by a single nucleotide substitution. The longer insert clones gave sequences 252 bp in length; again they were identical except for a single nucleotide substitution in one sequence. The consensus sequences were compared with entries in GenBank. In the case of the small fragment (GeneBank accession AY392149), the sequence exactly matched mitochondrial 16S sequences from *Architeuthis dux*. The next closes matches were from a variety of squid species all with >25 nucleotide differences over the region. The longer DNA sequence (GeneBank accession AY392146) matched most closely with mitochondrial 16S sequences from fish species in the order Gadiformes, with four of the top five matches being within the family Gadidae. An identity matrix (giving the proportion of identical residues between sequences) shows that the longer DNA sequence and the top five BLAST matches are about 80% identical, with none being a likely species match. Based on this information I obtained tissue samples from two local Gadiformes (*Moro moro* and *Macruronus novaezelandiae*) and one species belonging to the sister order Zeiformes (*Cyttus traversi*). DNA was extracted from these species and the 16S fragment amplified and sequenced (GenBank accession AY392146-48). Analysis of these sequences showed a perfect match between the unknown sequence from the squid gut and *M. novaezelandiae* (Figure 2.3).

(a)



(b)

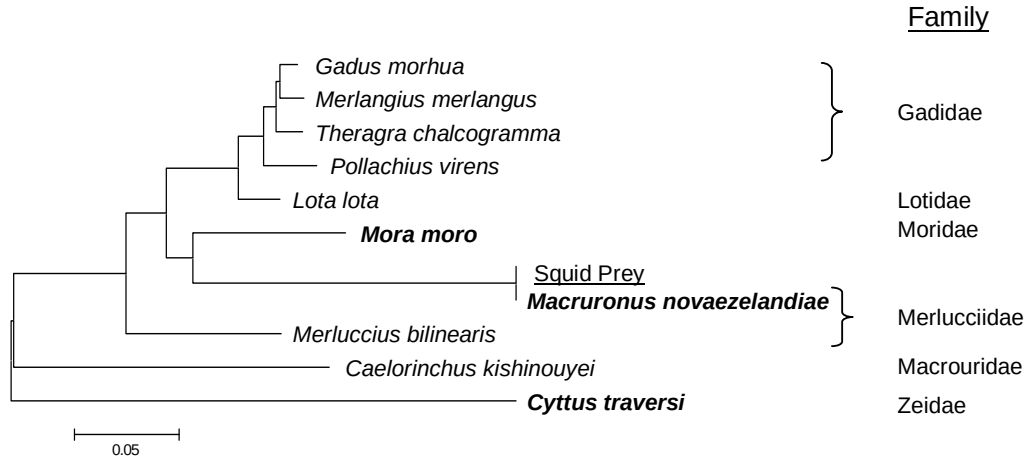


Figure 2.3 *Neighbor-joining tree based on:* ***(a)*** *the 252 bp mtDNA sequence obtained from the Architeuthis gut sample aligned with: sequences of the five closest Blast matches (G. morhua, M. merlangus, T. chalcogramma, P. virens and L. lota) and the sequence of a related fish belonging to a genus previously identified as Architeuthis prey (C. kishinouyei)* ***(b)*** *Additional sequences obtained from fish species during the present study (**bold**), and an additional GenBank sequence from the family merlucciidae (M. bilinearis). All of the species belong to the order Gadiformes with the exception of C. traversi which belongs to the sister order Zeiformes.*

The concentration of DNA purified from the scales, bone fragment and tentacle fragments was below the level measurable using our fluorometer (< 5 ng/μL). The amplification of DNA from the 10 scales resulted in six samples producing fish sized bands and weak squid sized bands. The remaining four scale samples gave weak or no obvious fish sized bands. The bone fragment and tentacle fragments produced only fish or squid sized bands respectively. Cloning and sequencing of these PCR products revealed all sequences matched either *M. novaezelandiae* (blue grenadier) or *Architeuthis* sequences previously obtained. While the bone and tentacles gave only blue grenadier or *Architeuthis* sequences, a mixture of sequences was obtained from scales indicating *Architeuthis* DNA present in the gut was associated with the fish scales. The intensity of fish sized bands from the scales and the bone fragment were much stronger than the fish bands observed in the amorphous PCR products, indicating that the vast majority of blue grenadier sequences obtained from these samples originated from the scales or bone.

2.3.2 Chordate primers

The major limitation in the use of conserved primers to screen the gut sample was the prevalence of DNA from the predator (78 out of 80 clones initially screened were squid sized and came from *Architeuthis*, see below). The chordate specific primer set will only identify a sub-set of potential prey species, but all sequences obtained should be from prey rather than the predator. Using this primer set I obtained amplification products from the amorphous component, the bone fragment and 6 out of the 10 scales. All of these PCR products were sequenced and gave identical sequences 312 bp long. This sequence matched perfectly with sequence of this region obtained from a blue grenadier tissue sample.

2.3.3 DGGE analysis

Analysis of theoretical melting profiles for fish and squid DNA fragments amplified by the conserved primers revealed significant differences between these groups due to differences in guanine and cytosine (GC) content (Figure 2.2). Not surprisingly then, DGGE conditions which resulted in good separation of fragments

was different for fish and squid: for the fish surveyed conditions were a 30-70% gradient at 50 V for 8 h on a 16 cm gel; for squid fragments best separation was achieved with a 0-50% gradient at 60 V for 8 h on a 16 cm gel (Figure 2.4). The analysis of the PCR amplifications from the amorphous component and the band-stab gave two bands corresponding to *Architeuthis* and blue grenadier bands (Figure 2.4: lane 9).

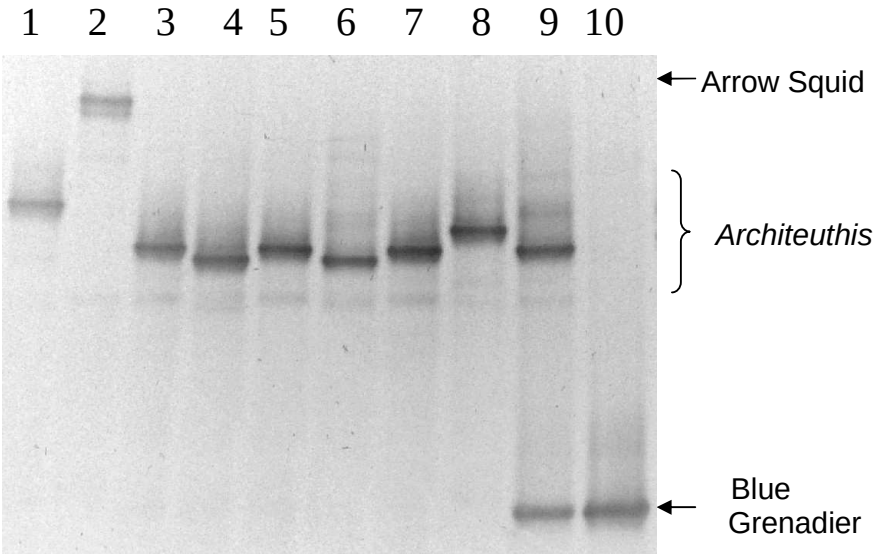


Figure 2.4 *Denaturing gradient gel electrophoresis separation of mtDNA 16S PCR products. Lanes 2 + 10 are amplified from genomic DNA of arrow squid (Nototodarus sp.) and blue grenadier (Macruronus novaezelandiae) respectively. Lane 9 is an amplification of DNA extracted from the amorphous slurry component of the Architeuthis gut contents, remaining lanes are amplified from clones derived from the same source. Sequences of the clones shown either match the Architeuthis consensus (lanes 3, 5, 7) or are closely related (lane 1, variant H; lane 4, variant B; lane 6, variant D; lane 8, variant F. See Table 2.2 for sequences).*

While these results indicate that the majority of DNA present in the sample comes from these two species there is potential for less abundant DNA sequences to be present. To check this possibility 26 fish sized and 80 squid sized clones were amplified and screened for sequence variation. The analysis of the fish sized clones showed 23 samples matched the electrophoretic mobility of the common blue grenadier sequence (GenBank accession AY392146); the three remaining clones had unique DGGE bands. Sequencing of these clones revealed one was the single base

pair variant of the blue grenadier sequence which had been previously identified and the two other sequences were unique, but differed by only 1 or 2 base substitutions from the common blue grenadier sequence (Table 2.2). Screening of 80 squid sized clones revealed 70 clones running parallel to the common *Architeuthis* sequence (GenBank accession AY392149) and 10 not matching the reference sequence (Figure 2.4). The 10 variant clones were sequenced revealing eight different sequences all closely related to the previously obtained *Architeuthis* sequences (Table 2.2).

Table 2.2 *Variable sites identified in nucleotide sequences obtained from amorphous slurry component of the Architeuthis gut contents.*

	Position of variable sites								# of sequences	Frequency from DGGE
Architeuthis										
	17	29	48	114	115	116	146	159		
Consensus	T	T	T	C	C	T	C	C	6	70/80
Variant A	C	.	.	.	.	.	.	.	1	1/80*
B	.	.	C	.	.	.	.	.	1	1/80*
C	.	C	.	.	.	.	.	.	1	1/80*
D	.	.	.	.	.	C	.	.	1	1/80*
E	.	.	.	.	.	.	.	T	2	2/80
F	.	.	.	.	T	.	.	T	2	2/80
G	.	.	.	T	T	.	.	T	1	1/80
H	.	.	.	T	T	.	T	T	1	1/80
Blue Grenadier										
	16	54	81							
Consensus	G	C	T						5	23/26
Variant A	.	T	.						1	1/26
B	.	.	G						1	1/26
C	A	.	G						1	1/26

*these variants were not separable from each other under the DGGE conditions used

It is interesting to note that change in the mobility of the sequence variants during DGGE corresponds well with predicted shifts. In the *Architeuthis* sequences multiple C-T transitions resulted in a stepwise decrease in mobility of the fragment due to progressive lowering of the fragments denaturation temperature (Figure 2.4, Table 2.2). Similarly T-C transitions resulted in increased mobility of the amplified fragment. One of the weaknesses of DGGE is that some sequence variants will not be detected since they do not affect the fragments denaturation temperature (for example A-T or C-G transversions). This feature may have resulted in an underestimation of

the total number of sequences present, but it is unlikely DNA from new species (expected to contain several base pair substitutions) would have been missed.

2.4 Discussion

This chapter reports on the development and application of genetic tools for the identification of prey remains recovered from *Architeuthis* gut contents. Primers were designed which amplify a conserved region of 16S mtDNA that differs in size between fish and squid, allowing separation of DNA recovered from these potential prey groups. The analysis of an *Architeuthis* gut sample revealed both fish and squid sized PCR products. These PCR products were screened for sequence variants (i.e. different species of fish or squid) using DGGE. Additional sequences were recovered from PCR amplifications using a chordate specific primer set.

Fish DNA was amplified from scales, bones and the slurry component of the *Architeuthis* gut sample using both primer sets. These sequences were initially characterised (based on publicly available sequence data) as belonging to a single Gadiforme species, sequencing of local Gadiformes allowed us to identify the prey species as blue grenadier (called hoki in New Zealand). This fish species occurs in waters of southern Australia and around New Zealand, and they are a dominant component of the upper continental slope fish fauna around Tasmania (May & Blaber 1989). The capture of *Architeuthis* by several commercial trawlers targeting blue grenadier in Australia (D. Pemberton, unpublished data) and New Zealand (Bolstad & O'Shea 2004) has suggested these fish form a component of the diet of *Architeuthis*; however, blue grenadier had not previously been recorded in gut contents of *Architeuthis*. The absence of direct evidence for this link had led previous workers to conclude that *Architeuthis* probably preys on the same food items as blue grenadier rather than the blue grenadier itself (Bolstad & O'Shea 2004). It is interesting to note that the three *Architeuthis* specimens found stranded in southern Tasmania (first in 1986, then again in 1992 and 2002) were all found between June and early September – which is the same time of year that blue grenadier from Australian waters gather to spawn in dense aggregations off western Tasmania (Gunn *et al.* 1989).

The squid DNA sequences that I isolated from the giant squid gut closely matched *Architeuthis* and the majority of these sequences are likely to have originated

from the gut lining of the predator. Detection of cannibalism using DNA-based methods is possible, but only through the development and use of individual specific DNA markers. Using only information from more conserved markers, such as the ones used in the current chapter, it is not possible to differentiate between predator and prey of the same species. However, the ability to identify morphologically ambiguous tissue fragments is a strong point of the genetic identification approach and amplification of DNA extracted from the small tentacle fragments found in the squid produced only *Architeuthis* sequences. This finding suggests that cannibalism has occurred, a conclusion further supported by the presence of crushed squid beak in the gut and the lack of any DNA from different squid species in the 80 clones which were screened using DGGE. Cannibalism has been widely reported in other squid (Santos *et al.* 1997; Quetglas *et al.* 1999; Phillips *et al.* 2003) and has recently been described in *Architeuthis* (Bolstad & O'Shea 2004). It should be noted that autophagy or accidental self-ingestion cannot be ruled out as a potential source of the tentacle fragments (see discussion in Bolstad & O'Shea 2004).

Genetic identification from amorphous gut material is appealing since data collection is not limited to undigested tissue and hard part remains. Since DNA from several species of prey may be present in this mixture, heterogeneous amplification products must be separated for identification. To identify different 16S PCR products I took advantage of the size differences in the amplification products and also applied DGGE (Muyzer 1999). Direct DGGE analysis of PCR products from the squid gut identified both the blue grenadier and *Architeuthis* amplification products. Since rare amplification products are likely to be hard to detect using direct PCR, I also screened individual clones derived from these PCR products. In the 80 squid sized and 26 fish sized clones analysed no new prey species were identified. However, this analysis did detect multiple sequences closely matching *Architeuthis* and blue grenadier. Possible origins of these sequences include: heteroplasmy, amplification from multiple genetically different individuals or PCR-induced mutations resulting from the amplification of degraded DNA. These possibilities are not mutually exclusive so it is difficult to discount any completely, however several facts indicate the majority of these sequences are PCR artefacts. First of all most of the changes are C-T transitions which is consistent with *Taq* polymerase errors generated from damaged template through cytosine deamination and jumping PCR (Hofreiter *et al.* 2001). Secondly, all alleles are separated from the next closest allele by a single nucleotide substitution

suggesting *in situ* generation. Finally the nucleotide substitutions are inconsistent with patterns of conserved versus variable sites observed in closely related species (5 of the 7 substitutions in the *Architeuthis* sequence occur in sites which are conserved among the 30 other species of cephalopod surveyed for primer design).

In the DGGE bands amplified directly from squid stomach DNA template I did not observe artefacts formed by PCR induced errors (Figure 2.4). This suggests that the ratio of undamaged to damaged template is high enough to ensure that the damaged templates are indistinguishable through direct PCR analysis. Through cloning of individual molecules I confirmed that the majority were undamaged; however, the frequency of false alleles was high enough to interfere with the screening for prey species represented by a low frequency of the DNA. It may be possible to lower the background level of false alleles through the use of a polymerase possessing 3'-5' exonuclease activity (proofreading) or by treatment of the DNA extraction with uracil N-glycolase (Hofreiter *et al.* 2001). Another way to detect prey DNA representing a small proportion of total extracted DNA is the development of group specific primers which exclude DNA from the predator and/or amplify only a portion of potential prey (Jarman *et al.* 2004). This approach is exemplified by the single-step isolation of blue grenadier DNA in this study, through the use of chordate-specific primers.

The scarcity of *Architeuthis* specimens necessitates a detailed analysis of each one if our knowledge of this species is to increase substantially. The use of DNA-based methods to study diet allows identification of prey recovered from gut contents – including prey remains which could not be identified using morphological methods. The universality of genetic methods could also allow a standard protocol of gut content analysis to be developed, maximizing information gain from sporadically collected samples. One of the factors currently limiting the use of this approach is that prey identification relies on DNA sequence data being available for a wide range of potential prey species. With the rapid increase in available DNA sequence data (e.g. Miya *et al.* 2003) and development of taxonomic systems based on DNA sequences (Hebert *et al.* 2003; Ross *et al.* 2003; Tautz *et al.* 2003), genus or species identification of DNA sequences should become increasingly possible.

Chapter 3

Analysis of prey DNA in faeces of captive Steller sea lions
Part I: DNA detection, distribution and signal persistence

In my village there is no shame in being named for the dung goddess… Dung, that fertilizes and causes the crops to grow! Dung, which is pattied into chapatti-like cakes when still fresh and moist, and is sold to the village builders, who use it to secure and strengthen the walls of Kachcha buildings made of mud. Dung, whose arrival from the nether end of cattle goes a long way towards explaining their divine and sacred status! Oh, yes, I was wrong, I admit I was prejudiced, no doubt because its unfortunate odours do have a way of offending my sensitive nose – how wonderful, how ineffably lovely it must be to be named for the Purveyor of Dung.

— SALMAN RUSHDIE,
Midnight's Children

Abstract

In this chapter, I presented results from a captive feeding trial carried out to test whether prey DNA could be reliably detected in scat samples from Steller sea lions (*Eumetopias jubatus*). Two sea lions were fed a diet of fish (five species) and squid (one species), and DNA was extracted from the soft component of collected scats. Most of the DNA obtained came from the predator, but prey DNA could be amplified using prey-specific primers. The four prey species fed in consistent daily proportions throughout the trial were detected in more than 90% of the scat DNA extractions. Squid and salmon, which were fed as a relatively small percentage of the daily diet, were detected as reliably as the more abundant diet items. Prey detection was erratic in scats when the daily diet was fed in two meals that differed in prey composition, suggesting that prey DNA is passed in meal specific pulses. Prey items that were removed from the diet following one day of feeding were only detected in scats collected within 48 hours of ingestion. Proportions of fish DNA present in eight scat samples (evaluated through the screening of clone libraries) were roughly proportional to the mass of prey items consumed, raising the possibility of using DNA quantification methods to provide semi-quantitative estimates of diet composition.

3.1 Introduction

Determining trophic relationships within an ecosystem is a key part of many ecological studies (Trites 2003); however, obtaining reliable data on diet composition for most species is fraught with difficulties. Pinnipeds are one group of vertebrates whose diet has been extensively studied due to population declines of some pinniped species (e.g. Merrick *et al.* 1997; Sinclair and Zeppelin 2002), declines in numbers of some of their prey (e.g. Orr *et al.* 2004), and because of the overlap between their prey and species targeted by commercial fisheries (e.g. Harwood and Croxall 1988). At present, pinniped diet is primarily studied by morphological identification of prey hard part remains found in scats (cephalopod beaks, fish otoliths and bones) (Olesiuk 1993; Tollit and Thompson 1996; Sinclair and Zeppelin 2002). There are several well documented problems with data from these studies chiefly resulting from the fact that prey species with robust hard parts, which can readily survive digestion, are likely to be over-represented in scat whereas prey species with less robust or no hard parts are likely to be under-represented or not represented at all (Harvey 1989; Tollit *et al.* 1997). While numerical correction factors can reduce these biases, it has been demonstrated that digestion rates are affected by many different factors (Bowen 2000; Tollit *et al.* 2003), making their application problematic. Alternative methodologies such as stable isotope and fatty acid signature analyses (Hobson *et al.* 1997; Iverson *et al.* 2004) provide less specific, longer term data that are useful in many situations. However, they typically require animal capture and they do not provide the taxonomic and/or the numerical resolution that is often obtainable from scat analysis.

Traditional mammalian diet studies (using morphological identification of prey remains in scats) have been carried out in combination with genetic analysis of the predator's DNA present in scats to identify which species of predator the scat originated from (Farrell *et al.* 2000) and to ascertain the species and sex of the defecator (Reed *et al.* 1997). The use of DNA-based techniques to study diet directly has been carried out mainly in terrestrial invertebrates, where predators are sacrificed and the prey present in their stomachs is identified using polymerase chain reaction (PCR) detection methods (Symondson 2002; Agustí *et al.* 2003; Kasper *et al.* 2004). This destructive approach is clearly less acceptable in mammalian studies and has led to the development of non-invasive methods to detect prey DNA in the scats of

vertebrate predators (Höss *et al.* 1992; Jarman *et al.* 2002; Purcell *et al.* 2004). Multi-copy nuclear ribosomal and mitochondrial DNA (mtDNA) prey markers have been obtained from whale and penguin scats (Jarman *et al*. 2002, 2004), and single copy nuclear DNA prey genes (Y-chromosome fragments from male white-tailed deer) have been detected in the soft component of female bear scats (Murphy *et al*. 2003). Purcell *et al*. (2004) used DNA extracted from bones found in Pacific harbour seal scats to obtain species-level identification of salmonids. This approach allowed greater resolution than morphological analysis of the salmon bones (which provided identification only to family level), but identification was still contingent on prey hard parts surviving digestion. Detecting the presence of prey DNA in the soft component of scats may provide an alternative means of determining diet that is less affected by biases associated with differential digestion and passage (see Tollit *et al*. 2003) and could also allow for the detection of soft bodied prey items.

I conducted a feeding trial with captive Steller sea lions (*Eumetopias jubatus*) to further investigate the capacity of genetic techniques to recover prey DNA from scats. The diet of this sea lion has been well studied using conventional methods making it a good model species for evaluation the DNA-based approach. Steller sea lions are found across the North Pacific Ocean rim (from northern Japan, through the Aleutian Islands, and south to California; Figure 3.1). There has been a dramatic decline of the western populations over the past 30 years (Trites and Larkin, 1996), and to determine the causes of this population crash, a number of studies have investigated what these sea lions eat (Merrick *et al.* 1997; Sinclair & Zeppelin 2002; Winship & Trites 2003). Captive feeding trials have also been carried out with Steller sea lions to examine the biases associated with conventional hard part dietary analysis (Tollit *et al*. 2003).

The approach used in the current genetic feeding trial was to look at the reliability of PCR amplification of prey DNA from the scats of animals fed a consistent daily diet made up of several prey species. I analysed several sub-samples of each scat to determine distribution of prey DNA in the scats and included novel "pulse" prey items to monitor the persistence of the genetic signal. Finally, I assessed whether some quantitative estimate of diet composition could be obtained by quantifying the amount of DNA present in the scat through the screening of PCR clone libraries.

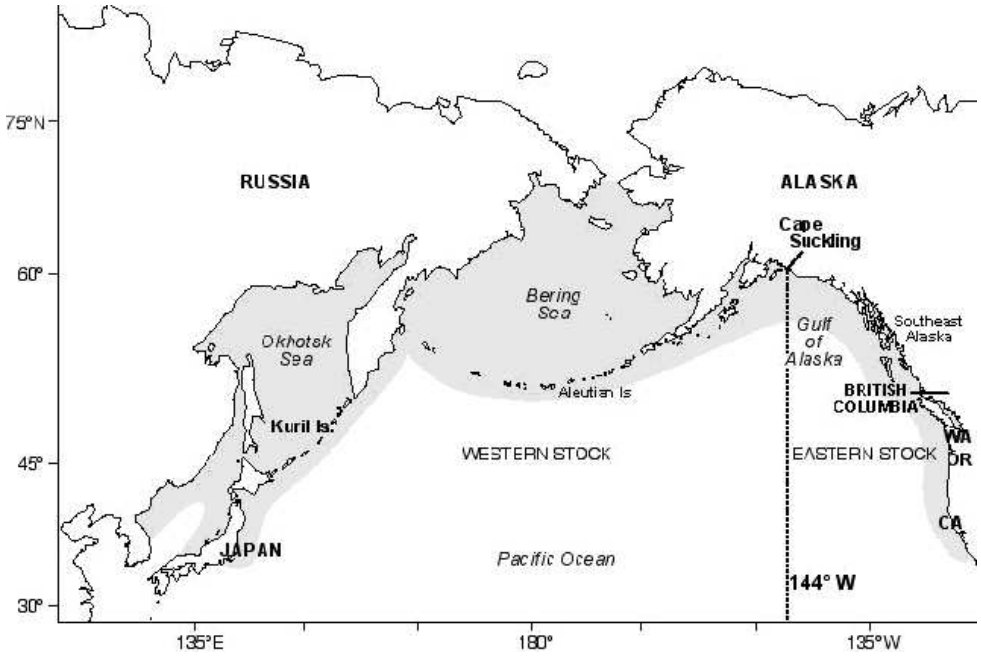


Figure 3.1 *Distribution of Steller sea lions and the delineation of the two distinct stocks. Map courtesy of the US National Marine Mammal Laboratory.*

3.2 Materials and Methods

3.2.1 Feeding trial and sample collection

Two female Steller sea lions participated in the trial from July to September 2003 at the Vancouver Aquarium Marine Science Centre. They were housed individually either in a continuously flowing 20,000 L saltwater swim tank with a 2 x 2 m haul-out platform or in a 1.8 x 2.5 m grated dry run. The first animal (Hazy, #F97HA, mean mass 146 kg, six years old) was in the feeding trial for 48 days and the second animal (Nuka, #F00NU, mean mass 131 kg, three years old) for 24 days. Six species of prey were used in the trial: Pacific herring (*Clupea pallasii*), surf smelt (*Hypomesus pretiosus*), sockeye salmon (*Oncorhynchus nerka*), walleye pollock *(Theragra chalcogramma*), capelin (*Mallotus villosus*) and Californian market squid (*Loligo opalescence*). The basic daily diet (7-8 kg per day, ~5.5% of body mass) was fed in two meals (at ~9:30 and 14:30) and consisted of herring (47% by mass), smelt (34%), salmon (13%) and squid (6%). This diet was initiated at least four days before the first scats were collected. Over most of the trial the diet was fed in two meals

which were equal in mass and had the same proportions as the total daily diet. However, during the first 14 days of the trial, Animal 1 was fed the basic daily diet in unequal meals, with only smelt being fed in both meals (Meal #1 was 6.25 kg consisting of 60% herring, 32% smelt and 8% squid; Meal #2 was 1.75 kg consisting of 57% salmon and 43% smelt). This regime allowed us to evaluate the extent of mixing of prey DNA from different meals. The other variation in the basic diet was the inclusion of novel prey species in place of herring. This was done on three occasions – one day where pollock was fed to Animal 1 in place of herring and another two days where capelin was fed to Animal 2 instead of herring. The purpose of these novel prey pulses was to determine how long prey DNA would be detectable in scats after consumption (see Figure 3.2 for overview of feeding regimes and Figure 3.3 for dates of sample collection).

During the feeding trial, samples were collected from either individual scats obtained on the dry-run/haul-out (n = 13), or from scats obtained through swim tank draining (n = 27) (Figure 3.3). Four samples were taken during each collection. Three small sub-samples (2-3 mLs faecal material) were obtained from distinct intact faecal lobes. The remainder of the scat (or random portions of several lobes when volume of scat was prohibitively large, >250 mL) was mixed to form a final blended sample. All scat samples were preserved in 95% ethanol 3-5 times the greater than the sample volume.

Several additional scat samples were analysed during the study. Four scat sub-samples were collected as control samples from captive sea lions not directly involved in the trial – three from an animal being fed a diet of solely Pacific herring and one from an animal being fed solely pollock. In order to investigate the degradation of DNA in unpreserved scats, components of two large scat samples were left at ambient temperature (high 26°C, low 13°C, mean 19°C) in an open container exposed to sunlight and sub-samples (n = 18) were preserved in ethanol at times ranging from 0 h to 18 days. Finally, twelve sub-samples were taken from various scats collected during the study period (from sea lions within and outside of the study that had known diets). I processed these samples using a blinded procedure (i.e. information on the prey DNA expected to be in these samples was withheld from me until I had finished the laboratory analysis).

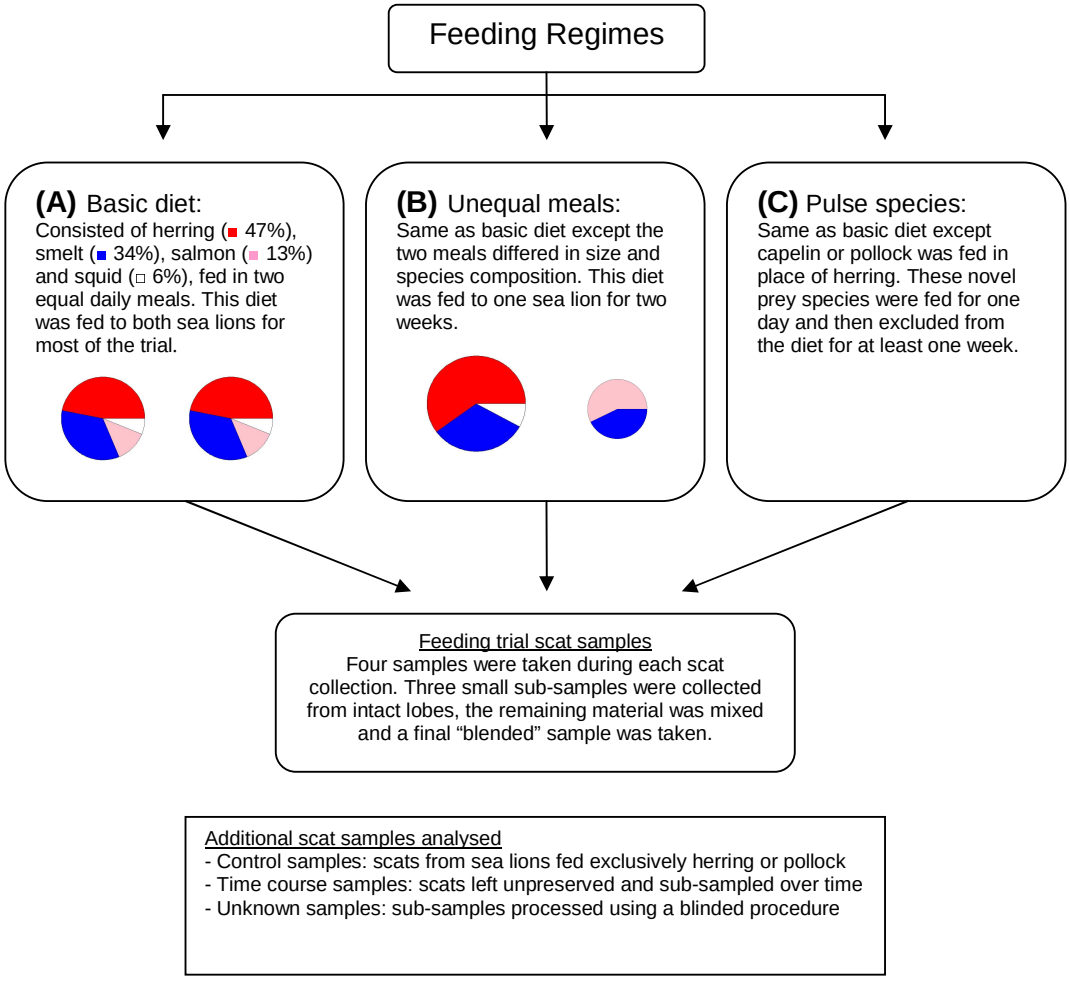


Figure 3.2 *Overview of feeding regimes and scat samples collected during the feeding trial.*

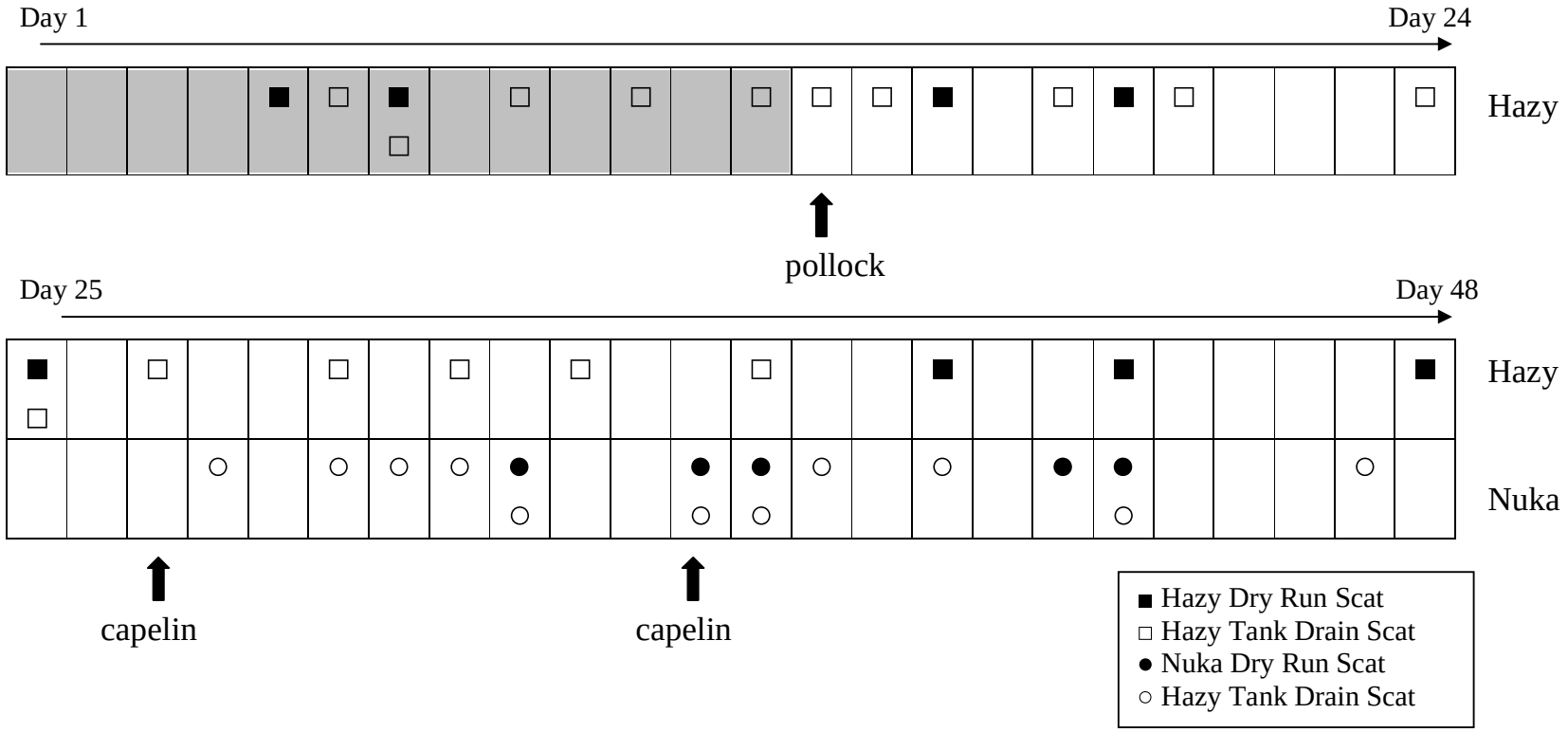


Figure 3.3 *Schematic of feeding trial experiment showing timing of scat collection and pulse feeding events (pollock and capelin arrows). Each block represents a day, for the first 24 days scats were only collected from one animal (Hazy), on days 25 – 48 scats were collected from two animals (Hazy and Nuka). Shaded days at the beginning of the trial (days 1-13) represent the unequal meal feeding regime. See text for further details.*

3.2.2 DNA extraction, PCR and sequencing

Extraction of DNA from scats was carried out using the QIAamp® DNA Stool Mini Kit (Qiagen). Samples were resuspended in the storage ethanol and then 1.5 mL of the ethanol/scat slurry was removed and centrifuged for 30 s at 6000 rpm in a microcentrifuge. The ethanol was poured off and the dry weight of the pellet was determined. All remaining steps followed the manufacturer's instructions, except that only half the recommended volume of buffers/ InhibitEX™ tablets was used. The buffer volumes were cut down to reduce the risk of crossover contamination by minimizing the number of pipetting steps and by reducing the volume of liquid loaded into spin columns and tubes.[6] The DNA was eluted in 100 μL Tris buffer (10 mM). In total, DNA was extracted from 194 samples (120 distinct sub-samples and 40 blended samples from the 40 feeding trial scat collections, three herring only scats, one pollock only scat, 18 exposure time course sub-samples and 12 blind sub-samples). Extraction blanks (containing no scat) were included (n = 8) to check for crossover contamination. Scat DNA extractions were done in a laboratory that had not previously been used for DNA analysis and were carried out before any prey DNA was extracted from tissue. Extraction of DNA from prey tissue was carried out using the Qiagen DNeasy tissue extraction kit (Qiagen).

The 3' end of the mitochondrial 16S ribosomal RNA gene was chosen as a PCR target since I had previously designed conserved primers which will amplify a short DNA fragment from the prey species used in the feeding trial (16S1F + 16S2R; Table 3.1) and I had also developed a DGGE based species identification method for this region (Deagle *et al.* 2005a). In addition to these "universal" primers, I amplified prey DNA using two sets of group-specific primers (Figure 3.4). One of these primer pairs specifically amplifies DNA from the fish prey and the other amplifies only squid DNA. The fish-specific forward primer was designed by aligning the Steller sea lion 16S mtDNA sequence (GenBank Accession NC 004030) with homologous sequences from the fish prey species fed in the feeding trial (GenBank Accession AY799999-AY800003). The resulting primer (16fishF, Table 3.1) was used in conjunction with 16SR (~ 250 bp product). It is completely conserved in the feeding trial fish species

[6] With the large volumes suggested in the kit the lids often came in contact with the sample and drops would be released on opening.

but only one out of five base pairs at the 3' end of the primer match the Steller sea lion and the primer is not conserved in squid. The squid PCR primers I used (Table 3.1) amplify a region of nuclear 28S ribosomal DNA (~180 bp product) from squid, but not from other molluscs or more distantly related animal taxa (K. Goldsworthy and S. Jarman, unpublished data).

Standard PCR reactions were performed on 1 µL of DNA extracted from scat in a 25 µL volume containing 0.4 µM of each primer, 0.2 mM dNTPs, 2.0 mM $MgCl_2$, 1x BSA (New England Biolabs),1x AmpliTaq Gold® buffer and 0.625 units AmpliTaq Gold® (Applied Biosystems). Cycling conditions were as follows: 94°C for 10 min then 35 cycles (94°C, 30 s / 55°C, 30 s / 72°C, 45 s) followed by 72°C for 2 min. Aerosol-resistant pipette tips were used with all PCR solutions and negative control reactions (extraction controls and a PCR blank) were performed with each batch of PCR amplifications. Samples were separated on a 1.8% agarose gel or gradient acrylamide gels. Sequencing was carried out using the CEQ Dye Terminator Cycle sequencing Quick Start Kit, employing half reactions; products were electrophoresed on a Beckman Coulter CEQ 2000 sequencer. Sequences from the 16S ribosomal RNA gene of the fish prey species were obtained through direct sequencing of PCR products amplified using the primers 16Sar-5' and 16Sbr-3' (Palumbi 1996)[7].

Table 3.1 *PCR primers used in this chapter (size of products given Appendix II).*

Pimer name[a]	Sequence 5'→3'	Target (gene: species)
16S1F	ggacgagaagaccct	mtDNA 16S: sea lion, fish, squid
16SfishF	agaccctatggagctttagac	mtDNA 16S: fish in feeding trial
16S2R	cgctgttatccctatggtaact	mtDNA 16S: sea lion, fish, squid
16S2R Clamp	gggcgggggcggcgggacgggcgcgggg cgcggcgggcgcgctgttatccctatggtaact	mtDNA 16S: sea lion, fish, squid
Squid28SF	cgccgaatcccgtcgcmagtaaamggcttc	nuclear 28S rDNA: squid
Squid28SR	ccaagcaacccgactctcggatcgaa	nuclear 28S rDNA: squid

[a] F and R denotes forward and reverse.

[7] The 16Sbr-3' primer was inadvertently refered to as 16Sa-3' in Deagle *et al*. 2005b.

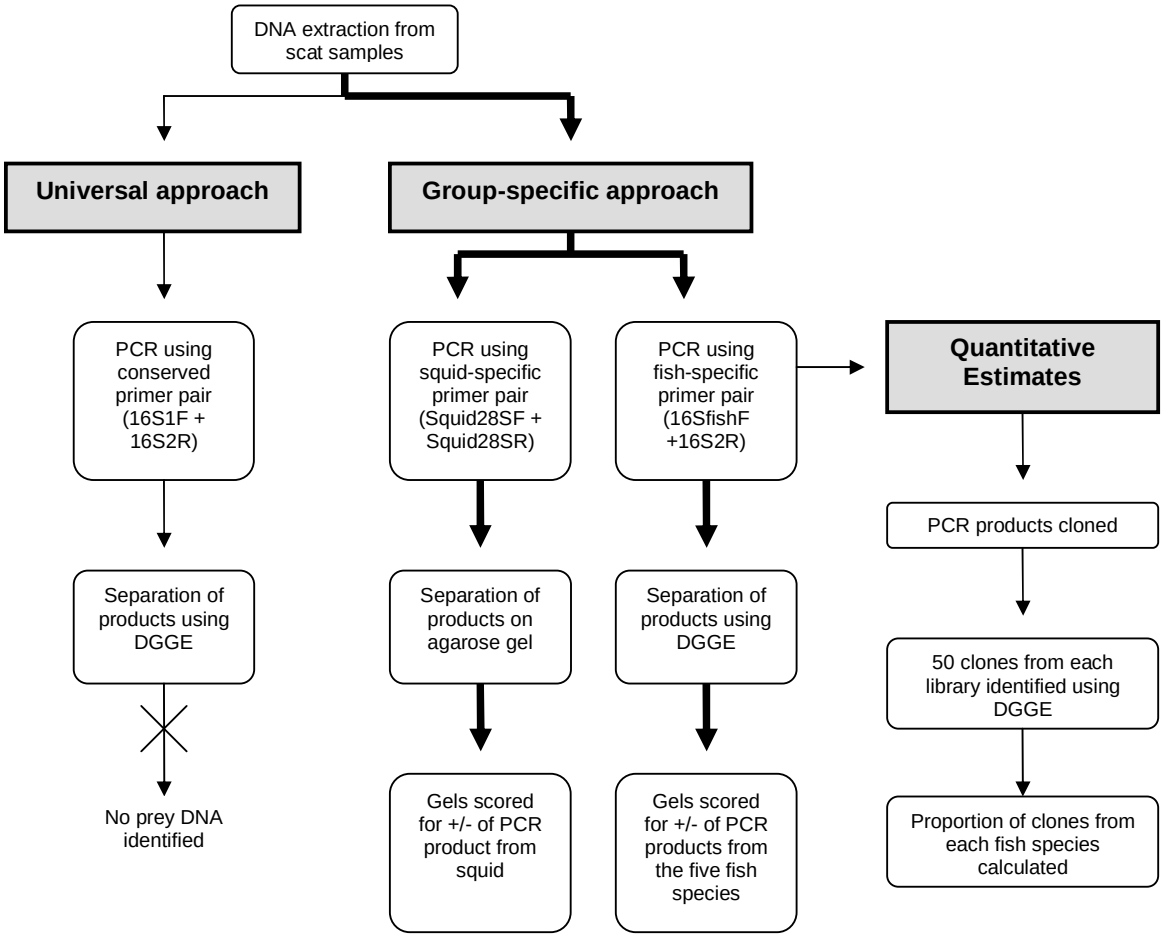


Figure 3.4 *Overview of genetic analysis performed in this chapter. Thick arrows represent analysis carried out on all scat samples, thin arrows show analysis carried out on a sub-set of samples.*

3.2.3 DGGE analysis

I used denaturing gradient gel electrophoresis (DGGE) to separate unique PCR products in amplifications expected to contain DNA from multiple species (see Myers *et al*. 1987; Lessa and Applebalm 1993). DGGE was performed using the DCode™ system (Bio-Rad). Acrylamide gels were prepared according to the manufacturer's instructions and poured using a Model 475 Gradient Delivery System (Bio-Rad).

For samples separated by DGGE, the16S2R primer was redesigned to incorporate a GC clamp (Sheffield *et al*. 1989). Nested PCR was carried out to improve the intensity of the bands obtained from the fish-specific PCR (using the clamped reverse primer) and scat DNA template.[8] Primary enrichment PCR was conducted using the unclamped primer pairs (16S1F and 16S2R). Cycling conditions were: 94°C for 10 min then 20 cycles (94°C, 30 s/ 56°C, 30s/ 72°C, 1 min), followed by 72°C for 2 min. The secondary PCR was carried out as for standard PCRs (see previous section) with the clamped reverse primer and 1 μL of the primary reaction as template. Electrophoretic conditions (percentage acrylamide, gradient range, voltage and length of run) that resulted in clear band separation were determined by experimenting with products amplified from genomic DNA of the prey species (see Results section).

3.2.4 Quantitative Estimates

Clone libraries were constructed from fish PCR products to quantify the proportions of fish DNA present in eight samples (Figure 3.4). The eight samples analysed included five DNA samples from scats collected when the sea lions were fed the basic diet in equal meals and three DNA samples from scats collected when the daily diet was fed in unequal meals. The clone libraries represent template DNA from single scat samples (n = 5, three from equal diet feeding regime and two from unequal diet feeding regime) or mixtures of DNA from seven scats (n = 3, two from equal diet feeding regime and one from unequal diet feeding regime). The DNA mixtures were included to determine if pooling of DNA from several extractions would provide an

[8] See discussion at the end of Appendix I on the use of nested PCR.

average view of diet rather than the snapshot that might be expected from an individual scat. Scats potentially containing pollock or capelin DNA were not included in this analysis. Standard PCR (as above) was carried out using 50 μL reaction volumes, 2 μL of scat DNA template and the 16SfishF and 16SR primer pair. PCR products were cloned into the pCR®2.1-TOPO TA cloning vector and transformed into TOP10 chemically competent *Escherichia coli* (Invitrogen). The bacteria were plated and positive transformants recognized using blue/white colour selection. For each sample, 50 white colonies were picked using a pipette tip and suspended in a 20 μL PCR mixture containing the primers (16SF and 16SR). Standard PCR was carried out to amplify DNA from each colony and the amplified product was identified by DGGE analysis. By tallying the identity of 50 clones in each library, I obtained an approximate estimate of the proportions of fish DNA present in each sample. The fish component of the daily diet comprised of herring (50%), smelt (36%) and salmon (14%). If the prey DNA in the clone libraries were present in identical proportions to the mass of the prey items in the diet, the expected range in the proportions estimated by the approach is quite wide simply due to sampling variability. For a random sample of 50 drawn from a multinomial distribution with the proportions 50%, 36% and 14%, the observed proportions have a 95% chance of falling in the range 50%±13.9%, 36%±13.3% and 14%±9.6% respectively (calculated using the formula $p \pm 1.96\sqrt{p(1-p)/50}$, where p is the true proportion).

3.2.5 Data analysis

For statistical tests, each DNA extraction was treated as an independent sample. Statistical differences in PCR detection rates were evaluated by Chi-squared contingency table tests for comparisons between (i) the different prey species (ii) the blended samples and the sub-samples and (iii) the samples collected during the basic diet equal and unequal meal feeding regimes. Chi-square tests were also carried out to evaluate whether estimates of the proportions of fish DNA were consistent between scat samples and whether these estimates were consistent with the proportions of fish mass in the diet. Statistical tests were performed using The R Foundation for Statistical Computing Version 1.9.1 alpha software (R Development Core Team 2004).

3.3 Results

3.3.1 DGGE separation of PCR products

The DGGE conditions that allowed separation of 16S1F and 16S2R Clamp PCR products for the six prey species were a linearly increasing 30-70% denaturing gradient (with 100% denaturants being 40% formamide and 7M urea) in a 7.5% polyacrylamide gel. PCR products amplified from the fish DNA using the primers 16SfishF and 16S2R Clamp were separated on a 6% polyacrylamide gel that also contained a linearly increasing 30-70% denaturing gradient (Figure 3.5a). The running temperature for both was 56°C and the 16 cm gels were run at 70 V for 8 h. Initial runs separating bands amplified from genomic DNA of the prey species identified two alleles for herring. These alleles ran further than bands from any other species on the gel and both were scored as herring in subsequent analysis.

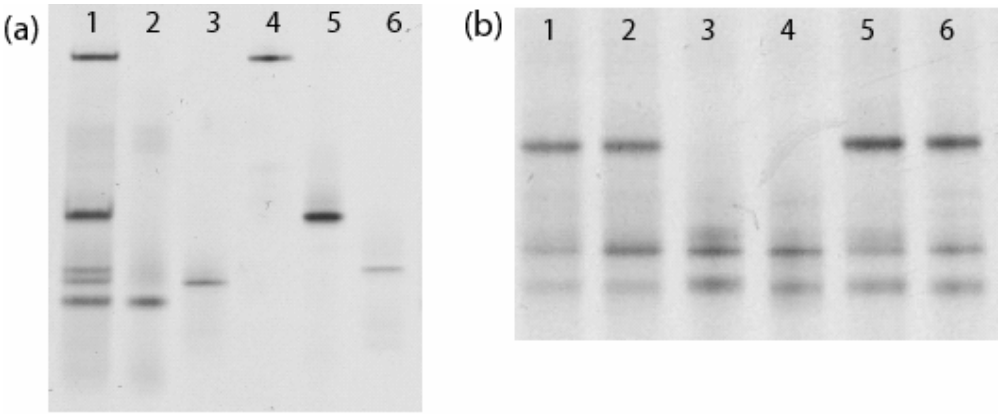


Figure 3.5 *Denaturing gradient gel electrophoresis of 16S mitochondrial DNA fragments amplified from fish prey species fed during the feeding trial. (a) PCR products amplified from genomic DNA template: 5 species mix (lane 1), Pacific herring (lane 2), smelt (lane 3), pollock (lane 4), sockeye salmon (lane 5) and capelin (lane 6) (b) PCR products amplified from DNA extracted from scat. Four samples amplified salmon, smelt and herring DNA (lanes 1, 2, 5, 6) and two samples amplified only smelt and herring (lanes 3, 4).*

3.3.2 Detection of prey DNA in feeding trial scat samples

The average mass of scat used in the DNA extractions was 79 ± 29 mg. The initial PCR tests were done on DNA from five scat samples using the 16SF and 16SR primers (universal approach, Figure 3.4). Amplification from each produced a single band which did not migrate with any of the prey bands on the DGGE gel. The five amplification products were cloned and four clones from each were sequenced. All 20 sequences matched perfectly with a Steller sea lion sequence from GenBank (NC 004030). An additional 50 clones were screened using DGGE in an attempt to identify DNA from prey which could potentially be present at low level in these PCR products. Each of these clones also originated from sea lion DNA.

To overcome swamping by the predator's DNA, all further prey detection attempts were done using the group-specific approach (Figure 3.4), which excluded sea lion DNA from amplification. When the sea lions were fed the basic diet (i.e. excluding the scats collected when Animal 1 was fed unequal meals as well as those collected within 48 h of the pulse species being fed), a total of 108 samples (27 blended and 81 sub-samples) were collected from 27 scats. Using squid-specific primers, PCR amplification of DNA extracted from these samples gave an overall percent frequency of detection (PFD) of 94% – with squid DNA being detected in at least one of the samples from every scat. Using the fish specific primers the PFD of fish DNA was 97%. The PFD levels for the individual fish species scored on DGGE gels (Figure 3.5b) were 94% for herring, 87% for salmon and 92% for smelt. As with the squid, DNA from each fish species was detected in at least one sample from every scat. The detection levels of the four prey species in the basic diet were not significantly different from one another ($\chi^2 = 0.4307$, df = 3, p-value = 0.9338). The PFD values were significantly higher for the blended scat samples (98%) compared with the sub-samples (90%) ($\chi^2 = 6.5$, df = 1, p-value = 0.011); results are summarized in Table 3.2.

In scat samples collected from Animal 1 while she consumed unequal meals, the overall PFD was 66% versus a PFD of 95% when she received the basic diet with meals of equal composition (Table 3.3). This difference in detection level was highly significant ($\chi^2 = 52.46$, df = 1, p-value < 0.001) and reflects patchy prey detection (which roughly corresponds to meal composition), not an increase in the number of

scat samples failing to produce any PCR products. Smelt was the only species included in both daily meals during the 14 day unequal meals feeding regime. In scats collected over this period, smelt had a PFD of 86% compared with 61% for herring, 68% for salmon and 50% for squid. Results from the pulse prey feedings (pollock fed for a single day and capelin fed for a single day on two occasions) showed that both pollock and capelin were only detected immediately following their inclusion in the diet and their detection was limited to scats collected within 48 h of consumption (Table 3.4).

Table 3.2 *Frequency of detection of prey DNA in scat samples collected during the basic diet feeding regime of the feeding trial (108 samples collected from 27 scats). This summary excludes results from scats collected within 48 h of the sea lions being fed pulse diet items and results from scats collected when diet was being fed in unequal meals.*

	Blended sample	Sub-sample	Total
Squid	100% (27/27)	93% (75/81)	94% (102/108)
Herring	100% (27/27)	93% (75/81)	94% (102/108)
Smelt	96% (26/27)	90% (73/81)	92% (99/108)
Salmon	96% (26/27)	84% (68/81)	87% (94/108)
Total	98% (106/108)*	90% (291/324)*	92% (397/432)

* Difference between blended and sub-samples $\chi^2 = 6.48$, df = 1, p-value = 0.011

3.3.3 Control samples

DNA extracted from the control scat samples (collected from sea lions fed only herring or pollock) produced no visible PCR products with the squid primer set and the fish primer set produced single bands on the DGGE gels which matched the expected prey species. All extraction blanks were negative.

Table 3.3 *Frequency of detection of prey DNA in scats collected from a sea lion while being fed the same basic diet in either: (1) two daily meals of equal size and species composition (60 samples from 15 scats) or, (2) two daily meals of unequal size and species composition (28 samples from 7 scats). Results exclude the scats collected within 48 h of the animal being fed pulse diet items.*

	Equal meals	Unequal meals
Squid	95% (57/60)	50% (14/28)
Herring	95% (57/60)	61% (17/28)
Smelt	95% (57/60)	86% (24/28) †
Salmon	97% (58/60)	68% (19/28)
Total	95% (229/240)*	66% (74/112)*

† Smelt was included in both meals, remaining species fed exclusively in meal 1 or meal 2

* Difference between equal and unequal meals $\chi^2 = 52.45$, df = 1, p-value < 0.001

Table 3.4 *Summary of the PCR detection results in days following inclusion of pulse species (pollock or capelin) in the diet for a single day. Each shaded block represents a scat sample and symbols show the presence (+) or absence (–) of the DNA in the blended sample and three sub-samples tested.*

	Day 1	Day 2	Day 3	Day 4	Day 5
Pollock	– – – +	+ + – –	– – – –		– – – –
Capelin 1	+ + + +		– – – –	– – – –	– – – –
Capelin 2	+ + – –	+ + + +		– – – –	
	+ + + +				

3.3.4 Time course samples

The two scat samples that were sub-sampled over an extended time period initially had detectable DNA present from each of the four prey species in the basic diet. The first scat was sampled from 0-8 days with most samples being taken over the first two days. The second scat was sampled every few days over an 18 day period. In both samples detection failed simultaneously for all prey markers (between day five and eight for the first scat and between day two and seven for the second scat). I tested all samples with the 16S1F and 16S2R conserved primer set to determine if the decomposing scats had produced chemicals with a strong inhibitory effect on PCR amplification. These primers produced PCR products for all of the sub-samples, indicating amplification was possible from the templates where detection of prey DNA had failed. The PCR product was presumably from sea lion template and the drop-off in amplification of prey but not predator DNA is likely due to the larger initial amount and higher quality of predator DNA in the samples.

3.3.5 Unknown samples

Using the blinded procedure I tested 12 scat samples for the presence of six potential prey items. DNA was extracted from each sample only once. Two of the 12 samples failed to produce any PCR products, 28 prey items were detected in the remaining 10 samples (Table 3.5). Subsequent comparison of the genetic test results with known prey species in the diets indicated that I identified 100% of the species eaten in eight of the scat samples and identified the major diet component but missed other minor prey species in two samples (Table 3.5). No false positives were obtained in this analysis.

Table 3.5 *Results of blind PCR tests preformed on 10 scat sub-samples. Symbols indicate presence or absence of DNA marker in scat / presence or absence of prey species in diet, incongruent results are shaded. Results are not shown for 2 scat samples which produced no PCR products.*

	Scat samples									
	1	2	3	4	5	6	7	8	9	10
Pollock	–/–	–/–	–/–	–/–	+/+	–/–	–/–	–/–	–/–	–/–
Salmon	+/+	+/+	+/+	+/+	–/–	–/+	+/+	+/+	–/+	–/–
Capelin	+/+	–/–	–/–	–/–	–/–	–/–	–/–	–/–	–/–	–/–
Smelt	+/+	+/+	+/+	+/+	–/–	–/+	+/+	+/+	–/+	–/–
Herring	–/–	+/+	+/+	+/+	–/–	+/+	+/+	+/+	+/+	+/+
Squid	+/+	+/+	+/+	+/+	–/–	–/+	+/+	+/+	–/+	–/–

3.3.6 Quantitative estimates

Analysis by PCR/cloning detected all of the fish prey species in the five scat samples collected when the sea lions were fed two equal meals per day (50 clones analysed per sample). The proportional estimates from these scats were consistent with each other ($\chi^2 = 9.47$, df = 8, p-value = 0.305) even though there was considerable variation (herring ranged from 54-72%, smelt ranged from 12-28% and salmon ranged from 16-28%); results summarised in Figure 3.6. Only one of the estimates from these five scat samples was consistent with proportions of fish mass in the diet (i.e. had a χ^2 value greater than 0.05). Herring and salmon tended to be over-represented while smelt was under-represented in the clone libraries (Figure 3.6).

The proportional estimates were much more variable for the three scat samples collected when the daily diet was fed in two unequal meals. The proportions of fish DNA in the two libraries produced from individual scats matched the composition of individual meals better than the overall daily diet (Figure 3.7). The daily proportion of herring was grossly under-estimated in one library and salmon was completely absent in the other. Only the proportion of smelt (which was fed in both meals) was estimated reasonably well. The quantitative estimate obtained from the clone library

produced by mixing DNA from seven of these patchy scats did provide an estimate more in line with the overall proportions of prey items in the diet (Figure 3.7).

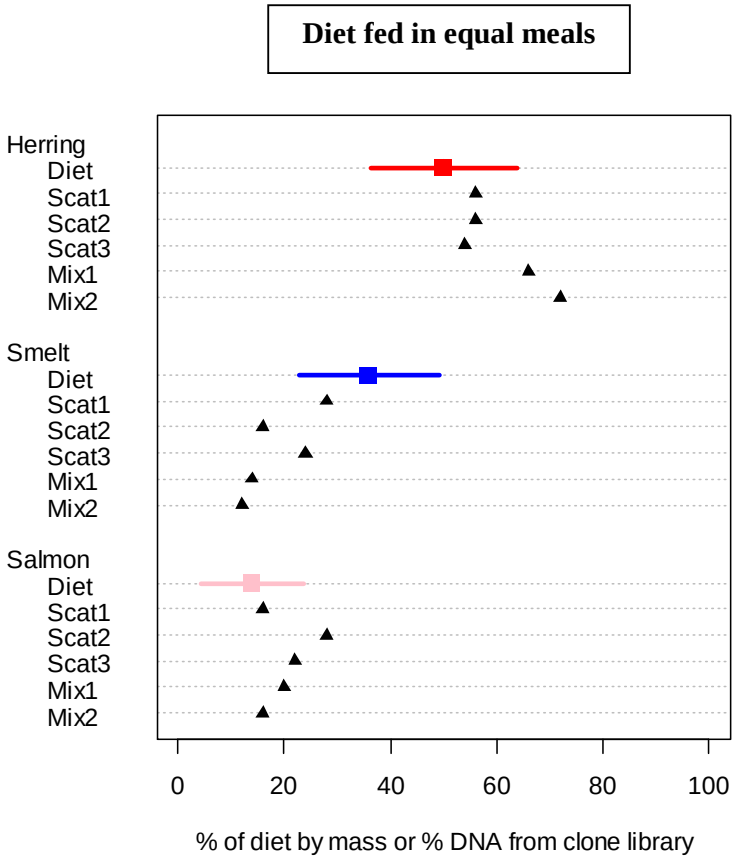


Figure 3.6 *Dot plot showing estimates of the proportions of fish species in diet obtained through analysis of DNA in clone libraries. These scat samples were collected when the two daily meals were identical in composition. Points show the proportion of each fish species by mass in diet (■) and DNA proportions in clone libraries (▲). Bars represent the 95% probable range of estimates (due to sampling error) if DNA proportions are equivalent to mass proportions (see text for details). Clone libraries represent DNA from single scat samples (Scat1-3) or mixtures of DNA from 7 scats (Mix1-2).*

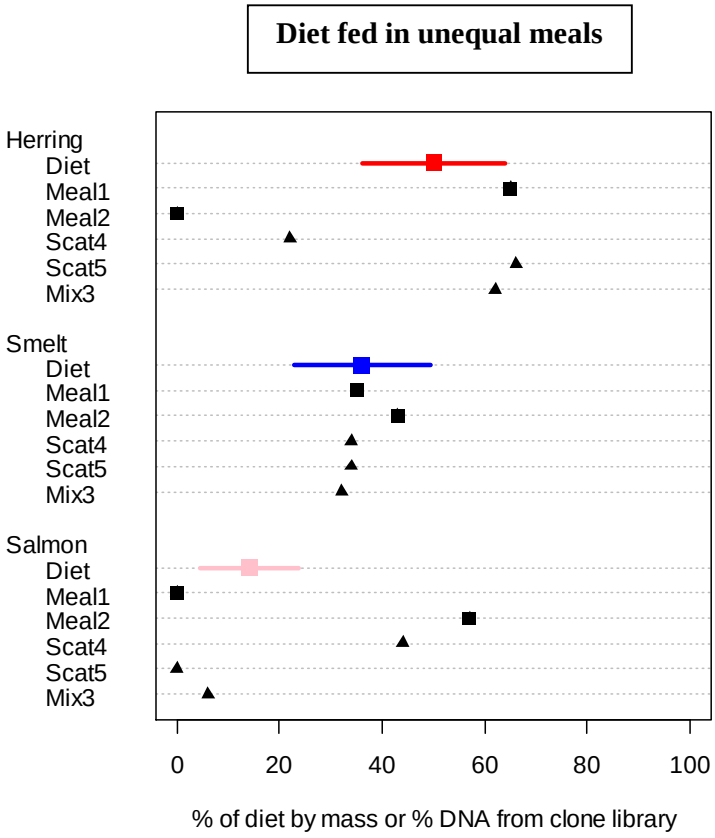


Figure 3.7 *Dot plot showing estimates of the proportions of fish species in diet obtained through analysis of DNA in clone libraries. These scat samples were collected when the two daily meals differed in composition. Points show the proportion of each fish species by mass in diet (■) and DNA proportions in clone libraries (▲). Bars represent the 95% probable range of estimates (due to sampling error) if DNA proportions are equivalent to mass proportions (see text for details). Clone libraries represent DNA from single scat samples (Scat 4-5) or mixtures of DNA from 7 scats (Mix 3).*

3.4 Discussion

Although previous studies have amplified prey DNA from vertebrate scat using PCR, this chapter reports the first results from a controlled feeding trial looking at detection of DNA from several prey species. I initially attempted to use a single PCR test to simultaneously amplify DNA from all prey items present in each scat sample. This approach has advantages, primarily because primer binding and PCR conditions will be consistent for all prey species, and the laboratory analysis is minimised. However, primers which are conserved among the target prey species usually also amplify DNA from the predator by necessity. This was a serious problem in the current study – by direct screening of PCR products generated using universal primers only sea lion DNA was detected in the scat samples. Predator DNA was expected to be prevalent since a previous study found that nearly one third of the PCR products generated from fin whale scats (using metazoan-specific primers) matched the fin whale DNA sequence (Jarman *et al*. 2004). However, the absolute dominance of sea lion DNA was unexpected and reinforced the need to actively exclude the predator DNA from analysis. This can be accomplished by designing PCR primers targeting evolutionary cohesive groups of prey that will not amplify predator DNA (see Jarman *et al*. 2004), or it can be achieved using subtractive screening methods (such as predator DNA specific restriction enzyme digestion or subtractive hybridization).

I chose to employ group-specific approach, using primers that targeted short regions of fish-specific (mitochondrial) and squid-specific (nuclear ribosomal) DNA. PCR products from the five fish species were separated using DGGE. Using this method the prey DNA detection success rates were very high (close to 100% for samples taken from scat samples that had been blended). Detection success rates for squid and sockeye salmon, which were fed as a relatively small percentage of the daily diet, were equivalent to rates for the more abundant diet items (smelt and herring). I did find that the detection of prey DNA was less consistent in samples taken from a small distinct part of individual scats compared with samples that were taken from blended scat. This finding suggests that DNA from different prey species is not distributed evenly within a single scat and has implications for the development of sampling protocols. The results also demonstrate that prey DNA from different

meals consumed on the same day is not well mixed among scats, implying that the DNA in each scat represents the prey species consumed over a short time interval. Analysis of scats produced after the inclusion of novel pulse prey items in the diet showed that detection of the novel prey DNA is restricted to scats produced within 48 hours of consumption. While these results need verification with other species of prey (and predator species), they indicate that prey DNA in scat samples can be reliably detected through PCR analysis and this can provide fine resolution data on recently consumed prey. The limited time that prey DNA can be detected after ingestion is a constraint shared to some extent by hard part analysis (Tollit *et al*. 2003); this is a serious limitation for studies of marine mammals that forage long distances from sites where scats can be collected. Estimates of diet over longer time periods may have to be obtained from alternative techniques such as analysis of fatty acids (Bradshaw *et al*. 2003; Iverson *et al*. 2004) or by using animal-borne video systems (Bowen *et al*. 2002).

Using DGGE to separate DNA markers from multiple fish species worked satisfactorily in the captive feeding trial but there were a few limitations. First, the presence of sequence diversity within the herring that were used in the experimental feedings resulted in two markers coming from this single prey item. This feature could confuse interpretation of results in a field application where more inter-specific prey diversity might be expected. Second, I found a bright heteroduplex band formed between capelin and smelt. This band migrated near the pollock-sized band and could be accounted for in this controlled study (since pollock and capelin were not fed together) – but again, this could cause difficulties in a field application. I suggest that DGGE should only be applied when the diversity of the prey targeted by the PCR tests is limited and well defined, otherwise cloning and sequencing is probably a better option (e.g. Jarman *et al*. 2004; Kasper *et al.* 2004).

PCR-based detection methods have several limitations in situations where target DNA is present in low amounts and the quality of samples is poor (Taberlet *et al*. 1999). The problem most likely to be encountered in the analysis of prey DNA in scat is the production of false positive and/or false negative results. Cooper and Poinar (2000) outline procedures for working with ancient DNA to help prevent the occurrence of false positives (amplification of small amounts of contaminant DNA when target DNA is absent in the sample). To apply all these procedures to studies of prey DNA from even moderate numbers of scat samples would be unfeasible.

Fortunately, molecular analysis of prey DNA in contemporary scat samples is not quite as extreme as ancient DNA research. Physical isolation of workspace for pre-amplification steps, proper use of negative control amplifications, verification of the reproducibility for a sub-set of samples and some cross-validation with independent hard part analysis are minimum precautions that need to be carried out to allow confidence in positive results from field studies (Taberlet *et al*. 1999).

The production of false negative results (failure of amplification when target DNA is or was present in the sample) could be due to a number of reasons. These include degradation of the DNA present in the sample, failure of the DNA extraction or failure of the PCR amplification. In the current study, the known diet and high prey DNA detection rates allowed us to rule out the occurrence of large numbers of false negatives. However, in studies where the diet is unknown, monitoring the incidence of false negatives is extremely difficult. An indication of the potential frequency of false negatives in molecular scatology studies can be obtained by looking at the amplification success rate of predator mtDNA from scat (Table 3.6).

Table 3.6 *Percentage detection of predator mtDNA from studies carrying out PCR analysis of DNA from mammalian scat.*

% Detection of Predator mtDNA	Sample Size	Species	Reference
70%	20	Harbour and Grey Seals	Reed *et al.* (1997)
80%	20	Black Bear	Wasser *et al.* (1997)
90%*	50	Sun Bear	Wasser *et al.* (1997)
100%	23	Mustelid Species	Hansen and Jacobsen (1999)
59%	34	Felid and Fox	Farrell *et al.* (2000)
53%	163	Marten and Fox	Davison *et al*. (2002)
90%*	30	Marten	Davison *et al.* (2002)
84%	130	Wolf	Lucchini *et al.* (2002)
88%*	300	Brown Bear	Murphy *et al.* (2003)
77%	128	Fox, Wolverine	Dalén *et al.* (2004)

* captive animals

These data show that it can be difficult to amplify predator mtDNA from scats even though this target is expected to be ubiquitous and, based on results from the current study, present in higher quantities than prey DNA. It is also obvious from these data that the frequency of negative results obtained in different studies varies considerably, with field-based studies tending to have a higher incidence of negative results compared with captive animal studies.

Completely eliminating false negatives in DNA-based diet studies is not likely to be possible. However, there are ways to reduce the occurrence of this type of error. Obtaining fresh scat samples is of primary importance. I found that in unpreserved scat samples, prey DNA was no longer detectable after five to seven days; this indicates that while scats do not have to be collected immediately after defecation, they should be as fresh as possible. The identification and exclusion of samples of very poor quality can be done by pre-screening of DNA extractions (e.g. Morin *et al*. 2001) or through the use of internal positive controls (i.e. primer sets which target prey groups expected to be represented in all scats). The number of false negatives obtained when focusing on good quality scats can be reduced by processing several samples per scat. In the current study, I failed to detect some prey DNA in scat sub-samples collected when the sea lions were being consistently fed the basic diet. However, when I pooled data from the four samples analysed from each scat, DNA from all prey species was identified in every scat.

Estimating not only prey diversity but also relative amounts of prey eaten is the goal of many pinniped diet studies. This information can be used in conjunction with estimates of the predator species' energetic requirements and prey energy density to obtain overall consumption estimates (e.g. Olesiuk 1993; Winship and Trites 2003). Presence/absence data from hard part studies has been used to estimate the relative frequency of occurrence of different prey in the diet (Sinclair and Zeppelin 2002). Genetic data seems suited to this type of analysis since large numbers of samples can be screened once appropriate tests have been developed (although the potential occurrence of false negative results must be carefully considered – see discussion above). The heterogeneous distribution of prey DNA within a scat and the short detection period of prey DNA are also advantageous in this type of study since the detection of prey in different scat samples will likely represent independent observations (see Tollit *et al*. 2003 for discussion).

Estimates of the relative proportions of different prey species in the diet can be improved through the reconstruction of prey biomass. In studies that rely on recovering and identifying hard parts, this is accomplished by estimating the number and size of prey consumed based on counts and measurements of hard parts recovered in scat. These estimates are biased since the percentage of hard parts recovered from different prey species varies considerably and the size of hard parts is often reduced due to digestion (Tollit *et al.* 1997; Bowen 2000). In captive feeding trials, the recovery rates for fish otoliths range from 0-89% and otolith digestion results in the underestimation of fish length by 16-51% (Tollit *et al.* 1997). Correction factors have been developed to account for these biases (Harvey 1989; Tollit *et al.* 1997), but final estimates remain limited by wide confidence intervals.

Genetic analysis of scats could potentially provide biomass estimates if the amount of DNA from each prey species is proportional to the mass of the prey in the diet. Even if it is only possible to classify prey biomass present in each scat into a few categories using DNA quantification (e.g. < 20%, 20-50%, 50-80%, >80%), these data would allow a substantial improvement in estimates of overall proportions of the prey in diet compared with simple presence and absence data (S. Jarman, unpublished data). The quantitative estimates of the proportions of fish prey DNA in scats did provide a rough estimate of the proportion of the fish present in the meals fed during the trial. Averaged over a number of samples this level of accuracy would provide useful data on diet composition and will likely provide better quantitative estimates than obtained from hard part analysis under some scenarios (Da Silva and Neilson 1985; Jobling 1987). By pooling DNA from several "patchy" scats collected when a sea lion was fed two meals with different prey composition, I obtained a composite picture of overall diet rather than of individual meals. This sample averaging approach could be useful for reducing the number of samples that need to be analysed[9]. I did observe some directional bias in the amount of DNA coming from different prey species. This could be due to a methodological factor such as PCR bias (Suzuki & Giovannoni 1996), or it could be due to species-specific differences in mtDNA copy number, cell density or DNA survival during digestion. Presumably correction factors similar to those used with hard part dietary data could be developed and applied. Based on these initial quantitative results, further assessment of the

[9] See further discussion of this idea in Chapter 4 (p. 82).

ability of DNA amounts in scats to estimate proportional diet composition seems warranted. The clone screening quantification technique I have used here is straight forward to apply; however, it is laborious (precluding the analysis of a large number of samples) and allows only end-product analysis. Real-time PCR quantification could be a more productive approach since it will allow rapid simultaneous quantification of DNA from multiple prey species. Real-time PCR will also allow quantitative comparisons between prey species targeted by different PCR tests (e.g. the amount of fish compared with squid in the current study); this type of comparison was not possible using the clone library approach I employed.

The use of molecular scatology to study diet has the potential to provide new insight into the diet of vertebrate predators. More accurate taxonomic identification of prey remains in scats can be obtained (Hofreiter *et al*. 2000; Purcell *et al*. 2004) and employing genetic methods has obvious benefits in cases where soft bodied prey or prey with fragile bones are suspected to be an important part of the diet. The technique also provides a means to carry out an independent dietary analysis of scats. The combination of genetic and hard part data sets should also help clarify the errors associated with each method. Measures of diet diversity and studies focusing on small groups of prey species are currently feasible using genetic techniques. However, genetic markers will need to be characterised for many prey species if the approach is to be widely applied, and several aspects of the methodology will need further development before a complete picture of diet can be constructed through molecular scatology.

Chapter 4

Analysis of prey DNA in faeces of captive Steller sea lions
Part II: DNA quantification using real-time PCR

One fish
two fish
red fish
blue fish

— Dr. SEUSS

Abstract

Following on from Chapter 3, I continued to investigate the possibility of using the relative amounts of DNA recovered from different prey in faeces to obtain quantitative diet composition data. Here, the fish mtDNA in faeces obtained from the Steller sea lion captive feeding trial were analysed using quantitative real-time PCR (qPCR). The faecal samples were collected when the animals were being fed a diet consisting of 50% Pacific herring (*Clupea pallasii*), 36% surf smelt (*Hypomesus pretiosus*) and 14% sockeye salmon (*Oncorhynchus nerka*) by mass. Quantitative real-time PCR was used to measure the amount of mtDNA from the three fish species in: (i) a blended tissue mix representative of the sea lion diet and (ii) the sea lion faecal samples. The percent composition of fish mtDNA extracted from the undigested tissue samples (n = 10) corresponded reasonably well to the mass of fish in the mixture (58.6 ± 4.6% for herring, 27.9 ± 2.2% for smelt and 13.5 ± 3.1% for salmon). In the faecal samples (n = 23) the absolute amount of fish mtDNA recovered varied 100-fold, but the percent composition of the three fish was relatively consistent (57.5 ± 9.3% for herring, 19.3 ± 6.6% for smelt and 23.2 ± 12.2% for salmon). Differences between the mtDNA proportions in the tissue samples compared to the faecal samples indicate there are prey-specific biases in DNA survival during digestion. Nevertheless, these biases may be less than those commonly observed in the conventional analysis of prey hard remains and further investigation of this approach is warranted.

4.1 Introduction

Pinnipeds are important top level predators in many marine ecosystems (Bowen 1997). Their population size can be affected by changes in prey availability (Soto *et al.* 2006), and when pinniped populations are large, they can have significant impacts on populations of their prey (Browne *et al.* 2002; David *et al.* 2003). Reliable methods of estimating pinniped diet composition are critical for the appropriate management in situations where populations of pinnipeds or their prey are endangered (Sinclair & Zeppelin 2002; Purcell *et al.* 2004). Diet composition is usually determined by analysis of hard part remains in faeces, but estimates can be inaccurate due to variation in the recovery of hard parts from different prey species (Tollit *et al.* 1997; Bowen 2000).

Recently a number of studies have demonstrated the feasibility of using DNA-based methods to identify prey hard remains in pinniped faeces (Purcell *et al.* 2004; Kvitrud *et al.* 2005), or to identify prey from DNA present in the soft matrix of faeces (Deagle *et al.* 2005b; Parsons *et al.* 2005). Obtaining quantitative diet composition data from DNA-based studies could be possible using frequency of occurrence data from a large number of samples, but this data can be misleading if prey occurring in faecal samples are consumed in different amounts. Estimates could be improved through quantification of the DNA present in faeces if the amount of DNA from prey species in the samples is proportional to their mass in the diet (Deagle *et al.* 2005b; Kvitrud *et al.* 2005). Through the analysis of PCR clone libraries, I previously found that the proportions of fish DNA in faeces from captive sea lions were roughly proportional to the mass of the prey items consumed (Deagle *et al.* 2005b). While these results indicate some quantitative signature is present, they were based on the analysis of a small number of samples, and they also show an apparent bias in the relative amount of DNA recovered from different fish species. This bias could be an artefact of the clone library method used – caused by variation in amplification efficiencies between different prey species (von Wintzingerode *et al.* 1997). Alternatively, the bias could be due to species-specific differences in DNA density of the prey, or differential DNA survival during digestion. If differences in prey DNA density can explain the bias, it would be feasible to develop and apply prey-specific correction factors. If there are differences in prey DNA survival during digestion, it

may still be possible to develop experimentally derived correction factors but this would be more difficult.

Here, I use quantitative real-time PCR (qPCR) to investigate these issues. This method can be applied to measure the amount of DNA from a target of interest in dilute DNA samples (Miller *et al.* 2003). The technique measures the fluorescence produced by an increase in number of DNA copies during PCR amplification. The time taken for the fluorescence to increase to a threshold level is directly related to the initial amount of DNA present in the sample (Wittwer *et al.* 1997). Comparisons with a standard curve, generated with known amounts of starting DNA template, allow the amount of DNA in samples to be determined. Once suitable standards and species specific primers have been developed, the use of a qPCR approach to quantify the amount of DNA in faecal samples is much faster than the clone library method used in the previous chapter. In addition, potential differences in amplification efficiencies between the DNA from different prey species are corrected for when using this method.

In this chapter, I determined the relative amounts of mitochondrial DNA in a blended tissue mix representative of the diet being fed to captive Steller sea lions (E*umetopias jubatus*) (i.e. what goes in). I then determine the amount of DNA recovered from these prey items in the sea lions' faeces (i.e. what comes out).

4.2 Materials and Methods

4.2.1 Faecal and tissue samples

The sea lion faecal samples were collected during the feeding trial outlined in the previous chapter (Deagle *et al.* 2005b). The 23 samples analysed here are separate faecal samples from two captive sea lions which consumed a constant diet fed in two equal daily meals (i.e. independent blended samples, collected when animals were fed the basic diet; Figure 3.2). The fish component of the diet consisted of 50% Pacific herring (*Clupea pallasii*), 36% surf smelt (*Hypomesus pretiosus*), and 14% sockeye salmon (*Oncorhynchus nerka*) by mass. Faecal samples were blended to break up separate lobes and preserved in 95% non-denatured ethanol. DNA was extracted from

the amorphous soft matrix using the using the QIAamp® DNA Stool Mini Kit (Qiagen) as described previously (Deagle *et al.* 2005b).

A tissue mix equivalent to the sea lions' diet was produced by mixing 150 grams of Pacific herring, 108 grams of surf smelt and 42 grams of sockeye salmon. Fish were taken from the frozen stock fed to the sea lions; the salmon had been cleaned with head and fins removed, the herring and smelt were whole. Samples were thawed and blended in a chopping blender, and the tissue was further homogenised in ethanol using a micro blender (CAT brand, model X520D). From this finely homogenised tissue, ten aliquots were removed (72 ± 7 mg) and DNA was extracted following the same procedure as for the faecal extractions. These samples were diluted to approximately 1 ng per μL before being used in qPCR.

4.2.2 Quantitative PCR

The quantity of extracted mtDNA was estimated using SYBR® Green based qPCR assays (Wittwer *et al.* 1997). For each of the three fish species I designed PCR primers that amplify products (65-69 bp in size) from the 3' region of the mitochondrial 16S rDNA gene: Herring-1F (5'-ACCAATCACGAAAAGCAGGT-3') and Herring-69R (5'-CGAAGACGTTTGTGCCAGTA-3'); Smelt-1F (5'-ACGTCAAACTTCCCCTTTCA-3') and Smelt-65 R (5'-CCAACCGAAGACAGGAGAGA-3'); Salmon 1F (5'-GGCAGATCACGTCAAAAAC-3') and Salmon 65R (5'-AGACATATGGGCTAGGGGTC-3'). The primers were designed with reference to alignments of the sequences from the sea lion and all three fish species in order to ensure they were specific to the target species.

Amplifications were run using the Chromo4™ detection system (MJ Research). The PCR mix (20 μL) consisted of 10 μL QuantiTect® SYBR® Green RT-PCR mix (Qiagen), 0.5 μM of each primer, 1 x BSA (New England Biolabs) and 4 μL template DNA (diluted 1:5). Thermal cycling conditions were: 94ºC for 15 min followed by 35 cycles of: 94ºC, 30 s / 55ºC, 30 s / 72ºC, 45 s; optical data were acquired following each 72ºC extension step. A sub-set of samples were separated on 1.8% agarose gels to confirm product size and to check for primer dimers.

To ensure accurate relative quantification of the DNA from the prey species I used a single recombinant plasmid containing the relevant 16S mtDNA region from

each of the target species as a common standard (Miller *et al.* 2003). The plasmid was constructed using conserved primers (16SfishF and 16S2R; Deagle *et al.* 2005b), modified to include restriction sites on the 5' end (Figure 4.1; primer sequences given in Appendix II). PCR products generated from the relevant genomic DNA were digested to produce cohesive ends and then ligated into the *Kpn*I/*Xho*I sites of the pCR®2.1 TOPO® cloning vector (Invitrogen). The resultant plasmid was sequenced to confirm the presence of one copy of each of the inserts and the concentration of plasmid DNA was determined by fluorescence of PicoGreen (Molecular Probes) in a PicoFluor fluorometer (Turner Designs).

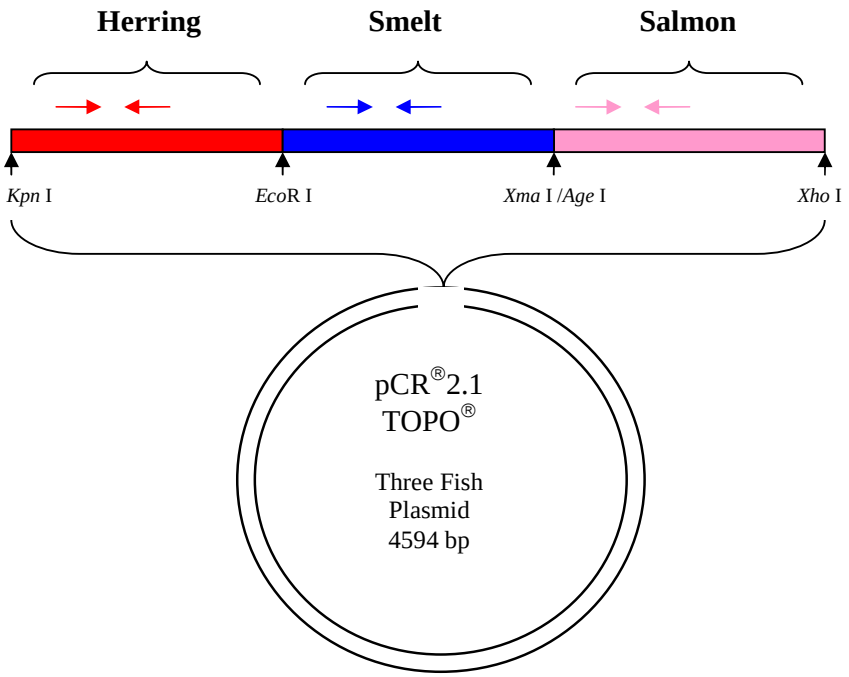


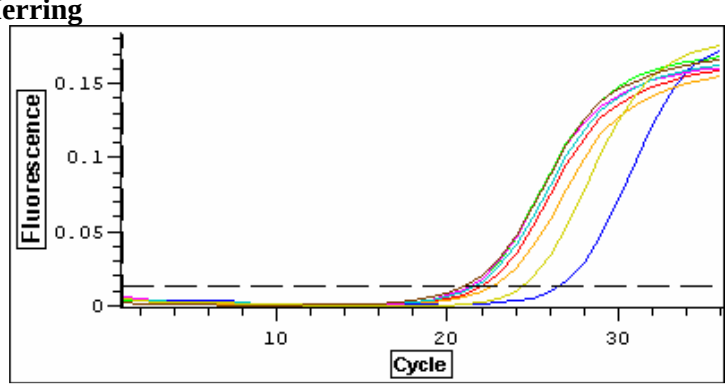
Figure 4.1 *Schematic of the plasmid used as a standard in qPCR. The plasmid insert contains mtDNA 16S gene fragments from three fish species ligated into the polylinker region of pCR®2.1 TOPO® (Invitrogen). Arrows represent the binding sites of the species-specific primers used during qPCR.*

Independent standard curves were generated within each PCR run using concentrations of the three-fish plasmid which encompassed the range of target (a 2-fold dilution series for tissue and a 5-fold dilution series for faecal DNA; determined in preliminary experiments). For individual extractions, DNA copy numbers for all

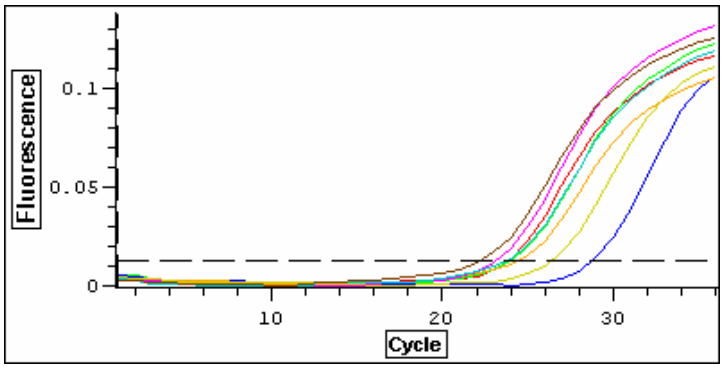
three target species were quantified in a single run (using a PCR reagent mix differing only in primer composition). Separate standard curves were initially constructed for each primer set; however, there were no consistent differences in these curves, so the data were pooled to create one standard curve per run based on a minimum of 15 reference points. To assess inter-run variability, two independent runs were carried out for each sample. Template free negative control reactions were included for each PCR mix within every PCR run to ensure that reaction batches were not contaminated. For quantitation, the threshold cycle (C_t) was set at 10 standard deviations above the mean fluorescence over cycle range 1-10 (Figure 4.2).

Likelihood ratio tests (LRT) (Hogg & Tanis 2005) were used to test whether the proportions of mtDNA in the tissue or faecal samples were statistically different than a specified set of proportions (e.g. the proportions of fish by mass in the diet). It was assumed that the proportions of mtDNA followed a Dirichlet distribution, which is the multivariate generalization of the beta distribution used to describe a set of proportions that sum to one (Kotz *et al.* 2000). The Dirichlet distribution was parameterized in terms of the expected proportions and an additional variance parameter. In calculating the likelihood under the null hypothesis, the variance parameter was allowed to be estimated freely. Thus, the LRT statistic has a Chi-squared distribution with k−1 degrees of freedom, where k is the number of proportions (in this case k = 3).

(a) Herring



(b) Smelt



(c) Salmon

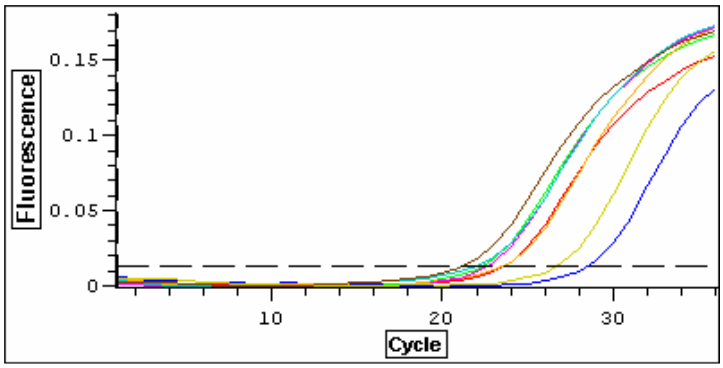


Figure 4.2 *Fluorescence measurements obtained during SYBR® Green real-time PCR amplification of DNA from three fish species (a,b,c) in eight representative faecal samples (each sample represented by a distinct colour). The dotted line on each plot shows the fluorescence level used for determination of the threshold cycle (C_t) for DNA quantitation, shifts of curve to the right indicate lower amount of DNA template in a sample. Amplifications were run using the Chromo4™ detection system (MJ Research).*

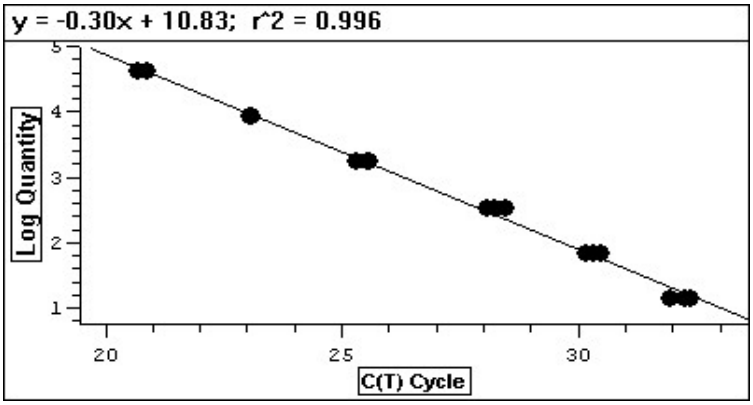


Figure 4.3 *Example of a quantitative real-time PCR standard curve generated through amplification of the three fish plasmid. Plot shows log of initial copy number (five fold dilution series, starting at approximately 1 $\times 10^5$) versus the PCR cycle at which the threshold fluorescence level was attained. Data from all three primers sets is plotted along with the common standard cure used for quantification.*

4.3 Results and Discussion

In the standard curves produced from the qPCR amplifications there was a linear relationship between the log of the plasmid DNA copy number and the C_t value over the concentration range of the standards (R^2 values ≥ 0.989; Figure 4.3). Within tissue samples, the copy number estimates were relatively consistent in the two replicates measurements (Table 4.1a). Overall, the percent composition of mtDNA in the tissue was 58.6 ± 4.6% for herring, 27.9 ± 2.2% for smelt and 13.5 ± 3.1% for salmon (mean ± sd, calculated using mean copy number data; Figure 4.4a and Table 4.2a). Compared to mass composition, herring proportions were somewhat overestimated (8-9%), smelt were correspondingly underestimated and salmon proportions were well matched. The tissue DNA proportions are statistically different from the mass proportions (LRT, P value < 0.001); nevertheless they indicate that the percent composition of fish mtDNA in undigested tissue is a reasonably good proxy for the relative mass of the corresponding fish (Figure 4.4a). One weakness of our methodology is that the tissue samples analysed are pseudoreplicates (i.e. DNA was extracted from sub-samples removed from a single tissue mix). Intra-specific

variation in tissue DNA density may exist, and it would be more appropriate to analyse independent tissue mixes in any future experiments.

In the faecal samples there was a large range in the total amount of fish DNA in different samples (848 – 150560 copies; Table 4.1b). There were also slightly larger errors in replicate qPCR measurements from individual samples, possibly due to the wide range covered by the standard curves. Despite this, the percentage composition of the fish DNA within the faeces was relatively consistent (57.5 ± 9.3% for herring, 19.3 ± 6.6% for smelt and 23.2 ± 12.2% for salmon; Figure 4.4b and Table 4.2b). These results match clone library estimates of the proportions of fish DNA in the faecal samples as determined in the previous chapter (61%, 19% and 20% respectively; Figure 3.6).

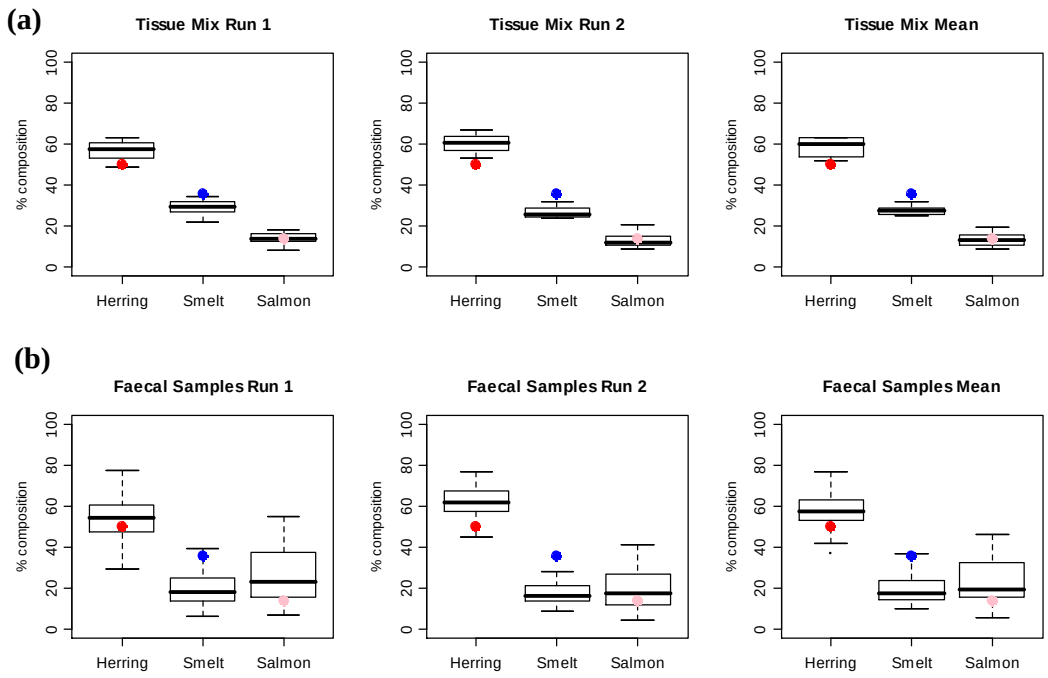


Figure 4.4 *Box plots showing a summary of the percentage DNA composition data (median, range and upper/lower quartiles) based on quantitative real-time PCR estimates:* ***(a)*** *results from tissue mixture (n = 10);* ***(b)*** *results from faecal samples (n = 23). Filled circles show percentage of each fish prey species (by mass) in the diet. Data for two independent runs and the mean values are shown.*

Table 4.1 *Estimated copy numbers of DNA template in each real-time PCR amplification (two replicate measurements).*

(a) Tissue Mix

Sample Number	Herring		Smelt		Salmon	
	Replicate 1	Replicate 2	Replicate 1	Replicate 2	Replicate 1	Replicate 2
1	66144	72683	29390	28062	13944	13304
2	23909	24026	11416	12177	7647	6068
3	77211	73359	35145	29778	10221	11837
4	25800	20041	9110	9852	6254	4145
5	68303	69280	37568	28094	15879	12677
6	83087	89079	41434	32177	12964	14991
7	55454	65952	25952	24176	12067	8844
8	18734	18009	11337	10818	5259	5196
9	22654	26951	15380	12279	8380	10293
10	16579	22448	11518	10853	5416	5972

(b) Faecal samples

Sample number	Herring		Smelt		Salmon	
	Replicate 1	Replicate 2	Replicate 1	Replicate 2	Replicate 1	Replicate 2
1	1343	1246	290	401	540	368
2	18787	32492	9424	9225	8430	5543
3	14763	21744	3681	3353	10511	11433
4	84590	77566	38252	41454	30456	28803
5	21347	35755	11850	14802	11337	8748
6	428	658	188	136	116	170
7	4105	4155	1060	1082	817	1056
8	26800	39488	9530	8724	17480	14461
9	5528	8983	2121	2208	7292	4688
10	621	783	441	464	83	59
11	959	1368	370	286	134	126
12	15870	23383	6627	6126	15179	9401
13	6772	6723	1708	1959	2250	2204
14	1681	2069	397	441	1356	1193
15	12531	19324	5076	3785	21657	13606
16	13570	13887	1951	4189	13226	12667
17	3420	4208	1417	1436	908	656
18	2508	2952	823	655	504	569
19	31792	33851	5237	6590	4044	4089
20	27548	58398	29084	15406	16932	9843
21	19421	28447	12077	10526	34086	24150
22	15124	13772	6395	6636	5235	2936
23	557	733	236	293	205	163

Table 4.2 *Percent composition of fish DNA in samples (calculated using mean copy number from the two qPCR runs).*

(a) Tissue Mix

Sample Number	Herring %	Smelt %	Salmon %
1	62.1	25.7	12.2
2	56.2	27.7	16.1
3	63.4	27.3	9.3
4	61.0	25.2	13.8
5	59.4	28.3	12.3
6	62.9	26.9	10.2
7	63.1	26.0	10.9
8	53.0	31.9	15.1
9	51.7	28.8	19.5
10	53.6	30.7	15.6
Mean	**58.6**	**27.9**	**13.5**

(b) Faecal samples

Sample Number	Herring %	Smelt %	Salmon %
1	61.8	16.5	21.7
2	61.1	22.2	16.7
3	55.7	10.7	33.5
4	53.9	26.5	19.7
5	55.0	25.7	19.3
6	64.0	19.1	16.9
7	67.3	17.5	15.3
8	56.9	15.7	27.4
9	47.1	14.0	38.9
10	57.3	36.9	5.8
11	71.8	20.2	8.0
12	51.3	16.7	32.1
13	62.4	17.0	20.6
14	52.5	11.7	35.7
15	41.9	11.7	46.4
16	46.2	10.3	43.5
17	63.3	23.7	13.0
18	68.2	18.4	13.4
19	76.7	13.8	9.5
20	54.7	28.3	17.0
21	37.2	17.6	45.2
22	57.7	26.0	16.3
23	59.0	24.2	16.8
Mean	**57.5**	**19.3**	**23.2**

If there is no differential digestion of mtDNA for the different fish species then the mtDNA proportions found in faeces should match the mtDNA proportions in the tissue mix. The data show this is not the case – the faecal proportions are significantly different from the mean proportions in the tissue (LRT, P value < 0.001). This is due to an overestimation of the proportion of salmon and an underestimation in the proportion of smelt mtDNA in the faecal samples. The data also indicate that correction factors based on tissue DNA density are useful (i.e. would allow for closer estimates of the proportions in the diet), even though they are not sufficient on their own to account for the biases observed in faeces (see Table 4.3). I also tried further adjusting the corrected faecal DNA proportions (Table 4.3) by applying the mean number correction factor (NCF) proposed by Bowen (2000) to account for variation in hard-part recovery: NCF=3.0 for Atlantic herring; 4.3 for Surf smelt; 1.6 for Chinook salmon. Interestingly, the percentage composition estimates become 50.3%, 36.6% and 13.1% – almost exactly matching the diet composition. This suggests that hard-part remains and DNA from a particular species may be affected by digestion in a similar manner, although further data will be required to see if this observation is a general phenomenon.

Table 4.3 *Summary of proportional composition data with faecal estimates corrected to account for differences in tissue DNA density.*

	Diet species (*i*)		
	Herring	Smelt	Salmon
Diet mass proportions (m_i)	0.50	0.36	0.14
Tissue DNA proportions (t_i)	0.586	0.279	0.135
Faecal DNA proportions (x_i)	0.575	0.193	0.232
Corrected faecal DNA proportions ($\hat{x}_i$) [a]	0.501	0.254	0.245

[a] Proportions corrected for differences in prey tissue DNA density using the formula:

$$\hat{x}_i = \frac{x_i c_i}{\sum_{i=1}^{n} (x_i c_i)} \quad \text{where} \quad c_i = \frac{m_i}{t_i} \quad \text{and} \quad n = \text{number of prey items}$$

Correction factors could be developed specifically for DNA by measuring the relative recovery rates of DNA from various prey items in further captive feeding trials. However, the recovery of hard parts has been shown to be affected by numerous variables, such as species of predator, activity level of predator and size of prey; see reviews by Bowen (2000) and Tollit *et al.* (2003). Therefore, investigating these variables in relation to DNA recovery would not be a simple task.

It is not clear what implications the observed bias in mtDNA recovery will have for the future of DNA quantification in molecular diet studies. All methods used to estimate pinniped diet have limitations and are characterised by relatively large amounts of error (Figure 4.5). Quantification of DNA in faecal samples can provide some informative data; for example, based on the qPCR results, herring was correctly predicted as the dominant prey in 21 of the 23 samples. In situations where large uncertainties surround conventional hard part faecal analysis (Gales & Cheal 1992; Casper *et al.* 2006), measuring prey DNA amounts in faeces may still provide a useful, albeit potentially somewhat biased, alternative metric. In the current study, I used species-specific primers to compare consumption of individual species, but more general comparisons could be made (e.g. amount of salmonid DNA versus total fish DNA) by using group-specific primers (Jarman et al. 2004). The proposed quantitative DNA-based faecal analysis could also be used to examine the diet of species whose faeces do not contain hard-parts (e.g. seabirds or cetaceans) and could be modified to quantify prey DNA in stomaches of invertebrates where few non-molecular approaches to studying diet exist (Symondson 2002).

Perhaps the biggest limitation to the application of this approach at present is the technical difficulty of quantifying DNA from multiple potential prey species within samples. In addition, prey DNA would need to be quantified in a large number of faecal samples to obtain an average view of diet. My previous suggestion, that DNA extracted from multiple faecal samples could be pooled to reduce the amount of laboratory analysis (Deagle *et al.* 2005b), should be reconsidered due to the large variation in the total amount of prey DNA in different faecal samples[10].

[10] If the cloning approach of Chapter 3 is used to estimate proportions of prey DNA, the PCR products amplified from individual faecal extractions could be quantified and mixed in equal concentrations before cloning. This would give equal influence to each sample in the pooled DNA clone library.

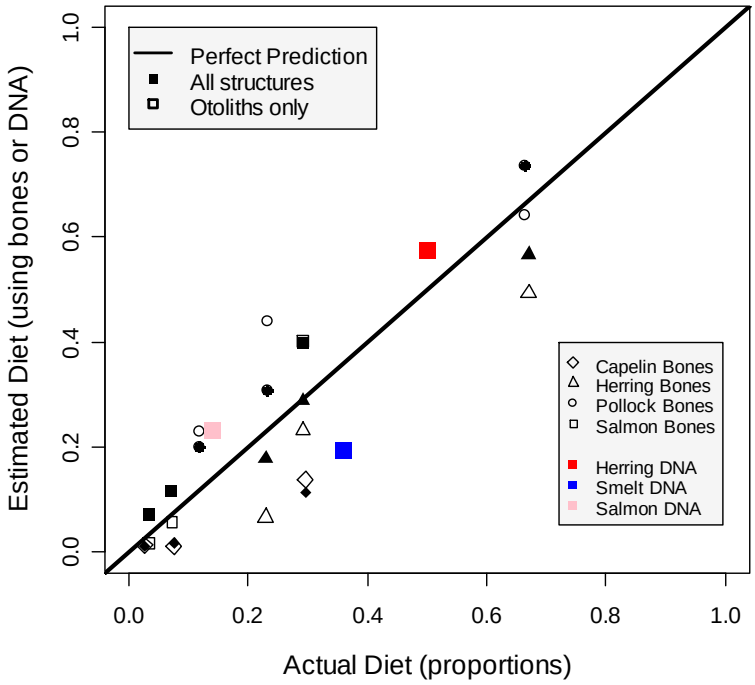


Figure 4.5 *Plot of actual diet versus estimated diet using various biomass reconstruction methods. Estimates for various prey species are based on recovery of all hard-parts (solid black symbols), recovery of otoliths only (unfilled symbols), or uncorrected DNA proportions in faeces (coloured squares). All data are from captive Steller sea lions and hard part data includes regurgitated hard remains collected during tank drains. Figure modified from Tollit et al. (2006).*

The differential survival of DNA from different fish species will not necessarily translate into significant biases in quantitative diet estimates obtained based on presence/absence genetic data (since the results show prey DNA can still be detected when present at low levels). The potential benefits of using DNA quantification, versus a DNA-based frequency of occurrence approach, will depend on the composition of faecal samples collected in the field. If most faeces contain DNA from only one species or similar amounts of DNA from several species, then presence/absence data would provide nearly the same information as a quantitative analysis. However, if the opposite scenario is true, and most faecal samples contain DNA from several prey species consumed in different amounts, quantification of the relative amount of DNA becomes much more important. In a diet study focussed on the impact of pinnipeds on salmonid prey, Laake *et al.* (2002) highlighted tenfold

differences in diet estimates across species when comparing frequency of occurrence and biomass reconstruction indices. Differences were strongly related to size of prey, with differences greatest when prey were small or large. Reliable diet estimates based on biomass reconstruction require numerical correction factors to take into account differential digestion, but reliable estimates are not available for many species (Bowen 2000). Thus, in cases where pinnipeds are preying upon large prey such as salmon and are also consuming other smaller prey, the ability to estimate diet composition using alternate methods, in addition to those based on frequency of occurrence, are especially beneficial.

The use of genetic analysis of faeces for studying diet of pinnipeds and other predators is likely to increase since the approach provides species-level identification of prey (not always possible using prey hard remains) and because it is possible to detect prey species that are poorly represented by hard remains in faeces (Purcell *et al.* 2004; Parsons *et al.* 2005). This chapter demonstrates the feasibility of quantifying the amount of DNA from several prey species in predators' faeces and evaluates the possibility of using this data to obtain quantitative diet composition data. The results indicate that diet composition estimates based on the relative amounts of DNA in pinniped faeces can be biased due to the differential survival of DNA from different fish species during digestion. Additional DNA-based diet studies with both captive and wild animals will be needed to evaluate the significance of this bias and further evaluate the potential to obtain quantitative diet estimates through the genetic analysis of faeces.

Chapter 5

Quantification of damage in DNA recovered from faecal samples

"You want a story without animals."
"Yes!"
"Without tigers or orang-utans."
"That's right."
"Without hyenas or zebras."
"Without them."
"Without meerkats or mongooses."
"We don't want them."
"Without giraffes or hippopotamuses."
"We will plug our ears with our fingers!"
"So I'm right. You want a story without animals."

— YANN MARTEL,

Life of Pi

Abstract

A novel method to quantify the frequency of polymerase-blocking DNA damage present in specific gene-regions is outlined. Unlike conventional PCR-stop assays this is accomplished without reliance on a corresponding undamaged DNA control or a dose-response curve. The approach involves using quantitative PCR to measure the amount of DNA present at several fragment sizes within a single sample, then fitting the resultant data to a model of random fragmentation to obtain an estimate of the frequency of DNA damage (λ). The method was assessed by estimating the amount of DNA damage in two components of DNA extracted from the Steller sea lion faecal samples these faeces: prey DNA (expected to be highly degraded) and predator DNA (expected to be slightly less degraded). The distribution of fragment lengths for each target fit well with the assumption of a random fragmentation process and, in keeping with expectations, the estimated frequency of DNA damage within a sample was always higher for the prey DNA than for the predator DNA (mean $\lambda_{prey} = 0.0176$ per nucleotide; mean $\lambda_{predator} = 0.0106$ per nucleotide). The results clarify the relative nature of template quantity measurements obtained when analysing degraded templates using qPCR (i.e. the estimated amount of DNA will vary with marker size in a sample-specific fashion). In addition, the findings highlight the benefit of targeting small fragment sizes in DNA-based diet studies. For example, in Chapter 2, I targeted a 250 bp prey DNA fragment – by targeting a 150, 100 or 60 bp long fragment I would have had 6, 15 or 31 times more starting template respectively (based on the average amount of damage estimated in herring DNA). More generally, this method provides a useful approach for characterizing mixed, highly degraded PCR templates such as those often encountered in forensics, ancient DNA research and ecological studies using non-invasive samples as a source of DNA.

5.1 Introduction

It has become common practice to use DNA derived from poorly preserved, decomposed or ancient tissue sources in evolutionary, forensic, ecological and medical studies (Golenberg *et al.* 1996; Coombs *et al.* 1999; Glenn *et al.* 1999; Pääbo *et al.* 2004). Often only small amounts of DNA can be extracted from such samples and it is invariably highly damaged. In the absence of normal cellular processes, DNA strand breakage rapidly begins to occur as a result of endogenous endonuclease activity and spontaneous depurination (Lindahl 1993). Depending on the ambient conditions, further strand breaks, oxidative damage and molecular crosslinks accumulate (Pääbo 1989; Höss *et al.* 1996; Mitchell *et al.* 2005). Assessing the extent of damage is difficult, especially when the DNA of interest is present in a sample containing DNA from several different sources. However, determining DNA quality is desirable in many situations, as reflected by the variety of approaches that have been used to measure DNA damage (Pääbo 1989; Govan *et al.* 1990; Cedervall *et al.* 1995; Glenn *et al.* 1999; Ayala-Torres *et al.* 2000; Hoogendoorn & Heimpel 2001; Fernando *et al.* 2002; Gilbert *et al.* 2003; Wandeler *et al.* 2003; Mitchell *et al.* 2005).

Qualitative estimates of DNA fragment sizes can be obtained through gel electrophoresis followed by visualisation of fragments (Pääbo 1989; Marota *et al.* 2002). This approach is simple but has limited sensitivity and, because it does not differentiate between fractions of the DNA extractions, it is generally only useful if all DNA present has been equally degraded. Furthermore, interpretation of fragment size distribution on a gel is hindered because DNA fragments separate on a logarithmic scale and the detection signal typically varies in proportion to fragment size.

Another approach that has been used to assess DNA quality is through observations of the decrease in PCR amplification signal from PCR targets of increasing sizes (Pääbo 1990; Glenn *et al.* 1999; Poinar *et al.* 2003). Since many forms of DNA damage block the extension step of PCR (Höss *et al.* 1996; Gilbert *et al.* 2003), the ability to recover large fragments via PCR indicates relatively low levels of DNA damage. By determining the maximum amplifiable fragment size in different samples it is possible to compare relative amounts of DNA degradation (Hoogendoorn & Heimpel 2001). There are several related PCR-based methods used to measure DNA damage incurred by exposure to mutagenic compounds (Jennerwein & Eastman 1991; Ayala-Torres *et al.* 2000; Fernando *et al.* 2002; Mambo *et al.* 2003).

These techniques, often called PCR-stop assays, measure gene-specific damage by quantifying the decrease in the number of molecules that can be amplified following a particular genotoxic treatment. A limitation of the currently used PCR-stop assays is that the total amount of target DNA has to be quantified using PCR independent means, or else a dose-response curve needs to be constructed. This precludes their use in a number of situations.

Here, I extend the existing PCR-based methods by proposing an experimental strategy that can be used to quantify gene-specific DNA damage in dilute samples without reference to the total amount of DNA in the sample. The approach uses quantitative PCR (qPCR) to measure the amount of amplifiable DNA present in a single sample for fragments of various sizes. If DNA degradation is occurring in a random fashion, it is possible to estimate the frequency of polymerase blocking DNA damage (or DNA breakage) that would explain the observed distribution of DNA fragment sizes.

In order to assess the proposed method, I estimate the frequency of DNA damage in DNA extracted from faeces. Faeces contain DNA from a variety of sources, including DNA from the defecating animal, ingested food, parasites and gut bacteria (Poinar *et al.* 2001; Jarman *et al.* 2004). Of particular interest to ecologists is the DNA from ingested food, which can be used to study diet (Symondson 2002), and the DNA from the defecating animal, which can be used as a non-invasive source of DNA from wild species (Taberlet *et al.* 1996; Morin *et al.* 2001). The different components of faecal DNA are expected to sustain varying amounts of damage. DNA from animal food sources are expected to be highly degraded since these tissues are usually fully digested after passing through the complete digestive system. In comparison, DNA from the defecating animal should be slightly less degraded because this component largely originates from cells shed along the lower digestive tract. I examine both prey and predator DNA extracted from faeces of captive Steller sea lions (*Eumetopias jubatus*). In each case the amount of DNA is quantified using primer sets that amplify fragments of five different lengths (ranging between 61 and 327 bp). I evaluate if a model of random degradation fits the data and then estimate the frequency of damage in DNA from each of the targeted components.

5.2 Materials and Methods

The method used to estimate the level of DNA degradation is outlined in Figure 5.1. Specific details are given in the sections that follow.

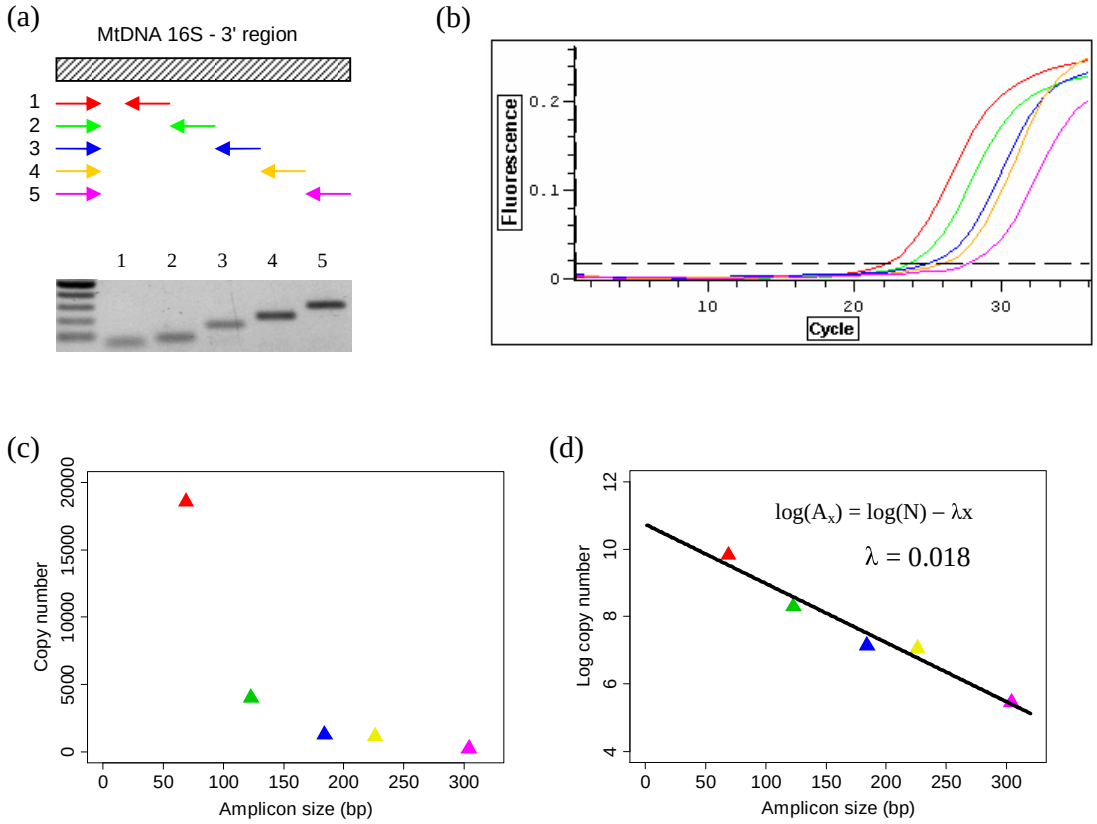


Figure 5.1 *Overview of the approach for quantification of DNA damage. (a) Schematic representation showing position of oligonucleotides designed to amplify fragments of increasing sizes and the corresponding PCR products amplified from genomic DNA (separated on a 1.8% agarose gel). (b) Representative plot of fluorescence observations from the quantitative PCR amplification of various sized fragments of herring DNA. Template for each is the same sea lion faecal DNA extraction. PCR fragment sizes from left to right are 69 bp, 123 bp, 184 bp, 226 bp and 304 bp. (c) Plot of the estimated copy number versus amplicon size for herring DNA in a sea lion faecal sample (# 7). (d) The same data log-transformed and fitted to a linear model in order to estimate the probability of a nucleotide being damaged (λ).*

5.2.1 DNA Samples

The sea lion faecal samples are a subset of those from Chapter 3. Ten samples were analysed for endogenous DNA from sea lion and Pacific herring (*Clupea pallasii*). These samples were collected from captive sea lions being fed a diet consisting of 47% herring by mass for a period of at least 48 hours before collection (i.e. blended or sub-samples from different days, collected when animals were fed the basic diet; Figure 3.2) and were previously shown to contain herring DNA. Three additional sea lion faecal samples were collected from animals being fed a diet of 100% walleye pollock (*Theragra chalcogramma*). These samples were used in spiking experiments for the analysis of length inhibition (see section below). Sample storage and DNA extraction for the sea lion samples has been described in Chapter 3.

5.2.2 Primer design

PCR primers which amplified fragments from the 3' region of the mitochondrial 16S gene (large subunit rDNA gene) were designed using Primer3 (Rozen & Skaletsky 2000). A common forward primer was selected for use with five reverse primers, producing products in the range of 60-350 bp (Figure 5.1a). Primer sequences and product sizes are given in Table 5.1. The forward primers were designed with reference to aligned sequences from additional fish species that were present in the sea lion diet to ensure they were specific to the target species. The specificity of the primer sets was also tested empirically against non-target DNA that may have been present in the faecal extracts. The herring primer sets were tested against the three faecal extracts from sea lion fed only pollock, and the sea lion primers against herring genomic DNA.

Table 5.1 *Primer sequences used to quantify DNA degradation.*

Target	Forward Primer (5' → 3')	Reverse Primers (5' → 3')	Product Length (bp)
Sea Lion	CAAGTCAACCAAAACGGGATA	CACCCCAACCTAAATTGCTG	61
		TCACTCGGAGGTTGTTTTGTT	91
		CTTGTTCCGTTGATCAAAGATT	163
		TCGAGGTCGTAAACCCTGTT	230
		GATTGCTCCGGTCTGAACTC	327
Herring	ACCAATCACGAAAAGCAGGT	CGAAGACGTTTGTGCCAGTA	69
		TAGGGTAGCCCCAATCCTCT	123
		GCATGTAGCCGGATCATTTT	184
		GGATTGCGCTGTTATCCCTA	226
		AATAGCGGCTGCACCATTAG	304

5.2.3 Quantification of mtDNA

The quantity of extracted 16S mtDNA was estimated using SYBR® Green based qPCR assays. Amplifications were run using the Chromo4™ detection system (MJ Research). The PCR mix (20 μL) consisted of 10 μL QuantiTect® SYBR® Green PCR mix (Qiagen), 0.5 μM of each primer, 1 x BSA (New England Biolabs) and 4 μL template DNA (diluted 1:5). Thermal cycling conditions were: 94ºC for 15 min followed by 35 cycles of: 94ºC, 30 s / 55ºC, 30 s / 72ºC, 45 s; optical data was acquired following each 72ºC extension step (Figure 5.1b). A subset of samples was separated on 1.8% agarose gels to confirm product size and to check for primer dimers. Product homogeneity was also confirmed by melting curve analysis.

A plasmid standard encompassing the relevant 16S mtDNA region was generated from genomic DNA for each target species. This was accomplished by amplifying the region using the conserved primers (16S1F GGACGAGAAGACCCT and 16Sbr-3' CCGGTCTGAACTCAGATCACGT) and cloning the PCR products using the TOPO TA cloning kit (Invitrogen). Plasmid DNA was isolated by alkaline lysis and the concentration of plasmid DNA was determined by fluorescence of PicoGreen (Molecular Probes) in a PicoFluor fluorometer (Turner Designs). Standard curves were generated using a 5-fold dilution series of plasmid encompassing the

concentration range of the faecal template. Separate standard curves were constructed for each of the different sized PCR amplifications and for each target species. Independent curves were calculated during each PCR run. The binding site for the 327 bp sea lion reverse primer was incomplete in the plasmid control, so quantification of this DNA fragment was based on the standard curve generated for the 230 bp sea lion fragment.

For individual extractions, the complete set of fragment sizes for a particular target was quantified in a single run (using a PCR reagent mix differing only in primer composition). This meant variation in reaction conditions between the different sized fragments that were being compared was minimized. Two independent runs were carried out for each assay. For quantitation, the threshold cycle (C_t) was set at ten standard deviations above the mean fluorescence over cycle range 1-10. To avoid contamination with undamaged DNA, faecal DNA template was added to tubes first and their caps were sealed before plasmid DNA was added to appropriate standard tubes in a separate room. Aerosol-resistant pipette tips were used with all PCR solutions, and template free negative control reactions were included for each PCR mix within every PCR run to ensure that reaction batches were not contaminated.

5.2.4 Analysis of length-inhibition

A decrease in PCR signal with an increase in the length of the product could result from the absence of high molecular weight DNA in samples, or could result from selective inhibition of longer PCR amplifications caused by the coextraction of inhibitory chemicals, as found in some purified ancient DNAs (Pusch *et al.* 2004). To determine if extract induced inhibition was a factor in the current study, I carried out spiking experiments. Known amounts of herring DNA (3380 or 13520 copies of the plasmid control) were added to DNA extracted from sea lions faeces which contained no herring (n = 3) and the amount of recoverable herring DNA of the five sizes was estimated as outlined above.

5.2.5 Model for quantitative estimates of DNA damage

DNA damage resulting in strand breaks or chemical modifications that would prevent PCR amplification can be caused by a number of mechanisms. I assume that in highly degraded samples such DNA damage occurs according to a random Poisson process at a rate of λ per nucleotide (i.e. λ is the probability of a nucleotide being damaged). The resulting distribution of undamaged fragment sizes (x) is defined by an exponential distribution with parameter λ:

$$f(x) = \lambda e^{-\lambda x} \quad \textbf{1}$$

This model has been used to characterise DNA damage induced by some mutagenic agents (Ayala-Torres *et al.* 2000; Fernando *et al.* 2002), and a very similar model has been used to describe random fragmentation resulting from DNase I digestion (Moore & Maranas 2000). It follows from the properties of an exponential distribution that the average undamaged fragment size is $1/\lambda$, and that the variance of undamaged fragment sizes is $1/\lambda^2$.

Since PCR will amplify any DNA which is undamaged in a region equal to or greater in size than the target region, it is appropriate to consider the probability of a fragment of size x or greater being present. This is given by $e^{-\lambda x}$, the complement of the cumulative exponential distribution. In a PCR designed to amplify a target region of size x, the expected proportion of amplifiable copies is $e^{-\lambda x}$. Thus, there is an exponential decline in the amount of amplifiable product with increasing product size, and the rate of decline is sharper for higher values of λ (Figure 5.2). It follows that if the total DNA copies present in the sample is N, then the expected number of amplifiable copies, denoted by A_x, is $Ne^{-\lambda x}$ (Figure 5.1c). Using a logarithmic transformation this relationship can be expressed in linear form as:

$$\log(A_x) = \log(N) - \lambda x \quad \textbf{2}$$

The actual number of amplifiable copies for a given size (A_x) will vary due to the random nature of the degradation process. If the process is truly Poisson, then the

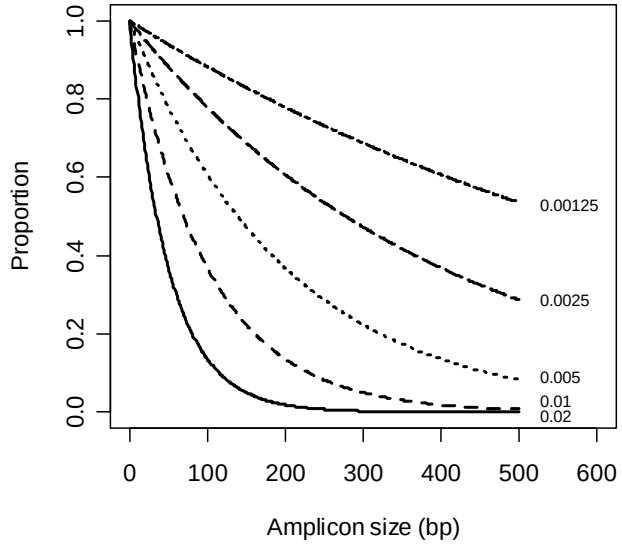


Figure 5.2 *The proportion of amplifiable fragments versus amplicon size after a random degradation process with the probability of a nucleotide being damaged (λ) of: 0.00125, 0.0025, 0.005, 0.01 or 0.02.*

amount of variance can be calculated theoretically – in theory, A_x is binomially distributed with sample size N and 'success' probability $e^{-\lambda x}$. Thus, the variance is $Ne^{-\lambda x}(1-e^{-\lambda x})$. However, the variability observed in practice is expected to be greater because the degradation process is not likely to follow a Poisson process exactly and, even if it did, there will be experimental measurement error. Here I assume that the error in $\log(A_x)$ is normally distributed with mean 0 and variance σ^2, this is consistent with previous studies (Ayala-Torres *et al.* 2000). Assuming this error structure, model (**2**) can be fit using simple least-squares regression (Figure 5.1d).

For each of the ten sea lion faecal samples, I obtained two estimates of copy number (A_x) corresponding to five fragment sizes (x) for both sea lion (predator) DNA and herring (prey) DNA. I fit model (**2**) to the data from each sample and target species to obtain estimates of log(N) and λ, with λ being the parameter of key interest. Coefficients of variation for the parameter estimates and R^2 values were also obtained for each of the model fits.

5.3 Results

5.3.1 Primer testing and DNA quantification

Electrophoresis of PCR amplifications from target genomic DNA extracts produced sharp bands of the expected sizes from all primer sets with no primer dimers visible (Figure 5.1a). None of the primers used in the quantitative assays amplified products from the non-target templates tested, and melting curve analysis performed on the products indicated that each primer set produced a single product from the mixture of DNA present in the faecal extractions.

For each of the quantitative PCR assays there was a linear relationship between the log of the plasmid DNA copy number and the C_t value over the concentration range of the standards (mean $R^2 = 0.994$). As expected with degraded DNA template, the amount of amplifiable DNA was inversely related to PCR product size for all targets amplified from the faecal DNA templates (Figure 5.1c; Table 5.2). The faecal extracts contained on average eight times more sea lion DNA than herring DNA at the smallest fragment sizes measured and 52 times more sea lion DNA than herring DNA at the largest fragment size. Expressed as copy number per milligram of extracted faecal matter, the samples contained on average 15109 copies (range 418 – 35498) of the 69 bp herring fragment compared to 123209 copies (range 19398 – 281880) of the 61 bp sea lion fragment, and 173 copies (range 38 – 395) of the 304 bp herring fragment compared to 8917 copies (range 692 – 26676) of the 327 bp sea lion fragment. The large variation in the amount of predator and prey DNA obtained from different faecal samples is consistent with another study that has quantified predator DNA in faeces (Morin *et al.* 2001). There was no clear relationship between the amounts of sea lion DNA and herring DNA purified from individual samples (Table 5.2).

Table 5.2 *Estimated copy numbers of template in each PCR amplification*[11] *and results from the model fits.*

(a) Sea lion DNA

Sample	Mean copy number at various amplicon sizes					Model parameters			
	61 bp	**91 bp**	**163 bp**	**230 bp**	**327 bp**	**λ [a]**	**CV(λ)**	**$1/\lambda$ [b]**	**R^2**
1	44727	22461	11965	3746	1387	0.0129	10.3	78	0.92
2	24347	15968	5393	1487	525	0.0148	6.9	67	0.96
3	236825	193410	121534	44745	25936	0.0088	9.1	113	0.94
4	35346	30846	20383	9914	5390	0.0074	10.9	135	0.91
5	43789	26107	11972	4232	1276	0.0135	7.6	74	0.96
6	179118	126256	57217	26322	8684	0.0115	7.9	87	0.95
7	167038	123458	62726	17076	7944	0.0121	7.0	83	0.96
8	29683	25006	14941	9007	2818	0.0089	8.7	113	0.94
9	198820	222720	146709	53608	19759	0.0093	11.0	108	0.91
10	58503	46534	34979	25738	9624	0.0066	11.7	153	0.90
Mean	101820	83277	48782	19588	8334	0.0106	9.1	101	0.94

(b) Herring DNA

Sample	Mean copy number at various amplicon sizes					Model parameters			
	69 bp	**123 bp**	**184 bp**	**226 bp**	**304 bp**	**λ [a]**	**CV(λ)**	**$1/\lambda$ [b]**	**R^2**
1	8157	522	143	93	42	0.0221	15.9	45	0.83
2	9927	994	257	240	120	0.0192	19.7	52	0.76
3	14005	2016	687	381	178	0.0179	11.2	56	0.91
4	1575	418	246	182	69	0.0123	10.9	81	0.91
5	11649	952	243	149	46	0.0222	12.5	45	0.89
6	11754	1700	605	307	120	0.0188	9.7	53	0.93
7	18588	4016	1272	1165	234	0.0180	11.7	56	0.90
8	25203	8883	2306	1366	429	0.0173	6.0	58	0.97
9	26295	5497	1299	655	266	0.0197	9.2	51	0.94
10	711	432	238	138	107	0.0086	11.9	117	0.90
Mean	12786	2543	729	468	161	0.0176	11.9	61	0.89

[a] λ is an estimate of the probability of a nucleotide being damaged
[b] $1/\lambda$ is an estimate of the average undamaged fragment size (in base pairs)

[11] Mean values from replicate runs are presented here, values for separate runs given in Appendix III.

5.3.2 Length-specific inhibition

There was no evidence of inhibitory effects caused by chemicals in the three sea lion faecal DNA extracts that I tested from sea lions fed only pollock. None of the samples contained endogenous herring DNA, and when spiked with known amounts of herring DNA the amount of DNA measured by the assay was not underestimated. This was true for all fragment lengths in the assay indicating there was no length-specific PCR inhibition (Figure 5.3).

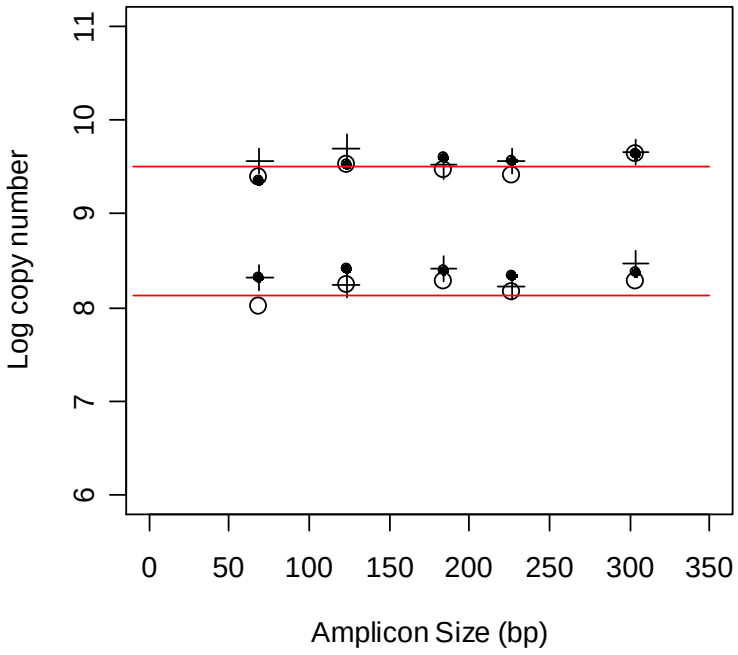


Figure 5.3 *Quantitative estimates of the amount of amplifiable herring DNA in three spiked faecal DNA extractions. Horizontal lines show the actual amount of herring DNA added as template (either 3380 or 13520 copies of the plasmid control). Symbols represent the corresponding estimates of the amount of herring DNA in three samples measured with assays targeting PCR products of five different sizes (69 bp, 123 bp, 184 bp, 226 bp and 304 bp).*

5.3.3 Model results

Results from fitting the random degradation model (**2**) to the data for each sample and target species indicate that the model describes the data well, with R^2 values generally above 0.90 (Figure 5.4; Table 5.2). In almost all cases there was a high level of agreement in the data from the two runs (Figure 5.4), and where the agreement was not as good is reflected in lower R^2 values (e.g. herring in samples 1 and 2).

The estimated probability of a nucleotide being damaged (λ) varied between samples for a given target species (0.0066 to 0.0148 for sea lion DNA; 0.0086 to 0.0222 for herring DNA); however, within a sample, the λ estimate for herring was always greater than that for sea lion (Figure 5.4). On average, the frequency of damage was 1.7 times greater for the herring DNA compared with the sea lion DNA from the same sample; a paired t-test indicates the difference in λ values is significant ($t = 8.4$ with 9 df, $p < 0.001$). The mean fragment size in each sample can be estimated by $1/\lambda$ (Table 5.2). Averaging over all samples, the mean fragment size for the herring DNA is 61 bp versus 101 bp for the sea lion DNA.

There is no clear relationship between amount of DNA (log(N)) and level of degradation (λ). Correlation between log(N) and λ for sea lion is -0.06; for herring it is 0.76 but this is being driven by two samples (4 and 10) with very low amounts of herring DNA. Leaving these two points out gives a correlation of -0.53, which is not only weaker but is also a change in direction.

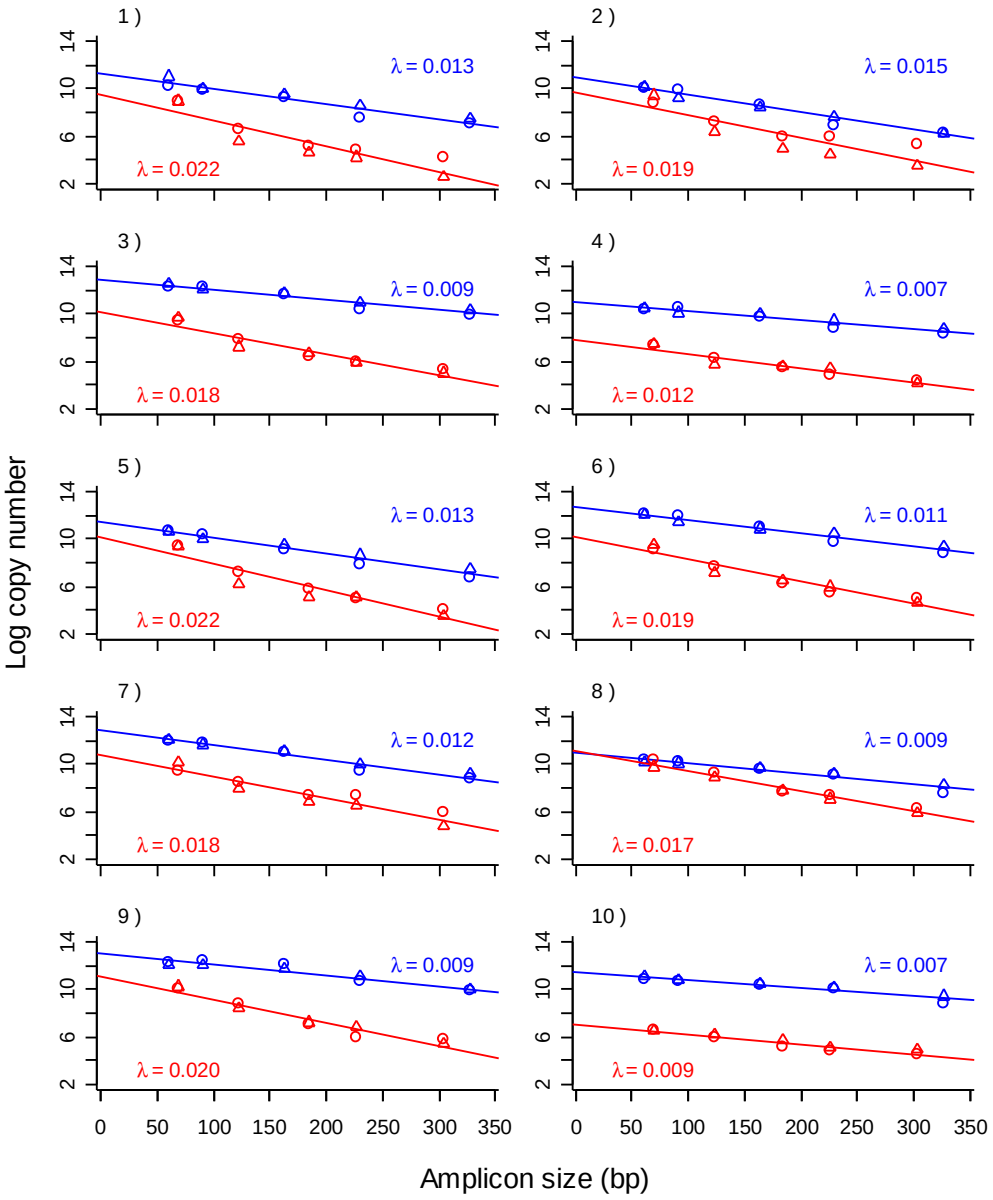


Figure 5.4 *Quantitative PCR results obtained for herring DNA (red) and sea lion DNA (blue) extracted from ten sea lion faecal samples. Shown is the number of amplifiable copies (logarithmic scale) verus amplicon size for each target species in each sample. The estimated probability of a nucleotide being damaged (λ) are also shown for each target species in each sample.*

5.4 Discussion

In this chapter a novel PCR-based method for quantitative measurement of DNA damage is presented. The approach circumvents a problematic requirement of conventional PCR-stop methodology – the need to know the total amount of target DNA (damaged and undamaged) present in a sample. I make use of internal references by measuring the amount of target DNA present at several fragment sizes within a single sample using quantitative PCR. According to a model of random degradation the amount of available template will decline exponentially with increasing fragment size in damaged samples. The amount of DNA damage can be estimated by determining the rate of decline. This approach could be applied in a wide range of projects where quantification of DNA damage is required. The method will be particularly useful in experimental work dealing with highly damaged DNA from biological samples that contain DNA from a several sources since alternative methods are not easily applied. These types of samples are commonly encountered in forensics, ancient DNA research and ecological studies using non-invasive samples as a source of DNA.

The approach was assessed by estimating the amount of DNA damage in two components of DNA extracted from sea lion faeces: prey DNA (expected to be highly degraded) and predator DNA (expected to be slightly less degraded). The distribution of fragment lengths in these faecal DNA templates fit well with the assumption of a random fragmentation process, and differences in the estimated frequency of predator versus prey DNA damage within samples were congruent with expectations. To my knowledge, this study is the first to quantify the amount of prey DNA present in faecal samples (with the exception of the previous chapter), and the first to explicitly define the amount of template damage in any DNA extracted from faeces. The results clarify the relative nature of template quantity measurements obtained when analysing degraded templates using qPCR (i.e. the estimated amount of DNA will vary with marker size in a sample-specific fashion). This means that comparisons of DNA quantity (within and between samples) are dependent on the size of the fragments targeted by qPCR. This can have important implications – for example, a previous study (Morin *et al.* 2001) used qPCR targeting an 81 bp nuclear gene fragment in order to determine the amount of chimpanzee nuclear DNA present in faeces collected

from wild chimpanzees. One of the purposes of their assay was to predict the recovery of chimpanzee microsatellite markers which were 101-266 bp in size. They found that in samples where the measured amount of DNA was low, the quantity of DNA measured with the 81 bp marker was not a good indicator of the ability to recover microsatellites. This finding was likely due to differences in the amount of DNA degradation between samples; if they had measured the amount of DNA present at a size equal to the fragments amplified during the microsatellite genotyping the predictive power of their pre-screening assay would likely have been stronger.

The quantitative estimates show that there is less prey DNA compared to predator DNA in sea lion faeces for all PCR fragment sizes tested. Previous studies (Taberlet *et al.* 1996) have found the low quantity of predator DNA in faeces problematic, which suggests that the even more limited amount of prey DNA may be a serious difficulty for DNA-based diet studies relying on faecal samples. Fortunately, multi-copy nuclear or mitochondrial genes are usually appropriate markers for diet studies (as opposed to the single-copy markers which are often targeted for studies on the predator); this advantage may allow for reliable recovery of prey-specific DNA sequences from faecal samples.

When the amount of available template is low, knowledge of template quality is important to allow for an objective appraisal of optimal PCR target size. The appropriate size of a PCR target is a trade off between the amount of information obtained from the DNA (directly related to fragment size) and the quantity of template DNA available (inversely related to fragment size). With no knowledge of template quality, I previously amplified a 250 bp fragment in order to detect fish DNA in faecal samples (Deagle *et al.* 2005b). Because of the low amount of DNA, I needed to use nested PCR to consistently obtain strong signals. In light of the amount of DNA damage in faecal samples, I would have had 6, 15 or 31 times more starting template if I had used amplicons which were 150, 100 or 60 bp long respectively (based on the average amount of damage estimated in herring DNA). Since even the smallest amplicon size would have provided suitable taxonomic discrimination for my purpose, I potentially could have avoided the use of nested PCR. Non-invasive studies focusing on DNA from the predator in faecal samples should also preferentially use short amplicons in order to maximize amplification success. However, other marker characteristics (such as information content of the fragment or the repeat motif of microsatellites) should also be considered when planning experiments.

In ancient DNA work it has been recognised that an assessment of both the amount of DNA present and the amount of damage present is useful in order to define the limits of subsequent analyses and the authenticity of the sample (Pääbo 1989; Mitchell *et al.* 2005). The only ancient DNA study I am aware of which has quantified the number of fragments of different lengths present in a sample is one which analysed three different fragments (114 bp, 252 bp and 522 bp) of sloth mtDNA from a late Pleistocene sloth coprolite (Poinar *et al.* 2003). The largest fragment size they used (522 bp) contained on average only 0.5 ± 0.5 copies of DNA. With almost zero copies and such a large relative error, this point is not informative for quantification of damage; however, based on the other two points I can estimate the probability of a nucleotide being damaged (λ) to be 0 .0331, meaning an average fragment size of about 30 bp (Figure 5.5).

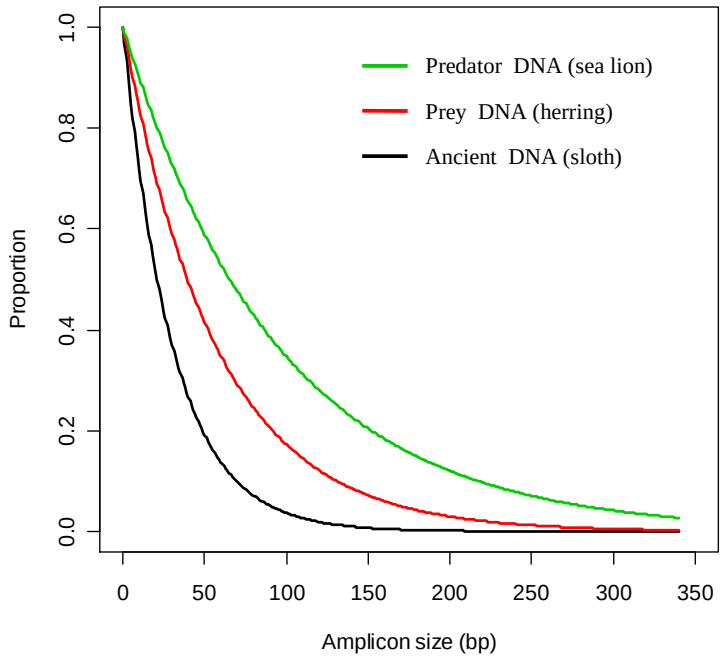


Figure 5.5 *Plots of the estimated proportion of amplifiable fragments versus amplicon size for various faecal DNA extracts. The predator and prey plots represent the mean lamda values for sea lion (λ = 0.0106) and herring DNA (λ = 0.0176) determined in the current study. The ancient DNA lamda was estimated from published data (Poinar et al. 2003) which quantified sloth mtDNA in a late Pleistocene sloth coprolite (λ = 0.0331).*

Although this estimate is likely inaccurate due to the limited data, it is consistent with *a priori* expectations that ancient faecal DNA would be considerably more degraded than modern faecal DNA, and it demonstrates the potential usefulness of my approach for determining DNA damage in molecules from ancient sources.

The ancient DNA data also highlight an important practical point – when using this method to quantify DNA damage in highly degraded samples, it is best to determine copy numbers for small fragments. This is because the copy number will decrease rapidly as fragment size increases and qPCR measurements at low copy numbers (< 100 copies per reaction) are inaccurate due to the larger relative impact of stochastic factors in PCR (Peccoud & Jacob 1996). Another concern when dealing with low copy numbers is the potential influence of reconstructive polymerization (i.e. fragments larger than the size of the DNA template being amplified through the annealing and extension of overlapping fragments). The difficulty in obtaining large PCR products from degraded DNA suggest the overall influence of this phenomenon is relatively minor; however, when the amount of template is low, competition for reaction components is minimal and this source of error does have the potential to bias result (Golenberg *et al.* 1996; Binaschi *et al.* 2000). A final reason to focus on shorter PCR fragments is that selective inhibition of longer PCR amplifications can be caused by some DNA extracts (Pusch *et al.* 2004).

A further potential application of the method could be for studies on the process of DNA decay. While DNA damage should correlate with age of template, the connection is often somewhat unclear (Pääbo 1989; Gilbert *et al.* 2003; Mitchell *et al.* 2005). A possible reason in some studies is that quantity is being used as a proxy for quality (Marota *et al.* 2002; Wandeler *et al.* 2003). The problem with this is that the high variance in the quantity of DNA between different samples can obscure the decrease in the amount of DNA over time. In this study there was roughly a 10-fold variance in amount of DNA between samples, whereas the variance in λ values was only 2-fold. This suggests that it might be better to determine DNA degradation in samples of different ages rather than focusing on the amount of DNA present. Several studies have measured DNA decay using various biochemical assays (Pääbo 1989; Höss *et al.* 1996; Mitchell *et al.* 2005). While these studies provide valuable information on the chemical process of DNA decay, the methods they employ are often not easily accessible. The currently outlined technique should be more accessible and could be modified to allow for the quantification of various forms of

DNA damage. For example, the frequency of cytosine deamination could be quantified through comparison of the original sample with aliquots treated with uracil N-glycosylase (Hofreiter *et al.* 2001). Other forms of damage could also be measured using other lesion-specific endonucleases (or chemical equivalents) or lesion-specific repair enzymes (Pääbo 1989).

The method presented in this chapter has several advantages over conventional PCR-stop assays even in situations where the latter are applicable. The main benefit of the approach is that it measures the decrease in available template within samples, rather than relying on external references. The use of external references requires that the amount of template be normalized in each sample; therefore the error in the total DNA concentration measurement (usually obtained using PCR-independent fluorometric or spectrophotometric methods) is incorporated into these assays. In addition, by relying only on comparisons within samples, absolute quantification of copy numbers is not critical. This means that non-DNA damage related inhibitory effects of treatment on PCR will be controlled for (i.e. the presence of inhibitors should affect each PCR from the same template equally, and the relative qPCR estimates will not change). The method outlined here should also be robust to errors in qPCR measurement because each sample is quantified using several assays (five in the current study) rather than relying on just one or a few assays.

Both the conventional PCR-stop assays and the method currently described have some limitations. First of all, the distribution of damage on the DNA molecules must be approximately random in order for the frequency of DNA damage to be quantified. This requirement will not be satisfied in all cases, but the current approach does have the advantage of being able to detect this problem (i.e. significant deviations from randomness will be apparent when attempting to fit model **2**). Another limitation of PCR-stop assays and the current method is that only damage which blocks DNA polymerase will be measured and damage that partially blocks polymerase during the extension step will be underestimated (Eckert & Opresko 1999). When template quality is being evaluated, focussing on the "amplifiability" of DNA in the sample seems appropriate; however, in other studies this measurement may not be relevant. Finally, the methods are not sensitive enough to measure very low levels of DNA damage. The sensitivity can be boosted by increasing the sizes of the amplified fragments (Jennerwein & Eastman 1991) – currently used PCR-stop assays have amplified PCR fragments up to 17.7 kb in size to measure DNA damage

induced by genotoxic agents at biologically relevant doses (Ayala-Torres *et al.* 2000). I have not investigated the limits of the approach outlined here, but it should be similar to the sensitivity of the conventional PCR-stop assay.

In summary, the method outlined in this chapter will be useful for quantifying gene-specific DNA damage in highly degraded, mixed template samples that cannot be analysed using conventional PCR-stop methodology. The approach should be more accessible than alternate biochemical methods of studying DNA damage and will allow researchers to measure template quality to evaluate alternate sources of DNA, different methods of sample preservation and different DNA extraction protocols. The technique could also be applied to study the process of DNA decay. Lastly, even in situations where conventional PCR-stop assays are applicable, the approach may prove to be a more precise alternative.

Chapter 6

Studying seabird diet through genetic analysis of faeces: a case study on macaroni penguins (*Eudyptes chrysolophus*)

… the use of dung as an indicator of foods of any species is even less to be recommended.

— P.H.T. HARTLEY,

Journal article titled "The assessment of the food of birds"

Abstract

In the final data chapter I investigated the use of DNA-based faecal analysis to obtain dietary information from penguin faeces. Assessment of penguin diet usually relies on the analysis of stomach contents obtained through stomach flushing; this requires bird capture and the results can be biased due to differential digestion of prey species. Methods were developed and applied to detect prey DNA present in faeces collected from macaroni penguins (*Eudyptes chrysolophus*) breeding at Heard Island. For comparison, diet data were also obtained through conventional analysis of stomach contents. To obtain genetic data from faeces I used two approaches: (1) specific PCR tests were developed to detect DNA from five groups of potential prey, and (2) prey DNA obtained from a subset of the faecal samples was cloned and sequenced. Of the 88 faecal samples collected, 39 tested positive for one or more of the prey groups targeted with PCR tests. Euphausiid DNA was most commonly detected in early stages of chick rearing and DNA from the myctophid fish *K. anderssoni* was prevalent in samples collected in later stages; this trend mirrored the data obtained from the stomach contents. Analysis of prey sequences in "universal" clone libraries revealed a highly biased recovery of sequences from fish prey. This bias is most likely caused by the use of degenerate primers with a higher binding affinity for fish DNA template compared to DNA from other prey groups. In euphausiid-specific clone libraries the proportion of sequences from the two dominant euphausiid prey species (*Euphausia vallentini* and *Thysanoessa macrura*) changed over the sampling period; again, this reflected the trend seen in the stomach content data. Based on these findings, I conclude that DNA-based analysis of faeces has great potential for detecting specific prey species in seabird diet, and also for monitoring population level dietary trends. More studies that apply DNA-based methods in combination with conventional methods to analyse seabird diet would be useful for further validation of this new approach.

6.1 Introduction

Prey remains in faeces of predators have long been an important source of data for diet studies. In marine mammals, collection of faeces and identification of hard parts (cephalopod beaks and teleost otoliths and bones) has allowed large numbers of dietary samples to be analysed in population scale surveys (Olesiuk 1993; Tollit & Thompson 1996). This approach has not been used in studies of seabird diet because very few hard-parts are present in avian faeces and any recognizable material that is present in the faeces would almost certainly represent only the least digestible food items (Hartley 1948).

The conventional method for diet analysis in seabirds is therefore the examination of prey remains present in their stomachs. Historically these samples were obtained through lethal sampling (Hartley 1948; Croxall *et al.* 1985); however, during the last 20 years most studies have used stomach flushing as a non-lethal alternative method of obtaining stomach contents (Wilson 1984; Green *et al.* 1998; Hull 1999). While stomach flushing is certainly an improvement ethically, it is still intrusive because it requires animal capture and the procedure can have negative impacts on the sampled birds (Chiaradia *et al.* 2003). This means the number of stomach samples that can be obtained during a study is often limited and the approach may not be appropriate for studies of endangered species. The number of samples that can be collected is also often constrained by the operational difficulty of the flushing procedure, and studies are usually restricted to breeding birds that are feeding their chicks (as these are the only individuals that bring food back to the colony in their stomachs). Dietary studies based on stomach content analysis can be further hindered by a large number of unidentifiable remains in the stomach (Scribner & Bowman 1998) and recovery biases caused by differential digestion and/or retention of prey remains (Gales 1988).

Stable isotope analysis of tissue or feathers has also been used to study seabird diet. This approach provides information on trophic position of seabirds over relatively long periods of feeding. It has been useful for assessment of broad dietary shifts and changes in foraging location (Bocher *et al.* 2000; Cherel *et al.* 2000), but isotope analysis generally cannot provide fine-scale diet data.

The recent development of DNA-based methods to study diet (Symondson 2002) may provide an opportunity to retrieve dietary information from seabird faeces. This methodology does not rely on the survival of visually identifiable prey remains during digestion, only the survival of small amounts of prey DNA (Jarman *et al.* 2002; Symondson 2002). The majority of genetic diet studies have focussed on identification of prey remains in stomach contents (Scribner & Bowman 1998; Rosel & Kocher 2002; Agustí *et al.* 2003; Sheppard & Harwood 2005), but the approach has been applied to faeces in some instances (Sutherland 2000; Jarman *et al.* 2002; Jarman *et al.* 2004; Deagle *et al.* 2005b; Parsons *et al.* 2005). In birds, Sutherland (2000) showed that it is possible to recover insect DNA from the faeces of several species of passerine birds using polymerase chain reaction (PCR). Jarman et al. (2002, 2004) used a similar approach to amplify krill and fish DNA from the faeces of Adelie penguins. These studies clearly show that there is potential to obtain information about bird diet based on the DNA in their faeces, but they identified only a few prey species and analysed faeces collected from a small number of birds. Further assessment and validation of the methodology is needed before it can be more widely applied.

The objective of the present study was to compare results obtained from DNA-based faecal analysis with conventional stomach content analysis in a dietary investigation of the macaroni penguin (*Eudyptes chrysolophus*). The diet of this penguin has been the subject of several studies due to the importance of this species in the Southern Ocean ecosystem (Croxall & Prince 1980; Brown & Klages 1987; Klages *et al.* 1989; Green *et al.* 1998; Crawford *et al.* 2003). Macaroni penguins have an estimated population size of more than 11 million breeding pairs throughout their subantarctic distribution, and more than a million pairs are estimated to breed on Heard Island, the location of the current study (Cooper *et al.* 1990). Previous diet studies of the Heard Island macaroni penguin population using conventional stomach content analysis have shown they feed primarily on euphausiids, myctophid fish and to a lesser extent amphipods (Klages *et al.* 1989; Green *et al.* 1998). Here, I report on the diet of penguins during chick rearing period in the summer of 2003/2004. Diet was examined through conventional analysis of stomach contents and through the analysis of prey DNA extracted from the faeces. To obtain genetic data from the faecal samples I used two approaches. First, I determined the presence or absence of DNA from some potential diet items by applying PCR tests that specifically amplify

DNA from targeted groups of prey. Second, the prey DNA obtained from some of the faecal samples was cloned and sequenced. Sequences obtained were then compared to those in GenBank to further identify prey taxa.

6.2 Material and Methods

6.2.1 Study site, sample collection and DNA extraction

The macaroni penguin diet samples analysed in this chapter were collected between December 20th 2003 and February 16th 2004 at the Capsize Beach breeding colony on Heard Island (Figure 6.1). The breeding cycle in these penguins is highly synchronous within the colony. The sampling period covered two distinct phases of chick rearing: guard stage, when females take short foraging trips to provision the young while the male remains at the colony with the chick (December 20th – January 15th); and crèche stage, where both sexes take longer foraging trips and both provision the chick (January 16th – February 16th).

Stomach contents were collected from 69 adults returning to the colony using the water-offloading technique (Hull 1999). A maximum of two stomach flushes were performed during the procedure and individual birds were sampled only once (breast feathers of flushed birds were marked with Niazonal dye to ensure they were not re-sampled – this dye is permanent and bird remain marked until feathers are moulted at the end of the breeding season). The recovered stomach contents were drained through a sieve with 0.5 mm mesh size to remove excess water, and then preserved in 70% ethanol.

I had planned to obtain faecal samples for genetic analysis from the same birds that were stomach flushed, but very few of the birds defecated on capture. Those that did defecate almost invariably produced white fluid faecal pellets which contain no detectable prey DNA (unpublished data). Faeces were therefore collected from penguins present on the beach in front of the colony over the same time period (n = 88). The faecal samples were collected immediately after defecation and stored for approximately one year in 70% ethanol at 4 ºC before extraction. DNA was extracted from approximately 100 mg of faecal material from each sample, using the QIAamp DNA Stool Mini Kit (Qiagen) with the minor modifications described previously in

(Deagle *et al.* 2005b). The DNA was eluted in 100 μL Tris buffer (10 mM). Blank extractions were included in each batch of extractions to monitor for cross-over contamination.

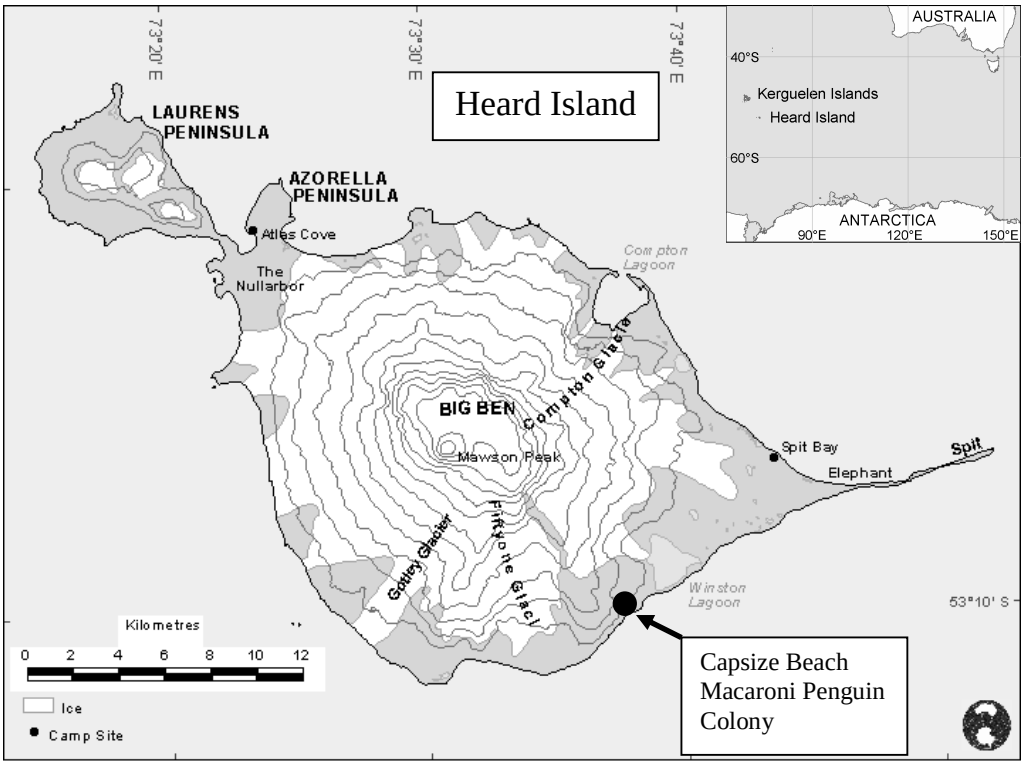


Figure 6.1 *Location of Capsize Beach penguin colony on Heard Island. Inset map shows location of Heard Island in relation to Australia and Antarctica. Maps courtesy of the Australian Antarctic Division.*

6.2.2 Stomach content analysis

In the laboratory, the stomach samples were sorted by elutriation to remove fish otoliths and squid beaks. Once hard parts were removed, the samples were drained on a 0.5 mm sieve, blotted dry and the total mass of the sample was taken. To determine the composition of different prey groups by mass, a 30 g sub-sample was

analysed in detail (for samples < 30 g the whole stomach sample was analysed). Each sub-sample was examined under a dissection microscope and divided into broad prey classes (euphausiids, fish, amphipod, cephalopods or unrecognisable material). The mass of each component was recorded. For the calculation of composition by mass for each sample, the unidentifiable component of the sub-sample was assumed to contain the same proportions of prey as the identifiable component, and the sub-sample was assumed to be representative of the entire sample (Croxall *et al.* 1985). The reconstituted mass of the diet is calculated in many diet studies, but was not determined here for two reasons: (1) in early foraging trips the relatively low level of digestion for most samples meant that the composition by mass could be determined directly; (2) in many of the later samples which were more completely digested, the otoliths that had accumulated likely represent more than one meal and calculation of mass based on these relatively robust structure is unlikely to provide a balanced view of the diet.

To further identify prey present in the diet, otoliths and squid beaks were identified where possible using published keys (Clarke 1985; Williams & McEldowney 1990). To identify the euphausiids, and to determine their species composition by number, up to 100 randomly selected individuals were identified in each sample (Croxall *et al.* 1985; Fischer & Hureau 1985). Amphipods were identified using unpublished reference material obtained from the Australian Antarctic Division.

6.2.3 Genetic presence/absence detection in faecal samples

For each faecal sample, the presence/absence of DNA from particular prey item was determined with five separate PCR assays using group-specific primers. The PCR assays were chosen to detect prey items that were previously identified in the diet of macaroni penguins at Heard Island. The following prey groups, or species, were tested for: (i) euphausiids; (ii) the myctophid fish, *K. anderssoni*; (iii) fish from the suborder Nototheniodei; (iv) amphipods; and (v) cephalopods. Primers details are given in Table 6.1.

Table 6.1 *PCR primers used in this chapter to detect DNA from five groups of potential prey in macaroni penguin faecal samples.*

Target (**taxon** – gene)	Primer name[a]	Sequence 5'→3'	Product size (bp)	Annealing temperature	Reference
Euphausiid [b] – mitochondrial 16S rDNA	EuphMLSUF	tttattggggcgataaaaat	169	54 ℃	This study
	EuphMLSUR	tcgaggtcgyaatctttcttgt			This study
Krefftichthys anderssoni [b] – mitochondrial 16S rDNA	KaMLSUF	cccacatcaaataccccta	169	55 ℃	This study
	KaMLSUR	gggtcattggtggtcagaag			This study
Nototheniodei – mitochondrial 16S rDNA	NotoMLSUF	ccctatgaagcttyagacrta	~ 275	55 ℃	Jarman *et al.* 2004
	NotoMLSUR	ccttgttgatawggtctctaaaa			Jarman *et al.* 2004
Amphipoda – nuclear 18S rDNA	AmphNSSF1	ctgcggttaaaaggctcgtagttgaa	204–375	51 ℃	Jarman *et al.* 2006
	AmphNSSR1	actgctttragcactctgatttac			
Cephalopoda nuclear 28S rDNA	Squid28SF	cgccgaatcccgtcgcmagtaaamggcttc	~ 180	55 °C	Deagle *et al.* 2005b
	Squid28SR	ccaagcaacccgactctcggatcgaa			Deagle *et al.* 2005b

[a] F and R denotes forward and reverse.

[b] The specificity of these primer pairs was not tested exhaustively. The EuphMLSUF primer binding site is conserved in the euphausiid genera I was targeting and is not conserved in non-euphausiid crustaceans. The KaMLSUF primer binding site is present in *Krefftichthys anderssoni*, but is not present in several closely related myctophid fish. The KaMLSUF primer set was tested on genomic DNA from the myctophid fish *Electrona carlsbergi* and a channichthyid *Champsocephalus gunnari*; neither of these templates produced positive PCR amplifications.

PCR amplifications were performed in 25 µl reactions containing 0.4 µM of each primer, 0.2 mM dNTPs, 2.0 mM $MgCl_2$, 1X BSA (New England Biolabs),1X AmpliTaq Gold buffer and 0.625 units AmpliTaq Gold (Applied Biosystems). Template was 1 µL of the DNA extract. Thermal cycling conditions were as follows: 94°C for 10 min then 35 cycles (94°C for 30 s / primer-specific annealing temperature for 30 s / 72°C for 45 s) followed by 72°C for 2 min. Aerosol-resistant pipette tips were used with all PCR solutions and negative control reactions (extraction control and a distilled water blank) were performed with each set of PCR amplifications. PCR products were separated by electrophoresis in 1.8% agarose gels and visualised by staining with ethidium bromide and transillumination with UV light.

6.2.4 Genetic clone library analysis of faecal samples

PCR clone libraries were produced from representative faecal samples which contained prey DNA and clones from these libraries were sequenced. Two primer sets were used to produce clone libraries:

(A) The first primer set targets a region of the mitochondrial 16S rDNA gene present in sequences from fish, cephalopods and crustaceans (16S1F-degenerate gacgakaagacccta and 16S2R-degenerate cgctgttatccctadrgtaact, further details in Table A2.2; based on primers described in Deagle *et al.* 2005a). The adenine nucleotide on the 3' end of the forward primer does not match the primer binding region in birds, but is conserved in the target prey groups; this single base mismatch was effective in preventing the amplification of penguin DNA present in the faecal samples (unpublished data). Using this primer set, with an annealing temperature of 54°C, I amplified DNA from ten faecal samples and cloned the products. Six sequences were obtained from each sample, giving a total of 60 sequences from these libraries.

(B) The second primer set was the euphausiid primer pair (Table 6.1). With this primer set I amplified DNA from ten faecal samples and ten sequences were obtained from each sample, giving 100 sequences from these libraries. PCR amplifications were carried out following the protocol outlined in the previous section. Products were cloned using the TOPO TA cloning system following instructions of the manufacturer (Invitrogen). Colonies containing recombinant clones were cultured in

LB broth and plasmid DNA was purified by alkaline lysis (Birnbolm & Doly 1979). Sequencing was carried out on 300 ng of plasmid DNA using the BigDye Terminator Version 3.1 cycle sequencing reagents (ABI). Capillary separation was performed on AB3730xl sequencing platforms at the Australian Genome Research Facility. Chromatograms were examined by eye to check base calling in the program Chromas2 (Technelysium). In a few cases, I generated my own sequence data to aid in clone identification. This was accomplished by direct sequencing of PCR products generated from genomic DNA of known species using the primers 16Sar-5' and 16Sbr-3' (Palumbi 1996).

6.2.5 Data analysis

To compare diet composition between the two stages of chick rearing (guard and crèche) I applied the non-parametric method, ANOSIM (analysis of similarity) to both the stomach content data and the faecal genetic data, using PRIMER statistical software (version 5.2.9). The procedures used followed those outlined in (Clarke 1993). For the stomach content data, the mass of prey groups present within each sample was converted to percentage composition. These data were used in preference to frequency of occurrence data since the incorporation of mass measurements provides a more accurate account of diet (by preventing prey taken in small quantities being over-represented). For the genetic results, no weighting of the detection results were possible, so comparisons were carried out using the presence/absence detection data. For each dataset a similarity matrix was generated using the Bray–Curtis similarity measure. ANOSIM tests were run on the matrices using 999 permutations to test for statistically significant differences in diet composition between samples collected during guard and crèche stage. The contribution of each prey category to the average dissimilarity between the chick rearing stages was calculated using the similarity percentages procedure (SIMPER).

To identify sequences obtained in the clone library analysis, sequences were aligned and grouped into clusters of nearly identical sequences (i.e. sequences differing by only a few base substitutions). All unique sequences were compared with sequences in GenBank using the BLAST program (Altschul *et al.* 1990). Matches that were identical with sequences present in GenBank over the entire fragment and that

were different from other species within the same genera were considered to provide species level identification. If all sequences within a nearly identical cluster shared a common closest match they were considered to be the same species. If no 100% matches were present in the database for any members of a nearly identical cluster, the consensus sequence was classified to the lowest taxonomic level possible with reference to the closest BLAST matches.

6.3 Results

6.3.1 Stomach content analysis

Of the 69 stomach samples, 11 were empty and an additional five were excluded from further analysis because they had a mass of less than 5 g. The remaining 53 stomach samples had a mean mass of 75 g (s.d. = 51 g, range = 12- 216 g); these included 37 collected during guard stage and 16 collected during crèche stage of chick rearing. Overall, euphausiids formed the largest component of the stomach samples by mass (69%). Fish ranked second (22%), followed by amphipods (8%) and then cephalopods (< 1%); (Table 6.2). The same ranking of relative importance of these prey groups was obtained based on the frequency of occurrence data (Table 6.3). ANOSIM tests detect significant differences in the prey mass proportions during guard and crèche stages (global R of 0.441; p = 0.01) This difference was due to an increase in the amount of fish and amphipods, with a corresponding decrease in the importance of euphausiids, during the later stages of chick rearing (Table 6.2). SIMPER analysis show a percentage dissimilarity of 58% between stages, and the contribution of the prey categories to the discrimination were: euphausiids (46%), fish (31%) and amphipod (22%).

The prey species identified in the stomach included at least three species of euphausiids, three fish, two amphipods and one squid (Table 6.4). Two species of krill (*Euphausia vallentini* and *Thysanoessa macrura*) made up the vast majority of the identified euphausiids. Both of these species have closely related sister taxa which occur in the vicinity of Heard Island (*E. frigida* and *T. vicina*) and the fragile taxonomic features distinguishing these species were missing from many of the partially digested samples, therefore the occurrence of these sister taxa in the samples

could not be discounted. Another species of euphausiid, *E. tricantha*, was found in very small numbers (only four specimens out of more than 3000 euphausiids identified). *E. vallentini* was the dominant species consumed during the early part of the study and it was almost completely replaced by *T. macrura* in samples collected during the latter part of the study. One intact fish was recovered and identified as *Channichthys rhinoceratus*; four additional digested channichthyid icefishes were recovered but could not be further identified, and several small unidentifiable fish were also present in the samples. Of the 3255 fish otoliths recovered from the stomachs which could be identified to species, all belonged to the myctophid *K. anderssoni*, with the exception of a single otolith identified as coming from *Electrona antarctica*. Likewise almost all identified amphipods belonged to a single species, *Themisto gaudichaudii*; the only exception was a single specimen identified as *Hyperia macrocephala*. From the squid remains present in the samples only one lower beak was large enough to allow identification, this beak came from the squid *Galiteuthis glacialis*.

Table 6.2 *Stomach sample composition of the main prey groups consumed by macaroni penguins during chick rearing (based on wet mass of prey components in samples). Unidentifiable material was assumed to be distributed in proportion to the composition of the identified components.*

	Total (n = 53)		Guard (n = 37)		Crèche (n = 16)	
	(g)	(%)	(g)	(%)	(g)	(%)
Euphausiids	2760.3	69	2279.0	84	481.4	38
Fish	884.2	22	424.5	16	459.7	36
Amphipods	327.4	8	7.6	< 1	319.8	25
Cephalopods	10.9	< 1	1.0	< 1	9.9	1
Total	3982.8	100	2712.1	100	1270.7	100

6.3.2 Genetic presence/absence detection in faecal samples

Slightly less than half of the faecal samples (39 out of 88) tested positive for one or more of the prey groups targeted with PCR tests. In these 39 samples, euphausiid DNA was detected in 15 samples (38%), *K. anderssoni* DNA in 24 samples (62%), Nototheniodei DNA in 6 samples (15%), amphipod DNA in 14 samples (36%) and cephalopod DNA in 4 samples (10%); (Figure 6.2). In these data there was a significant difference in the prey items detected between the guard and crèche stage (ANOSIM global R of 0.346; p = 0.01). As with the stomach mass data, euphausiids were detected more in guard stage samples compared with crèche stage samples and the opposite was true for fish and amphipods (Table 6.3). The contribution of these prey categories to the percentage dissimilarity between stages (73%; SIMPER analysis) were: euphausiids (36%), *K. anderssoni* (30%) and amphipod (20%).

Table 6.3 *Comparison of percent frequency of occurrence data (% FO) of prey groups identified through conventional stomach content analysis and genetic analysis of faeces.*

	Stomach data		**Faecal DNA data**	
Prey Item	Guard (n = 37)	Crèche (n = 16)	Guard (n = 13)	Crèche (n = 26)
	% FO	% FO	% FO	% FO
Euphausiids	97	100	85	15
K. anderssoni	62 [a]	100 [a]	31	77
Nototheniodei	5	6	0	23
Amphipods	54	75	15	46
Cephalopods	14	25	0	15

[a] Based on otolith recovery

Table 6.4 *Comparison of the prey identified by conventional stomach content and faecal DNA analysis.*

Prey Group	Species ID	Stomach contents	Faecal DNA presence/absence	Faecal DNA clone libraries
Euphausiids		+	+	+
	Euphausia vallentini	+		+
	Euphausia tricantha	+		
	Euphausia frigida			+
	Thysanoessa macrura	+		+
Fish		+	+	+
	Krefftichthys anderssoni	+		+
	Electrona antarctica	+		+
	Nototheniinae sp.			+
	Harpagifer sp.			+
	Champsocephalus gunnari			+
	Channichthys rhinoceratus	+		
Amphipods		+	+	
	Themisto gaudichaudii	+		
	Hyperia macrocephala	+		
Cephalopods		+	+	+
	Gonatus antarcticus			+
	Galiteuthis glacialis	+		

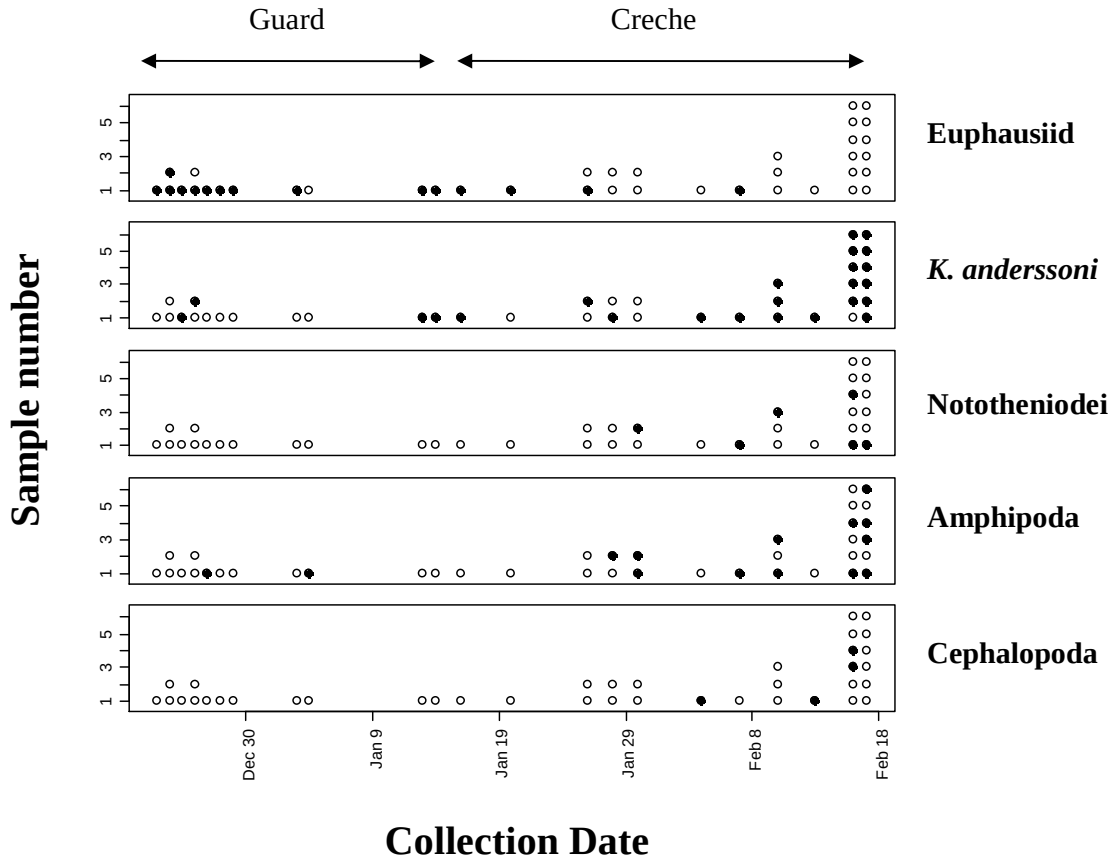


Figure 6.2 *Summary of the detection data for five prey groups targeted with specific PCR tests. Each box represents results from groups labeled on right. Each dot represents a faecal sample which tested positive for at least one prey item (39 in total); a filled dot indicates detection of the particular prey group. The horizontal axis shows the date the samples were collected.*

6.3.3 Genetic clone library analysis of faecal samples

Seven prey items were identified in the 60 sequences obtained from clones produced from PCR amplifications using the conserved primer set (16S F + R degenerate); results are summarised in Table 6.5. These clones were almost entirely derived from fish, with the majority of sequences matching the myctophid fish *K. anderssoni*. Other fish represented in the clone libraries include another myctophid (*E. antarctica*), a channichthyid (*Champsocephalus gunnari*), two species from the family Nototheniidae and one fish species whose sequence does not closely match any of the species represented in GenBank. The only non-fish prey detected was from a single clone identified as the squid *Gonatus antarcticus*.

In the clone libraries produced from PCR amplifications using the euphausiid primer set, three species of krill were identified in the 100 clones sequenced. Seventy of the clones matched *T. macrura*, 28 matched *E. vallentini* and 2 matched *E. frigida* (Table 6.5). Initial identification based on data present in GenBank allowed only genus level identification; final classification required that additional sequence data be generated for these three species (Table 6.5). I classified the *T. macrura* sequences based on a 100% match but there is no sequence data available for the closely related species *T. vicina*; therefore, I cannot discount the possibility that this species is present but cannot be distinguished. The proportion of clones from the two euphuasiids changed over the sampling period, with *Euphausia* sp. being the dominant krill component of the diet in samples collected during the early part of the sampling period and *T. macrura* being identified exclusively in samples collected later. This mirrors the trend seen in the stomach content data (Figure 6.3).

Table 6.5 *Summary of the taxonomic classification of sequence data obtained through the analysis of clone libraries produced from penguin faecal DNA. Further information on the results from individual samples is provided in Appendix IV.*

Clone library	Group – Family	Species ID	# of clones	Accesssion #	% similarity of closest match
A	**Fish**				
Conserved primers	– Myctophidae	*Krefftichthys anderssoni*	42	AB042176	100%
		Electrona antarctica	2	AY141397	99%
	– Channichthyidae	*Champsocephalus gunnari* [a]	4	AY249471	100%
	– Nototheniidae	*Harpagifer* sp. [b]	1	AY520130	100%
		Notothenioidae sp. [c]	4	DQ356243	99%
	– Unknown	*Unidentified Acanthopterygii*	6	DQ356242	82%
	Cephalopods				
	– Gonatidae	*Gonatus antarcticus*	1	AY681032	100%
			60		
B	**Euphausiids**				
Euphausiid primers	– Euphausiidae	*Euphausia frigida*	2	DQ356239	100%
		Euphausia vallentini	28	DQ356241	100%
		Thysanoessa macrura	70	DQ356238	100%
			100		

[a] sequence is also 100% match with *C. esox*, but this species not found near Heard Island

[b] sequence is 100% match with *H. kerguelensis* and *H. antarcticus*

[c] sequence is 98- 99% match with *Gobionotothen* spp. and *Notothenia coriiceps* both within the sub-family Notothenioidae

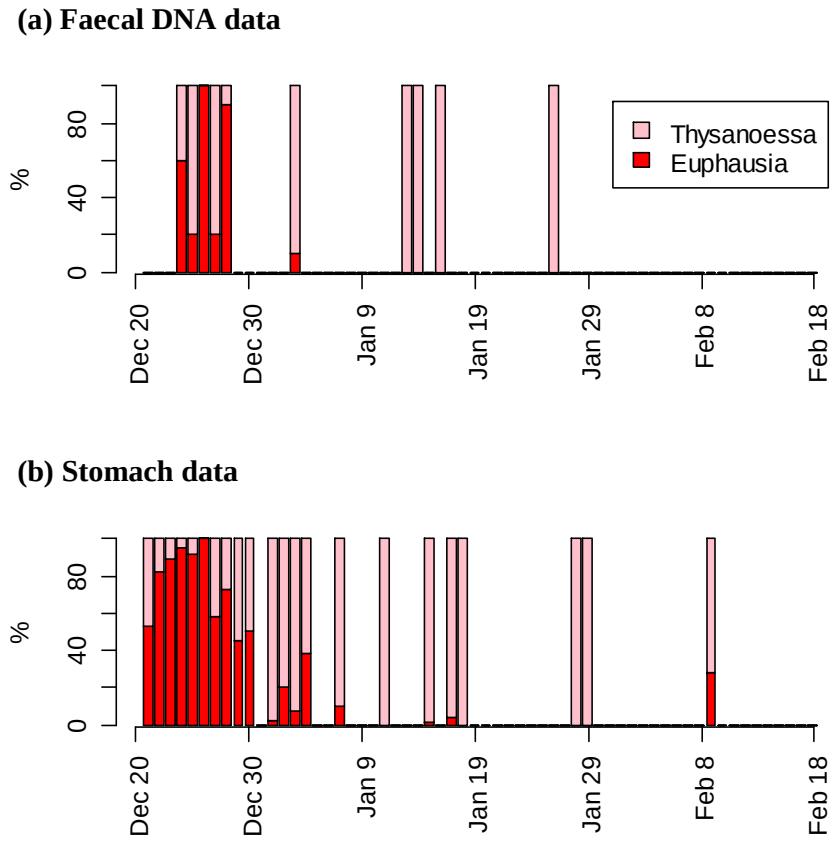


Figure 6.3 *Proportional breakdown of two euphausiid genera in diet samples collected over the sampling period: (a) Based on 100 sequences obtained from cloned PCR products amplified using a euphausiid-specific primer set (ten sequences from each of ten clone libraries); (b) Based on numbers present in stomach samples (data from multiple stomach samples collected on the same day were pooled).*

6.4 Discussion

The stomach content data I collected present a picture of macaroni penguin diet that is consistent with the results obtained in two previous diet studies carried out at Heard Island (Klages *et al.* 1989; Green *et al.* 1998). In all three studies the majority of the diet was made up of a combination of two euphausiids, *E. vallentini* and/or *T. macrura*, and one species of myctophid fish, *K. anderssoni*. All studies found that the amphipod *T. gaudichaudii* was taken in significant numbers by some

penguins and that squid was a very minor component of the diet. The most common species of euphausiids in the diet varied between studies: Klanges et al. (1989) reported both species were regularly consumed with *T. macrura* being marginally more important; Green et al. (1998) found almost exclusively *E. vallentini* in the diet; and I found a shift in the dominant species in the diet from *E. vallentini* to *T. macrura* over the course of the study. Shifts in the principal crustacean species being eaten have been reported previously in other penguin species and are likely due to temporal changes in prey availability (see discussion in Green et al. 1998). Dietary changes through the chick-rearing period of various penguin species have also been reported, often with a reduction in the reliance on euphausiids during later stages of chick rearing (Hindell 1988; Green *et al.* 1998). This trend was observed in the current study; euphausiids constituted 84% of the diet by mass during guard stage and 38% by mass during crèche stage. This shift is potentially due to a change in breeding constraints that allows for longer foraging trips by adults as the chick develops (Hindell 1988).

The DNA-based faecal analysis provided promising results. I was able to detect DNA from a range of pre-defined prey groups in the faecal samples using group-specific PCR assays. The fish *K. anderssoni* was the most common prey species detected, followed by the euphausiids. Compared with the stomach samples the order of importance of these key prey groups was reversed; this probably reflects the fact that more of the faecal samples were collected later in the season when fish were a more important part of the diet. The dietary shifts that were observed in the stomach content data are also apparent in the genetic data. Based on the presence/absence results there was a decrease in detection of euphausiid DNA and an increase in detection of fish and amphipod DNA during crèche stage. The change in the dominant euphausiid species being consumed by the penguins during the early part of the study was clearly seen in the euphausiid clone library data. Additional sequence results from the clone library analysis confirmed the dominance of *K. anderssoni* in the fish component of the diet and also revealed a larger diversity of fish prey in the diet than the stomach results. An increase in the importance of fish from the suborder Nototheniodei was apparent in the genetic analysis (in both the presence/absence and clone library data). The importance of this clade in the diet could easily have been underestimated in the stomach content analysis because there were a number of digested, unidentifiable remains from small fish present in the

stomach samples. Alternatively, the increased detection of the Notothenidei fish in the faecal samples could have resulted from real differences in the diet of the birds that were stomach sampled versus those whose faeces were sampled. The stomach samples were obtained from breeding birds (i.e. carefully selected birds with protrusive brood pouches that were purposefully returning to the colony). In contrast, the faecal samples were collected from any penguins present on the beach near the colony; these samples would have partly (or even mostly) come from non-breeding birds. Dietary differences between these groups is quite possible since non-breeding macaroni penguins tend to travel a lot further and are at sea for longer during foraging trips compared with breeding birds (Trebilco 2004).

I did encounter several difficulties in the genetic analysis that could be remedied in future studies. First, a relatively large number of samples contained no amplifiable prey DNA, resulting in a smaller than anticipated sample size. There are several possible reasons for this: many of the samples may have contained PCR inhibitors, DNA in some samples may have degraded during storage, or there may not have been any prey DNA in some samples when they were collected because the defecating bird had not fed recently. The last explanation is almost certainly true in some cases since nearly 20% of the breeding birds that were stomach flushed had empty stomachs. Non-breeding birds typically spend much longer ashore between foraging trips (Trebilco 2004), therefore faeces from many of these birds may not contain recoverable amounts of prey DNA. It would be useful to know the persistence of a detectable genetic signal in faeces during periods of fasting. This could be determined using captive birds and might clarify the reasons for the high incidence of negative results. These data are also needed to evaluate whether prey DNA present in faeces originates only from foraging that occurs relatively close to the colony, or if a genetic signal can be obtained from prey captured in more distant locations. Regardless of the length of time a genetic signal persists, the collection of faeces should focus on birds that have recently returned from foraging trips to maximize the chance of prey DNA being present. It would also be prudent to collect large numbers of samples to compensate for the proportion of samples that do not contain any amplifiable prey DNA. In the case of penguins at colonies, the samples are easily obtained and the initial PCR screening of samples for prey DNA can be done relatively quickly and cheaply.

Another technical difficulty was that the clone libraries generated using the conserved 16S primers (degenerate primers specifically designed to amplify DNA from a wide variety of prey) did not represent the diversity of prey in the penguins' diet – almost all the sequences obtained in this analysis came from fish. Two important prey groups, euphausiids and amphipods, were conspicuously absent in these libraries. I have previously shown that quantitative estimates of diet based on DNA amount in faeces can be biased due to differences in prey DNA density and DNA survival during digestion (Chapter 4; Deagle & Tollit 2006). However, the cause of the major bias found in the present study is likely methodological – the result of differences in primer binding efficiency for the different prey groups targeted. This is supported by the fact that DNA from euphausiids and amphipods were present in some of these samples based on the presence/absence data (Figure 6.1 and Table A4.1). The degenerate primer set used to create the library was capable of producing PCR products in preliminary testing carried out using genomic DNA template from fish, euphausiids and squid. However, the annealing temperature of both primers matching the euphausiid binding regions is lower than the comparable annealing temperature for the fish primers (see primer details in Appendix II: Table A2.2). This could cause differential PCR amplification of the mixed template present in the faecal samples (von Wintzingerode *et al.* 1997). With regards to the lack of representation of amphipods in this library, the primers were designed with reference to amphipod sequences from the gammaroidean suborder (these were the only amphipod mitochondrial 16S sequences available in GenBank). Testing carried out at the end of the study revealed that PCR products could not be amplified from genomic DNA of the hyperiidean species, *T. gaudichaudii,* with this primer set. To avoid biased conclusions in future studies that use a "universal" primer approach, an increased emphasis should be placed on primer selection. In an ideal situation, completely conserved primer binding sites could be targeted so DNA from all prey would be amplified simultaneously, with equal efficiency. Unfortunately very few primers meet this criterion and also amplify short, informative DNA regions that are well represented in GenBank. One way to safeguard against spurious results caused from primer-specific bias would be to analyse clone libraries produced from multiple conserved primer sets to allow for cross-validation (Blankenship & Yayanos 2005).

It might be assumed that the identification of sequences obtained in faecal analyses would be quite limited due to the short length of DNA that can be amplified

from degraded prey DNA. However, in the current study the taxonomic resolution obtainable in some groups based on sequence data from the penguin faeces was very good. In the euphausiid clone library, I was able to distinguish between short mitochondrial 16S sequences from two species in the genus *Euphausia* (*E. vallentini* and *E. frigida*). These species were not distinguished morphologically in the stomach content analysis, even though it is highly probable that both species were present in the samples. In some groups (e.g. notothenioid fish) differentiation between some closely related species was not possible due to lack of genetic variation in the targeted mitochondrial 16S region. Before a study is initiated, *a priori* analysis of genetic variation in potential prey groups should be carried out to determine if the taxonomic resolving power of a particular marker is suitable for the question being addressed.

The major limitation I encountered in the identification of DNA sequences resulted from a lack of reference sequence data[12] for some prey groups. One DNA sequence I obtained is only distantly related to several ray-finned fish (approximately 20% divergence) and could not be classified further. This was surprising given the relatively good coverage of this group in the database. As discussed above, an entire suborder of amphipods is unrepresented by mitochondrial 16S DNA sequences in GenBank. This is likely to be the case for many groups of marine invertebrates, making identification in diet samples possible only with concurrent sequencing efforts on samples of the relevant potential prey taxa. One of the compensatory features of DNA-based identification is that sequence data obtained in different studies is easily catalogued and available for future comparisons and scrutiny. This means the taxonomic classification of sequences can be improved retrospectively as the number of sequences available for comparison in public database increases. The increase in available sequence data for some genes, such as the mitochondrial cytochrome oxidase subunit I gene, favoured by the international DNA barcoding effort, will be rapid (Hebert *et al.* 2003). This protein coding gene is not ideal for the design of non-degenerate group-specific primers (Vences *et al.* 2004), but it has found some use in diet studies (Agustí *et al.* 2003; Blankenship & Yayanos 2005). The continued development of the barcode of life database (or similar databases that focus on

[12] I used sequences from GenBank for taxonomic identification. These sequences are not always reliably linked to the respective taxonomic specimens, but until better databases become available this remains the primary source of data for DNA-based diet studies.

taxonomy and accuracy of the sequence data generated) will be vital to allow for DNA-based identification to be more generally applied in diet studies.

An established tradition exists of studying the feeding habitats of seabirds by stomach content analysis, and this will likely remain the technique used by most researchers. Stomach content analysis is technically relatively straightforward and provides some data that are unattainable using other methods of dietary analysis (such as information on prey size and meal size). However, the approach does have many drawbacks, such as: the invasiveness of bird capture and the stomach flushing procedure; the time-consuming nature of sample collection and analysis; the limited opportunities for sample collection; and the biases of the technique caused by variable digestibility of different prey. Because of these problems, the use of alternative approaches such as DNA-based faecal analysis will likely become more common as supplements to, or even in place of, traditional stomach content analysis.

The data in this chapter indicates that DNA from a variety of prey is recoverable from penguin faeces and dietary information can be obtained using DNA-based species detection methods. The most obvious advantage of genetic faecal analysis is that a large number of samples can be collected with minimal amount of disturbance to the birds and these samples can be rapidly screened using presence/absence PCR tests. The application of “universal” primers in diet studies can be used to detect DNA from a wide range of prey species and this may provide a powerful method for determining the prey consumed by seabird species that have a broad and/or unknown diet. However, results obtained using primers targeting a broad range of taxa should be interpreted cautiously since differential amplification of DNA from different prey species can cause major biases.

In summary, the DNA-based analysis of faeces has great potential for use in detection of specific prey species in seabird diet, and also for monitoring population level dietary trends. More studies applying DNA-based methods in combination with conventional methods of dietary analysis would be useful to allow for further validation this new approach for studying seabird diet.

Chapter 7

General Discussion and Future Directions

“Oh dear! I’d nearly forgotten that I’ve got to grow up again! Let me see—how is it to be managed?”

— LEWIS CARROLL,
Alice`s Adventures In Wonderland

7.1 Overview of chapter

This thesis consists of a series of studies investigating issues relevant to the development of DNA-based techniques for studying the diet of marine predators. The data chapters (Chapter 2 – Chapter 6) are self-contained; each includes a discussion of the results in context of the currently available literature (including references to other published chapters). In this final chapter, I will review two recent pinniped DNA-based diet studies, and conclude with some general discussion of potential directions for research in this field.

7.2 Recent DNA-based diet studies

A number of DNA-based diet studies have been published in the last two years (Table 7.1 and 7.2). Some of these have not been considered in the data chapters of my thesis because they were not available at the time these chapters were written. The majority of the recent DNA-based dietary research has focussed on terrestrial invertebrates, but there have also been studies on marine invertebrates, fish and marine mammals (see references in Table 7.1 and 7.2 and review by Sheppard & Harwood 2005). Two of these studies (Kvitrud *et al.* 2005; Parsons *et al.* 2005) are on pinnipeds and are directly relevant to the work in my thesis. As such, I will briefly consider them here.

At the same time that the results from the Steller sea lion captive feeding trial were being published (Chapter 3, Deagle *et al*. 2005b), another group of researchers published findings from a similar study (Parsons *et al.* 2005). This group developed PCR tests targeting salmonid mtDNA. They used their assay to test DNA recovered from the soft matrix of faeces collected from captive grey seals (*Halichoerus grypus*), which had been fed a monospecific diet of Atlantic salmon (*Salmo salar*). They were able to detect salmon DNA in 95% of the faecal DNA extracts, a finding that is virtually identical to the findings presented in Chapter 3 (Table 3.2). The ability of the DNA-based approach to detect salmon DNA in faeces from two pinniped species indicates the method will be generally applicable. These results open up exciting new possibilities for studying interactions between pinnipeds and salmon fisheries, an area of research that that has been hampered by the low recovery rates of salmonid otoliths in pinniped faeces.

The second recent pinniped DNA-based diet study also focused on salmonid prey, but was based on analysis of isolated hard-parts obtained from harbour seal (*Phoca vitulina*) faeces collected in the field (Kvitrud *et al*. 2005)[13]. These researchers extracted DNA from 751 salmonid bones removed from 72 faecal samples. With primers targeting the single-copy nuclear growth hormone type-2 gene they were able to amplify DNA from 421 (56%) of the bones. The PCR products varied in size according to species, allowing identification based on standard gel electrophoresis. A total of 237 bones (56%) were from coho salmon (*Oncorhynchus kisutch*), 181 (43%) were from chinook salmon (*O. tshawytscha*) and 3 (<1%) were from trout (*O. mykiss* or *O. clarki*). The chinook salmon bones were further analysed through microsatellite genotyping, which allowed discrimination between individual fish. These microsatelitte genotypes were used to show that in the faecal samples containing chinook bones, 32% had remains from more than one fish. The remains from one individual chinook were recovered in three different faecal samples (due to multiple samples having been collected from one seal or from prey sharing among seals). This study highlights the power of DNA-based identification techniques in diet studies to provide data that could not be otherwise obtained. However, by focusing only on recovered hard-part remains, the study does not take advantage of the ability to detect prey in the faecal soft matrix. The 72 harbour seal faecal samples that contained salmonid bones came from a total of 1,407 faeces examined for hard-parts. It would be interesting to know what percentage of the 1,407 samples contained salmon DNA in the faecal soft matrix.

[13] The study was similar to the one carried out by Purcell *et al*. (2000) – reviewed in Chapter 1 and published in the primary literature as Purcell *et al*. (2004) – who also used PCR to obtain species-level identification of salmonid bones recovered from seal faeces.

Table 7.1 *Recent dietary studies*[a] *of invertebrate predators investigating the use of DNA-based methods for prey detection in gut contents.*

Predator	Targeted Diet Items	Sample Tested	Gene	Marker Size	Sample Size[b]	Field-based	Reference
Wasp species (Vespidae)	Various prey items	Prey remains in mouthparts	mtDNA rDNA	500 – 650	70	Yes	Kaspar *et al.* (2004)
Predatory beetle larvae (*Curinus coeruleus*)	Geometrid moth caterpillars	Whole predator	mtDNA COI	140 – 170	46	No	Sheppard *et al.* (2004)
Marine invertebrates	Various prey items	Gut contents	Nuclear + mtDNA rDNA + COI	700 – 770	6	Yes	Blankenship & Yayanos (2005)
Giant squid (*Architeuthis* sp.)	Various prey items	Gut contents	mtDNA rDNA	190 – 250	1	Yes	Deagle *et al.* (2005a)
Carabid beetle (*Pterostichus melanarius*)	Aphids and slugs	Gut contents	mtDNA COI + rDNA	110 – 250	70	No	Foltan *et al.* (2005)
Carabid beetle (*Pterostichus melanarius*)	Earthworms, aphids, weevils and molluscs	Gut contents	mtDNA COI + rDNA	80 – 240	>200	Yes	Harper *et al.* (2005)
Carabid beetle larvae (*Poecilus versicolor*)	Beetle larvae (*Melolontha melolontha*)	Whole predator	mtDNA COI	180 – 590	445	No	Juen & Traugott (2005)
Predatory bug and spider	Dimondback moth (*Plutella xylostella*)	Whole predator or gut contents	Nuclear rDNA	280 – 430	158	Yes	Ma *et al.* (2005)
Carabid beetle (Pterostichus melanarius)	Spider and aphid	Whole predator or gut contents	mtDNA COI	110 – 250	>300	No	Sheppard *et al.* (2005)
Carnivorous copepod (*Pareuchaeta norvegica*)	Copepod (*Calanus hegolandicus*)	Faecal pellets	mtDNA COI	170	14	Yes	Vestheim *et al.* (2005)

[a] Includes primary literature published in 2004 or 2005 and listed in the Web of Science© database. Diet studies on parasites are excluded (e.g. studies analysing the source of insects' bloodmeals).

[b] Indicates the approximate number of separate samples tested for prey DNA. Many studies analysed each sample with several different PCR tests.

Table 7.2 *Recent dietary studies*[a] *of vertebrate predators investigating the use of DNA-based methods for prey detection in stomach contents or faeces.*

Predator	Targeted Diet Items	Sample Tested	Gene	Marker Size	Sample Size[b]	Field-based	Reference
Whale, penguin, squid	Various prey items	Faeces soft matrix	Nuclear + mtDNA rDNA	220 – 370	5	Yes	Jarman *et al.* (2004)
Whale shark (*Rhincodon typus*)	Krill (Malacostraca)	Faeces soft matrix	Nuclear rDNA	330	3	Yes	Jarman & Wilson (2004)
Harbour seal *(Phoca vitulina*)	Salmonid species (*Onchorhynchus* spp.)	Faeces hard-parts	mtDNA COIII	370	39 (116 bones)	Yes	Purcell *et al.* (2004)
Steller sea lion (*Eumetopias jubatus*)	Fish and squid	Faeces soft matrix	Nuclear + mtDNA rDNA	180 – 250	194	No	Deagle *et al.* (2005b)
Harbour seal (*Phoca vitulina*)	Salmonid species (*Onchorhynchus* spp.)	Faeces hard-parts	Nuclear GH2 + Microsatellite	110 – 320	72 (751 bones)	Yes	Kvitrud *et al.* (2005)
Grey seal (*Halichoerus grypus*)	Atlantic salmon (*Salmo salar*)	Faeces soft matrix	mtDNA Cytb + rDNA	160 – 330	24	No	Parsons *et al.* (2005)
Large pelagic fish	Various prey items	Stomach prey items	mtDNA Cytb	450	12	Yes	Smith *et al.* (2005)
Macaroni penguin (*Eudyptes chrysolophus*)	Krill, fish and amphipods	Faeces soft matrix	mtDNA rDNA	170-280	88	Yes	Chapter 6 (Unpublished)

[a] Includes primary literature published in 2004 or 2005 and listed in the Web of Science© database, with the exception of information from Chapter 6 which was included for comparison. Studies based on the analysis of ancient remains were not included.

[b] Indicates the approximate number of separate samples tested for prey DNA. Many studies analysed each sample with several different PCR tests.

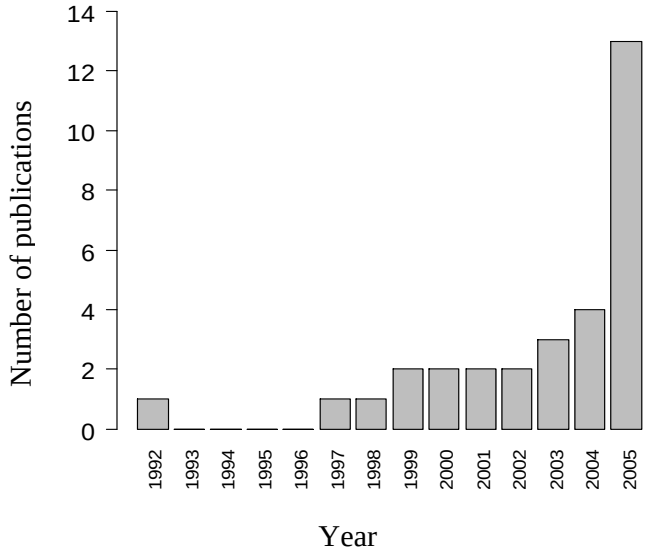


Figure 7.1 *The number of dietary studies using DNA-based identification methods published each year since the first study in 1992. This graph is based only on primary literature publications that are listed in the Web of Science© database. Diet studies based on analysis of ancient remains are excluded as are diet studies on parasites (e.g. studies analysing the source of insects' bloodmeals).*

7.3 General discussion and future directions

Interest in the use of DNA-based methods for dietary analysis has increased rapidly over the last few years (Figure 7.1), suggesting that these methods are likely to become much more commonly applied. Several options exist for studying the diet of large marine predators; each has strengths and weaknesses, and it is important to consider these in determining the best technique for any particular dietary study. DNA-based methods have some inherent limitations that will restrict their use. One obvious limitation, shared to some extent with non-genetic stomach content and faecal analysis, is that collection of samples (and therefore data) is restricted to times when

animals are feeding and accessible. In the case of many species of pinnipeds and seabirds, DNA-based diet studies will be possible only during the breeding season when the animals return to land. Some pinniped species do not feed during the breeding season (e.g. the southern elephant seal *Mirounga leonina*); it is unlikely that DNA-based methods could be used to study the diet of these predators. Alternative techniques such as analysis of fatty acids (Iverson *et al*. 2004; Bradshaw *et al*. 2003) or stable isotope analysis (Hobson *et al.* 1997) will need to be applied to determine diet when a predator's faeces or stomach contents cannot be collected. Another limitation of DNA-based methods is that the size and age classes of prey species being eaten cannot be determined. This information is often important (e.g. Reid *et al*. 2005) and can only be obtained in conventional stomach content or faecal diet studies.

The major strength of DNA-based techniques is they can provide rapid, accurate identification of otherwise unidentifiable prey remains. In dietary studies of large marine predators, the most straightforward application of DNA-based methodology has been identification of biological remains (tissue or hard-parts) recovered from stomach contents. This approach was applied to the giant squid gut sample analysed in Chapter 2, and has been used elsewhere to provide diet data in studies of seabirds (Scribner & Bowman 1998) and large pelagic fish (Smith *et al.* 2005). The quality and quantity of DNA recovered from partially digested remains in a predator's stomach should be significantly higher than the DNA recovered from faeces. Therefore, PCR amplifications from this template are generally reliable, and larger fragment sizes (i.e. >500 bp) can be targeted, allowing robust taxonomic classification of prey. In addition, the characterisation of the relatively pure DNA obtained from isolated prey remains is easier than similar analysis of mixtures of DNA (such as the prey DNA in faeces). Because this application is relatively uncomplicated and can increase the amount of data collected from each sample, the use of genetic identification methods will likely become routine in future studies that obtain stomach contents (especially when large numbers of prey remains are unidentifiable, or when samples from rare specimens are examined). DNA-based identification techniques have also been applied to hard-part remains in pinniped faecal samples (see section 7.1), and such applications are bound to continue. The drawback of focusing on physical prey remains is that the differential digestion of hard-parts will continue to bias results.

The greatest potential for DNA-based methods to advance dietary studies of pinnipeds and seabirds comes from their ability to retrieve information from the soft matrix component of faeces. In the case of pinnipeds, analysis of prey DNA in the faecal soft matrix will allow detection of prey species that are under-represented (or not represented) in conventional hard-part analysis. In seabirds, faecal analysis must rely on soft matrix (due to the absence of hard remains), and genetic methods provide a non-invasive and potentially less-biased approach compared to stomach content analysis. While the analysis of faecal soft matrix has been the primary focus of this thesis, the total number of studies carrying out this type of analysis has been very limited (see Table 1.2 and 7.2). Understandably, a large number of questions remain to be addressed in this new area of research.

7.3.1 Questions for future studies

An immediate focus for future research should be determining the generality of findings from initial diet studies focussing on prey DNA in the soft component of faeces. Additional controlled feeding trials similar to those on Steller sea lions (Chapter 2) and harbour seals (Parsons *et al.* 2005) should be carried out. This is especially important because the studies done to date have relied on only a few individual predators; therefore, the impact of intra-specific variation in predator physiology on DNA detection rates has not been considered. Feeding trials focusing on other pinniped species, and where possible on seabirds, are also of high importance. Ecological applications of the DNA-based techniques will undoubtedly focus on pinnipeds with low hard-part recovery rates (e.g. *Arctocephalus* seals; Casper *et al.* 2006) and on the detection of soft-bodied prey items, therefore trials with these predators and prey would be particularly useful. While feeding trials are the only way to answer some important questions, the critical test of the approach will come from its performance in field-based studies. There are currently no published field-based studies that have looked at recovery of prey DNA in the soft matrix component of pinniped faeces, addressing this deficit should be of top priority. Field studies should initially be carried out in parallel with other methods of diet analysis to allow for cross-validation of results (e.g. Chapter 6). After such validation, DNA-

based methods will be able to be used with confidence in situations where other methods of diet analysis cannot be applied.

A number of more specific questions could also be addressed to help advance the field. Potential detection of secondary predation events could be investigated (this has been shown to be a source of error in terrestrial invertebrate systems; Sheppard *et al.* 2005). Studies determining the duration of DNA signal in faeces during periods of fasting would also be informative (see discussion in Chapter 6). Due to the significance of sample quality in field-based studies, the post-defecation degradation of DNA in faeces should be further evaluated. One option for doing this would be by combining the time-course experiment of Chapter 3 with the method for quantification of DNA damage introduced in Chapter 5. The stability of DNA in samples should be considered both before and after preservation in ethanol.

7.3.2 Technical considerations

During the course of my thesis I have used several different genetic techniques to analyse faecal samples, and this raises the obvious question: what approach is the best one to use? The answer is dependent on the experimental question being asked and also on the scope of the planned study. Let's consider the options. All of the approaches that I applied involved amplification of the minute amount of prey DNA present in samples using PCR. The potential may exist for the development of non-PCR methods (e.g. ancient DNA from cave bear bones was recently cloned without amplification; Noonan *et al.* 2005), but it seems certain that PCR will continue to play a central role in most future DNA-based diet studies. I have previously distinguished between two general PCR approaches. In the first approach, PCR primers are designed to amplify DNA from a particular prey species (or wider taxonomic group) and successful PCR amplification from a sample indicates the presence of DNA from the target; I will refer to this as the targeted approach. The second approach utilizes primers which bind to DNA regions conserved in a broad range of prey items and the PCR products amplified with these conserved primers are subsequently characterised; this can be thought of as the exploratory approach. Both of these approaches are useful.

The targeted approach is useful for addressing specific dietary questions (i.e. does predator A consume prey B). The approach is appropriate when the range of potential prey is known *a priori*, or when particular prey species are of interest. Targeted PCR assays are relatively easy to develop and high-throughput sample processing can be implemented. For example, Harper *et al.* (2005) developed a multiplex PCR assay for detection of prey DNA in stomachs of invertebrate predators. In this multiplex PCR, tests for ten prey species are carried out simultaneously and a single person can process 200 samples per day. In the marine environment, these types of PCR tests could be developed for the detection of commercially exploited fish species and applied to faecal samples from various predators in order to determine the potential for direct competition between the predators and fisheries. Data from the macaroni penguin diet study (Chapter 6) show that the targeted PCR approach can also be useful for detection of dietary shifts, suggesting a potential application in long-term studies monitoring changes in diet (e.g. Reid *et al.* 2005). The obvious drawback of the targeted approach is that you only find prey species you specifically look for. Therefore, if the experimental system is not well understood, important prey items could be missed. An additional difficulty with the targeted approach is that a negative PCR result does not necessarily mean absence of that prey item in the sample, since the failure of PCR amplification can result from numerous factors. The use of appropriate internal positive control assays is the only way negative occurrences can be verified (this problem is discussed in Chapter 3).

In an ideal situation the exploratory PCR approach could be used to amplify all prey DNA in a sample and these molecules could then be identified in a single analytical step. This would allow for discovery of previously unknown prey items, and negative results could be logically interpreted (i.e. samples that contain no amplifiable prey DNA could be removed from further analysis). Unfortunately, this approach is hindered by the preponderance of DNA from the predator in dietary samples. I found that prey DNA could not be reliably detected using conserved primers (Chapter 2) due to swamping by predator DNA – there was, on average, 25 times more sea lion DNA than prey DNA in the sea lion faecal samples collected

during the feeding trial[14] (Chapter 5). Through the use of group-specific primers, prey DNA can be preferentially amplified (Jarman *et al.* 2004), but to characterise all potential prey, multiple group-specific primers need to be used. In addition, when predators consume prey closely related to themselves (e.g. squid commonly eat other squid) the group-specific approach becomes untenable. There are several other methods that could possibly be used to selectively remove predator DNA from samples. This includes using restriction enzymes to digest the predator DNA (Blankenship & Yayanos 2005), subtractive hybridization methods and amplification blocking procedures (Dominguez & Kolodney 2005). The development of these methods in the context of DNA-based diet studies would be very useful. The design of non-degenerate conserved primers amplifying short taxonomically informative DNA regions is also critical. Without these primers, data obtained from the exploratory approach can be biased by species-specific differences in primer binding efficiency (see discussion in Chapter 6).

If we assume that a wide range of prey DNA can be recovered from dietary samples using the exploratory PCR approach, the process of identifying the species represented can be carried out using a variety of methods. The most straightforward approach, and the one providing the maximum amount of data, is to clone and sequence the PCR products. Because these techniques are expensive, and sample processing is relatively time-consuming, this method is most suited to small studies looking at important dietary questions. Other methods used to characterise mixed PCR products require more initial investment in their development, but have the potential to be cheaper and faster for studies with large numbers of samples. Potential methods that could be used include: RFLP analysis, AFLP analysis, DGGE and hybridization techniques. RFLP analysis was used by Sutherland (2000); he found interpretation of multiple prey RFLP signatures could be difficult and confidence in the resulting identifications was often low. Therefore, this simple approach can only be recommended if the diversity of prey within samples is very limited. AFLP analysis relies on prey-specific differences in the amplified fragment length. High resolution separation of fluorescent-labelled fragments can be accomplished on a sequencer allowing simple interpretation of results (e.g. Harper *et al.* 2005). However,

[14] For a 300 bp marker, making the assumption that the herring fed as 50% of the diet represented 50% of the prey DNA in the faeces.

to obtain the reasonable taxonomic resolution, the PCR target would have to be a hyper-variable gene region. The size of the variable fragment amplified using the primers16S1F and 16S2R (Table 2.1) could be used to discriminate between fish and squid, but within these groups many species would not be differentiated. The AFLP approach is probably best suited to studies where the potential prey species are well characterised, and in these situations it is debatable whether it should be used preferentially over a multiplex-PCR targeted approach (Harper *et al.* 2005).

The DGGE technique for separating PCR products has very high resolution (i.e. PCR products differing by only a single base substitution can be detected). I used DGGE in the sea lion feeding trial for simultaneous detection of DNA from five fish species. The approach could be applied to much more diverse diets, and can be used to identify unknown prey species. The high resolution of the approach does mean that intra-specific variation may be detected[15], which will result in an increase in the amount of sequence analysis required to characterise these "new" bands. I also found that heteroduplex band formation can lead to misinterpretation of results, although the extent of this problem was not determined. Another difficulty is that nested-PCR may be required for consistent amplification when using a clamped primer (Sheffield *et al.* 1989). This leads to increased cost, and a greater risk of sample contamination (see Appendix I for discussion of nested PCR). Even with these complications, I think DGGE can be successfully applied in future diet studies of generalist predators. Particularly in large scale, or long-term, studies where a large amount of technique development can be justified. The primary reason I avoided the technique in the macaroni penguin study (Chapter 6) was lack of sufficient time for assay development. Lack of time was also why I did not explore the use of hybridization techniques. Hybridization techniques similar to those employed by Rosel & Kocher (2002), or using microarray technology (Gibson 2002), have great potential for use in prey DNA detection. Like DGGE, development of these methods will require considerable initial investment.

[15] When screening clone libraries, sequence variants created through PCR induced errors may also be detected (e.g. Chapter 2). However, my experience suggests that in products amplified directly from scat DNA template, the ratio of undamaged to damaged template is high enough to ensure that the damaged templates are not visible on standard gels.

7.3.3 Future of DNA-based quantitative diet estimates

Perhaps the biggest challenge for future DNA-based diet studies will be the development of methods for obtaining quantitative diet estimates. Without the ability to obtain at least semi-quantitative data, the application of DNA-based techniques will be limited. In this thesis, I provided an initial assessment of an approach for obtaining quantitative diet data based on measuring the amount of DNA in faecal samples (Chapter 4). The results were quite promising, but analysis was restricted to a small number of samples and a limited number of prey species. Further studies quantifying DNA from a diverse group of prey in faecal samples from a range of predator species are required to determine the generality of these preliminary findings. Particular emphasis should focus on determining tissue DNA density of different prey groups, and on investigating the magnitude of the bias caused by differential tissue digestion. The species-specific qPCR assays I employed would be impractical for field-based diet studies of generalist predators due to the cost and technical difficulty of quantifying DNA from numerous potential prey species. More general comparisons could be made using group-specific primers (Jarman et al. 2004) allowing broad dietary questions to be addressed with only a few qPCR assays. For example, where seal predation on salmon is a major concern (Kvitrud *et al*. 2005) the amount of salmonid DNA versus total fish DNA could be measured in seal faecal samples to better evaluate salmon consumption. Another currently feasible application using qPCR would be in the long-term Antarctic ecosystem monitoring program looking at the diet of Adélie penguins (Agnew 1997). The diversity of prey species in the diet of these penguins is limited, and simply measuring the amounts of krill and fish in the diet can provide useful ecological information. If many of the penguin faecal samples contained both prey groups a quantitative assay could be invaluable. Widespread application the DNA quantification approach to determining diet composition is ultimately dependent on methods becoming more accessible, through technological improvements in qPCR and/or microarray technology.

Assays based on presence/absence detection of prey DNA in faeces could also be used to obtain quantitative estimates of diet composition. The frequency of occurrence data produced by these assays will tend to give biased quantitative estimates because the importance of prey items taken in small quantities is

overestimated (e.g. in a sample containing 90% fish and 10% krill, the importance of these species will be considered equal). Still, these data have been widely used in traditional diet studies and strong biases in individual samples will become less important when a large number of samples are examined (Cullen *et al.* 1992; Sinclair & Zeppelin 2002). This frequency of occurrence approach cannot be directly evaluated through feeding trials because the feeding regime is artificial. In a case where all meals in a feeding trial contain all prey species in the diet, you would expect to get equal occurrences for prey regardless of relative amounts of the prey fed (e.g. results from Chapter 3). Alternatively, if diet is fed in monospecific meals (i.e. a prey item making up 10% of the diet is fed every tenth meal), then frequency of occurrence data would likely give good quantitative estimates of diet. Field studies comparing traditional diet data and genetic data from naturally foraging predators will be the best way to evaluate the approach. It should also be noted that if quantitative information is to be inferred from genetic presence/absence data, comparability of the data needs to be carefully considered (especially when it is obtained using multiple PCR primer sets). Differences in sensitivities of the PCR assays (due to variation in primer binding efficiencies, gene copy number and/or target fragment size) will result in the DNA from some prey being detected more easily (and over a longer time period) than others.

7.3.4 Concluding remarks

Despite the large number of questions that remain to be answered in the emerging field of DNA-based diet research, optimism for the future is warranted. The series of studies contained here show that valuable dietary information can be obtained from the DNA present in stomach contents or faeces of a variety of marine predators. Additionally, the rapid development of genetic technology promises a future where laboratory analysis is faster, cheaper and more powerful. A glimpse of this future can be obtained by looking at recent publications that have applied some emerging genetic techniques. In a study using microarray technology for DNA-based identification of microbes, a rapid assay has been developed that allows simultaneous detection and quantification of over 7000 unique DNA sequences (Palmer *et al.* 2006). This microarray assay is also sensitive enough to enable detection of DNA

from bacterial species present at a proportional abundance of < 0.l% of the total mixture. Recent advances in the development of ultrafast DNA sequencing technologies are equally impressive; one laboratory demonstrated a new method by sequencing the entire 580,000 bp genome of a bacteria species in a single four hour run, at a cost of less than 10 cents per 1000 bp (Margulies *et al.* 2005). While the advances in genetics that make DNA-based diet analysis possible are extraordinary, the use of elaborate technology in itself does not make the approach more exact or better than traditional methods of diet analysis. As pointed out by Avise (1994):

"*Molecular markers are used most intelligently when they address controversial areas or when they are employed to analyse problems of natural history and evolution that have proven beyond the purview of traditional non-molecular observation.*"

There are numerous questions in the field of ecological dietary research that are beyond the purview of traditional methods of diet analysis. The continued development and sensible application of DNA-based methods should allow many of these to be answered.

References

Abath FGC, Melo FL, Werkhauser RP, *et al*. (2002) Single-tube nested PCR using immobilized internal primers. *BioTechniques*, **33**, 1210-1214.

Abu Al-Soud W, Rådström P (1998) Capacity of nine thermostable DNA polymerases to mediate DNA amplification in the presence of PCR-inhibiting samples. *Applied and Environmental Microbiology*, **64**, 3748-3753.

Abu Al-Soud W, Rådström P (2001) Purification and characterization of PCR-inhibitory components in blood cells. *Journal of Clinical Microbiology*, **39**, 485-493.

Agnew DJ (1997). The CCAMLR Ecosystem Monitoring Programme. *Antarctic Science*, **9**, 235-242.

Agustí N, De Vicente MC, Gabarra R (1999) Development of sequence amplified characterized region (SCAR) markers of *Helicoverpa armigera*: a new polymerase chain reaction-based technique for predator gut analysis. *Molecular Ecology*, **8**, 1467-1474.

Agustí N, de Vicente MC, Gabarra R (2000) Developing SCAR markers to study predation on *Trialeurodes vaporariorum*. *Insect Molecular Biology*, **9**, 263-268.

Agustí N, Shayler SP, Harwood JD, *et al*. (2003a) Collembola as alternative prey sustaining spiders in arable ecosystems: Prey detection within predators using molecular markers. *Molecular Ecology*, **12**, 3467-3475.

Agustí N, Unruh TR, Welter SC (2003b) Detecting *Cacopsylla pyricola* (Hemiptera : Psyllidae) in predator guts using COI mitochondrial markers. *Bulletin of Entomological Research*, **93**, 179-185.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403-410.

Aron W, Burke W, Freeman MMR (2000) The whaling issue. *Marine Policy*, **24**, 501-502.

Asahida T, Yamashita Y, Kobayashi T (1997) Identification of consumed stone flounder, *Kareius bicoloratus* (Basilewsky), from the stomach contents of sand shrimp, Crangon affinis (De Haan) using mitochondrial DNA analysis. *Journal of Experimental Marine Biology and Ecology*, **217**, 153-163.

Ayala-Torres S, Chen YM, Svoboda T, Rosenblatt J, Van Houten B (2000) Analysis of gene-specific DNA damage and repair using quantitative polymerase chain reaction. *Methods-a Companion to Methods in Enzymology*, **22**, 135-147.

Barlow KE, Croxall JP (2002) Provisioning behaviour of Macaroni Penguins *Eudyptes chrysolophus*. *Ibis*, **144**, 248-258.

Binaschi M, Farinosi R, Borgnetto ME, Capranico G (2000) Detection of cellular DNA cleavage using non-proofreading thermostable DNA polymerases. *Biotechniques*, **28**, 1064-1066.

Birnbolm HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*, **7**, 1513-1523.

Bishop DG, Ritz DA, Hosie GW, Kenrick JR, Olley J (1983) Fatty-Acid Composition of the Lipids of *Puffinus tenuirostris* (Temminck) in Relation to Its Diet. *Journal of Experimental Marine Biology and Ecology*, **71**, 17-26.

Bjourson AJ, Cooper JE (1992) Band-stab PCR: a simple technique for the purification of individual PCR products. *Nucleic Acids Research*, **20**, 4675.

Blankenship LE, Yayanos AA (2005) Universal primers and PCR of gut contents to study marine invertebrate diets. *Molecular Ecology*, **14**, 891-899.

Blaxter M, Mann J, Chapman T, *et al.* (2005) Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **360**, 1935-1943.

Bocher P, Cherel Y, Hobson KA (2000) Complete trophic segregation between South Georgian and common diving petrels during breeding at Iles Kerguelen. *Marine Ecology-Progress Series*, **208**, 249-264.

Bolstad KS, O'Shea S (2004) Gut contents of a giant squid *Architeuthis dux* (Cephalopoda: Oegopsida) from New Zealand waters. *New Zealand Journal of Zoology*, **31**, 15-21.

Boon N, De Windt W, Verstraete W, Top EM (2002) Evaluation of nested PCR-DGGE (denaturing gradient gel electrophoresis) with group-specific 16S rRNA primers for the analysis of bacterial communities from different wastewater treatment plants. *FEMS Microbiology Ecology*, **39**, 101-112.

Bowen WD (2000) Reconstruction of pinniped diets: accounting for complete digestion of otoliths and cephalopod beaks. *Canadian Journal of Fisheries and Aquatic Sciences*, **57**, 898-905.

Bowen WD, Tully D, Boness DJ, Bulheier BM, Marshall GJ (2002) Prey-dependent foraging tactics and prey profitability in a marine mammal. *Marine Ecology Progress Series*, **244**, 235-245.

Bradshaw CJA, Hindell MA, Best NJ, *et al.* (2003) You are what you eat: Describing the foraging ecology of southern elephant seals (*Mirounga leonina*) using blubber fatty acids. *Proceedings of the Royal Society of London - Biological Sciences*, **270**, 1283-1292.

Braker G, Ayala-del-Rio HL, Devol AH, Fesefeldt A, Tiedje JM (2001) Community structure of denitrifiers, Bacteria, and Archaea along redox gradients in pacific northwest marine sediments by terminal restriction fragment length polymorphism analysis of amplified nitrite reductase (nirS) and 16S rRNA genes. *Applied and Environmental Microbiology*, **67**, 1893-1901.

Brix O (1983) Giant squids may die when exposed to warm water currents. *Nature*, **303**, 422-423.

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

Brown CR, Klages NT (1987) Seasonal and Annual Variation in Diets of Macaroni (*Eudyptes chrysolophus chrysolophus*) and Southern Rockhopper (*E. chrysocome chrysocome*) Penguins at Sub-Antarctic Marion Island. *Journal of Zoology*, **212**, 7-28.

Browne P, Laake JL, DeLong RL (2002) Improving pinniped diet analyses through identification of multiple skeletal structures in fecal samples. *Fishery Bulletin*, **100**, 423-433.

Casper RM, Gales NJ, Hindell MA, Robinson SM (2006) Diet estimation based on an integrated mixed prey feeding experiment using *Arctocephalus* seals. *Journal of Experimental Marine Biology and Ecology*, **328**, 228-239.

Cedervall B, Wong R, Albright N, *et al.* (1995) Methods for the Quantification of DNA Double-Strand Breaks Determined from the Distribution of DNA Fragment Sizes Measured by Pulsed-Field Gel-Electrophoresis. *Radiation Research*, **143**, 8-16.

Chen Y, Giles KL, Payton ME, Greenstone MH (2000) Identifying key cereal aphid predators by molecular gut analysis. *Molecular Ecology*, **9**, 1887-1898.

Cherel Y, Hobson KA, Weimerskirch H (2000) Using stable-isotope analysis of feathers to distinguish moulting and breeding origins of seabirds. *Oecologia*, **122**, 155-162.

Chiaradia A, Costalunga A, Kerry K (2003) The diet of Little Penguins (*Eudyptula minor*) at Phillip Island, Victoria, in the absence of a major prey - Pilchard (*Sardinops sagax*). *Emu*, **103**, 43-48.

Clarke KR (1993) Nonparametric Multivariate Analyses of Changes in Community Structure. *Australian Journal of Ecology*, **18**, 117-143.

Clarke MR (1966) A review of the systematics and ecology of the oceanic squids. *Advances in Marine Biology*, **4**, 91-300.

Clarke MR MacLeod N (1982) Cephalopod remains from the stomachs of sperm whales caught in the Tasman Sea. *Memoirs of the National Museum Victoria*, **43**, 25-42.

Clarke MR (1985) *A handbook for the identification of squid beaks.* Clarendon Press, Oxford (UK).

Clarke M, Young R (1998) Description and analysis of cephalopod beaks from stomachs of six species of odontocete cetaceans stranded on Hawaiian shores.

Journal of the Marine Biological Association of the United Kingdom, **78**, 623-641.

Coombs NJ, Gough AC, Primrose JN (1999) Optimisation of DNA and RNA extraction from archival formalin-fixed tissue. *Nucleic Acids Research*, **27**, e12.

Cooper A, Poinar HN (2000) Ancient DNA: Do it right or not at ALL. *Science*, **289**, 1139-1139.

Cooper J, Brown CR, Gales RP, *et al.* (1990) Diets and Dietary Segregation of Crested Penguins (*Eudyptes*). In: *Penguin Biology* (eds. Davis LS, Darby JT). Academic Press, San Diego, California.

Crawford RJM, Cooper J, Dyer BM (2003) Population of the Macaroni penguin *Eudyptes chrysolophus* at Marion Island, 1994/95-2002/03, with information on breeding and diet. *African Journal of Marine Science*, **25**, 475-486.

Croxall JP, Prince PA (1980) The food of Gentoo Penguins *Pygoscelis papua* and macaroni penguins *Eudyptes chrysolophus* at South Georgia. *Ibis*, **122**, 245-253.

Croxall JP, Prince PA, Baird A, Ward P (1985) The diet of the Southern rockhopper penguin *Eudyptes chrysocome chrysocome* at Beauchêne Islands, Falkland Islands. *Journal of Zoology*, **206**, 485-496.

Cullen JM, Montague TL, Hull C (1992) Food of Little Penguins *Eudyptula minor* in Victoria - Comparison of 3 Localities between 1985 and 1988. *Emu*, **91**, 318-341.

Dalén L, Gotherstrom A, Angerbjorn A (2004) Identifying species from pieces of faeces. *Conservation Genetics*, **5,** 109-111.

Da Silva J, Neilson JD (1985) Limitations of using otoliths recovered in scats to estimate prey consumption in seals. *Canadian Journal of Fisheries & Aquatic Sciences*, **42**, 1439-1442.

David JHM, Cury P, Crawford RJM, *et al.* (2003) Assessing conservation priorities in the Benguela ecosystem, South Africa: analysing predation by seals on threatened seabirds. *Biological Conservation*, **114**, 289-292.

Davison A, Birks JDS, Brookes RC, Braithwaite TC, Messenger JE (2002) On the origin of faeces: morphological versus molecular methods for surveying rare carnivores from their scats. *Journal of Zoology (London)*, **257**, 141-143.

Deagle BE, Bax N, Hewitt CL, Patil JG (2003) Development and evaluation of a PCR-based test for detection of Asterias (Echinodermata: Asteroidea) larvae in Australian plankton samples from ballast water. *Marine and Freshwater Research*, **54**, 709-719.

Deagle BE, Jarman SN, Pemberton D, Gales NJ (2005a) Genetic screening for prey in the gut contents from a giant squid (*Architeuthis* sp.). *Journal of Heredity*, **96**, 417-423.

Deagle BE, Tollit DJ, Jarman SN, *et al.* (2005b) Molecular scatology as a tool to study diet: analysis of prey DNA in scats from captive Steller sea lions. *Molecular Ecology*, **14**, 1831-1842.

Deagle BE, Tollit DJ (2006) Quantitative analysis of prey DNA in pinniped faeces: potential to estimate diet composition? *Conservation Genetics*, In press

Delong EF, Wu KY, Prezelin BB, Jovine RVM (1994) High Abundance of Archaea in Antarctic Marine Picoplankton. *Nature*, **371**, 695-697.

DeWoody JA, Fletcher DE, Wilkins SD, Avise JC (2001) Genetic documentation of filial cannibalism in nature. *Proceedings of the National Academy of Sciences of the USA*, **98**, 5090-5092.

Dominguez PL, Kolodney MS (2005) Wild-type blocking polymerase chain reaction for detection of single nucleotide minority mutations from clinical specimens. *Oncogene*, **24**, 6830-6834.

Doyle JJ, Doyle JL (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemistry Bulletin*, **19**, 11-15.

Ealey EHM (1954) Analysis of stomach content of some Heard Island birds. *Emu*, **54**, 204-210.

Eckert KA, Opresko PL (1999) DNA polymerase mutagenic bypass and proofreading of endogenous DNA lesions. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis*, **424**, 221-236.

Ellis, R (1995) *Monsters of the sea*. Robert Hale, London.

Farrell LE, Romant J, Sunquist ME (2000) Dietary separation of sympatric carnivores identified by molecular analysis of scats. *Molecular Ecology*, **9**, 1583-1590.

Fernando LP, Kurian PJ, Fidan M, Fernandes DJ (2002) Quantitation of gene-specific DNA damage by competitive PCR. *Analytical Biochemistry*, **306**, 212-221.

Fierer N, Jackson JA, Vilgalys R, Jackson RB (2005) Assessment of soil microbial community structure by use of taxon-specific quantitative PCR assays. *Applied and Environmental Microbiology*, **71**, 4117-4120.

Fischer W, Hureau JC (1985) *FAO species identification sheets for fisheries purposes. Southern Ocean (Fishing areas 48, 58 and 88 CCAMLR convention area).* Food and Agriculture Organization of the United Nations, Rome.

Foltan P, Sheppard S, Konvicka M, Symondson WOC (2005) The significance of facultative scavenging in generalist predator nutrition: detecting decayed prey in the guts of predators using PCR. *Molecular Ecology*, **14**, 4147-4158.

Förch EC (1998) *The marine fauna of New Zealand: Cephalopoda: Oegopsida: Architeuthidae (giant squid).* National Institute of Water and Atmospheric Research Biodiversity Memoir 110, Wellington, New Zealand.

Frantzen MAJ, Silk JB, Ferguson JWH, Wayne RK, Kohn MH (1998) Empirical evaluation of preservation methods for faecal DNA. *Molecular Ecology*, **7**, 1423-1428.

Gales NJ, Cheal AJ (1992) Estimating Diet Composition of the Australian Sea-Lion (*Neophoca cinerea*) from Scat Analysis - an Unreliable Technique. *Wildlife Research*, **19**, 447-456.

Gales RP (1988) The Use of Otoliths as Indicators of Little Penguin *Eudyptula minor* Diet. *Ibis*, **130**, 418-426.

Garcia SM, Cochrane KL (2005) Ecosystem approach to fisheries: a review of implementation guidelines. *Ices Journal of Marine Science*, **62**, 311-318.

Gibson G (2002) Microarrays in ecology and evolution: a preview. *Molecular Ecology*, **11**, 17-24.

Gilbert MTP, Bandelt HJ, Hofreiter M, Barnes I (2005) Assessing ancient DNA studies. *Trends in Ecology & Evolution*, **20**, 541-544.

Gilbert MTP, Hansen AJ, Willerslev E, *et al.* (2003) Characterization of genetic miscoding lesions caused by postmortem damage. *American Journal of Human Genetics*, **72**, 48-61.

Glenn TC, Stephan W, Braun MJ (1999) Effects of a population bottleneck on Whooping Crane mitochondrial DNA variation. *Conservation Biology*, **13**, 1097-1107.

Golenberg EM, Bickel A, Weihs P (1996) Effect of highly fragmented DNA on PCR. *Nucleic Acids Research*, **24**, 5026-5033.

Govan HL, Vallesayoub Y, Braun J (1990) Fine-Mapping of DNA Damage and Repair in Specific Genomic Segments. *Nucleic Acids Research*, **18**, 3823-3830.

Grahl-Nielsen O (1999) Comment: Fatty acid signatures and classification trees: new tools for investigating the foraging ecology of seals. *Canadian Journal of Fisheries and Aquatic Sciences*, **56**, 2219-2223.

Grahl-Nielsen O, Andersen M, Derocher AE, *et al.* (2003) Fatty acid composition of the adipose tissue of polar bears and of their prey: ringed seals, bearded seals and harp seals. *Marine Ecology-Progress Series*, **265**, 275-282.

Green K, Williams R, Green MG (1998) Foraging ecology and diving behaviour of macaroni penguins *Eudyptes chrysolophus* over the breeding season of 1992/93 at Heard Island. *Marine Ornithololog*y., **26**, 27-34.

Gunn JS, Bruce BD, Furlani DM, Thresher RE, Blaber SJM (1989) Timing and location of spawning blue grenadier, *Macruronus novaezelandiae* (Teleostei, Merlucciidae), in Australian costal waters. *Australian Journal of Marine and Freshwater Research*, **40**, 97-112.

Hall AJ, Watkins J, Hammond PS (1998) Seasonal variation in the diet of harbour seals in the south-western North Sea. *Marine Ecology-Progress Series*, **170**, 269-281.

Hammond PS, Rothery P (1996) Application of computer sampling in the estimation of seal diet. *Journal of Applied Statistics*, **23**, 525-533.

Hansen MM, Jacobsen L (1999) Identification of mustelid species: otter (*Lutra lutra*), American mink (*Mustela vison*) and polecat (*Mustela putorius*), by analysis of DNA from faecal samples. *Journal of Zoology (London)*, **247**, 177-181.

Harcourt RG, Bradshaw CJA, Dickson K, Davis LS (2002) Foraging ecology of a generalist predator, the female New Zealand fur seal. *Marine Ecology-Progress Series*, **227**, 11-24.

Harper GL, King RA, Dodd CS, *et al.* (2005) Rapid screening of invertebrate predators for multiple prey DNA targets. *Molecular Ecology*, **14**, 819-827.

Hartley PHT (1948) The assessment of the food of birds. *Ibis*, **90**.

Harvey JT (1989) Assessment of errors associated with harbour seal (Phoca vitulina) faecal sampling. *Journal of Zoology*, **219**, 101-111.

Harvey JT, Antonelis GA (1994) Biases Associated with Nonlethal Methods of Determining the Diet of Northern Elephant Seals. *Marine Mammal Science*, **10**, 178-187.

Harwood J, Croxall JP (1988) The assessment of competition between seals and commercial fisheries in the North-Sea and the Antarctic. *Marine Mammal Science*, **4**, 13-33.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London - Biological Sciences*, **270**, 313-321.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through DNA barcodes. *Plos Biology*, **2**, 1657-1663.

Hindell MA (1988) The Diet of the Royal Penguin *Eudyptes schlegeli* at Macquarie-Island. *Emu*, **88**, 219-226.

Hobson KA, Sease JL, Merrick RL, Piatt JF (1997) Investigating trophic relationships of pinnipeds in Alaska and Washington using stable isotope ratios of nitrogen and carbon. *Marine Mammal Science*, **13**, 114-132.

Hodum PJ, Hobson KA (2000) Trophic relationships among Antarctic fulmarine petrels: insights into dietary overlap and chick provisioning strategies inferred from stable-isotope (delta N-15 and delta C-13) analyses. *Marine Ecology-Progress Series*, **198**, 273-281.

Hofreiter M, Jaenicke V, Serre D, von Haeseler A, Paabo S (2001) DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research*, **29**, 4793-4799.

Hofreiter M, Poinar HN, Spaulding WG, *et al.* (2000) A molecular analysis of ground sloth diet through the last glaciation. *Molecular Ecology*, **9**, 1975-1984.

Hogg RV, Tanis EA (2005) *Probability and Statistical Inference*, 7th Edition. Prentice Hall, New Jersey.Hoogendoorn M, Heimpel GE (2001) PCR-based gut content analysis of insect predators: using ribosomal ITS-1 fragments from prey to estimate predation frequency. *Molecular Ecology*, **10**, 2059-2067.

Hooker SK, Iverson SJ, Ostrom P, Smith SC (2001) Diet of northern bottlenose whales inferred from fatty-acid and stable-isotope analyses of biopsy samples. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, **79**, 1442-1454.

Höss M, Jaruga P, Zastawny TH, Dizdaroglu M, Pääbo S (1996) DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Research*, **24**, 1304-1307.

Höss M, Kohn M, Pääbo S, Knauer F, Schroder W (1992) Excrement Analysis by PCR. *Nature*, **359**, 199-199.

Hull CL (1999) Comparison of the diets of breeding royal (*Eudyptes schlegeli*) and rockhopper (*Eudyptes chrysocome*) penguins on Macquarie Island over three years. *Journal of Zoology*, **247**, 507-529.

Hyslop EJ (1980) Stomach content analysis - a review of methods and their application. *Journal of Fish Biology*, **17**, 411-429.

Ichii T, Kato H (1991) Food and Daily Food-Consumption of Southern Minke Whales in the Antarctic. *Polar Biology*, **11**, 479-487.

Iverson SJ, Field C, Blanchard W, Bowen WD (2004) Quantitative fatty acid signature analysis: A new method of estimating predator diets. *Ecological Monographs*, **74**, 211-235.

Iverson SJ, Frost KJ, Lowry LF (1997) Fatty acid signatures reveal fine scale structure of foraging distribution of harbor seals and their prey in Prince William Sound, Alaska. *Marine Ecology-Progress Series*, **151**, 255-271.

Jarman SN, Gales NJ, Tierney M, Gill PC, Elliott NG (2002) A DNA-based method for identification of krill species and its application to analysing the diet of marine vertebrate predators. *Molecular Ecology*, **11**, 2679-2690.

Jarman SN, Deagle BE, Gales NJ (2004) Group-specific polymerase chain reaction for DNA-based analysis of species diversity and identity in dietary samples. *Molecular Ecology*, **13**, 1313-1322.

Jarman SN, Wilson SG (2004) DNA-based species identification of krill consumed by whale sharks. *Journal of Fish Biology*, **65**, 586-591.

Jarman SN, Redd KS, Gales NJ (2006) Group-specific primers for amplifying DNA sequences that identify Amphipoda, Cephalopoda, Echinodermata, Gastropoda, Isopoda, Ostracoda and Thoracica. *Molecular Ecology Notes*, **6**, 268-271.

Juen A, Traugott M (2005) Detecting predation and scavenging by DNA gut-content analysis: a case study using a soil insect predator-prey system. *Oecologia*, **142**, 344-352.

Jennerwein MM, Eastman A (1991) A Polymerase Chain Reaction-Based Method to Detect Cisplatin Adducts in Specific Genes. *Nucleic Acids Research*, **19**, 6209-6214.

Jobling M (1987) Marine mammal faeces samples as indicators of prey importance - A source of error in bioenergetics studies. *Sarsia*, **72**, 255-260.

Kanbe T, Yamaki K, Kikuchi A (2002) Identification of the pathogenic *Aspergillus* species by nested PCR using a mixture of specific primers to DNA topoisomerase II gene. *Microbiology and Immunology*, **46**, 841-848.

Kasper ML, Reeson AF, Cooper SJB, Perry KD, Austin AD (2004) Assessment of prey overlap between a native (*Polistes humilis*) and an introduced (*Vespula germanica*) social wasp using morphology and phylogenetic analyses of 16S rDNA. *Molecular Ecology*, **13**, 2037-2048.

Kear AJ (1992) The diet of Antarctic squid: comparison of conventional and serological gut contents analyses. *Journal of Experimental Marine Biology and Ecology*, **156**, 161-178.

Kimura M (1980) A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111-120.

Klages NTW, Gales RP, Pemberton D (1989) Dietary Segregation of Macaroni and Rockhopper Penguins at Heard Island. *Australian Wildlife Research*, **16**, 599-604.

Kohn M, Knauer F, Stoffella A, Schroder W, Paabo S (1995) Conservation Genetics of the European Brown Bear - a Study Using Excremental PCR of Nuclear and Mitochondrial Sequences. *Molecular Ecology*, **4**, 95-103.

Kotz S, Balakrishnan N, Johnson NL (2000) *Continuous Multivariate Distributions*, 2nd edition. John Wiley & Sons, New York.

Kubodera T, Mori K (2005) First-ever observations of a live giant squid in the wild. *Proceedings of the Royal Society B-Biological Sciences*, **272**, 2583-2586.

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: Molecular Evolutionary Genetics Analysis software. *Bioinformatics*, **17**, 1244-1245.

Kvitrud MA, Riemer SD, Brown RF, Bellinger MR, Banks MA (2005) Pacific harbor seals (*Phoca vitulina*) and salmon: genetics presents hard numbers for elucidating predator-prey dynamics. *Marine Biology*, **147**, 1459 -1466.

Laake JL, Browne P, DeLong RL, Huber HR (2002) Pinniped diet composition: a comparison of estimation models. *Fishery Bulletin*, **100**, 434-447.

Lake S, Burton H, van den Hoff J (2003) Regional, temporal and fine-scale spatial variation in Weddell seal diet at four coastal locations in east Antarctica. *Marine Ecology-Progress Series*, **254**, 293-305.

Lescroel A, Ridoux V, Bost CA (2004) Spatial and temporal variation in the diet of the gentoo penguin (*Pygoscelis papua*) at Kerguelen Islands. *Polar Biology*, **27**, 206-216.

Lessa EP, Applebalm G (1993) Screening techniques for detecting allelic variation in DNA sequences. *Molecular Ecology*, **2**, 119-129.

Lindahl T (1993) Instability and Decay of the Primary Structure of DNA. *Nature*, **362**, 709-715.

Lordan C, Collins MA, Perales-Raya C (1998) Observations on morphology, age and diet of three *Architeuthis* caught off the west coast of Ireland in 1995. *Journal of the Marine Biological Association of the United Kingdom*, **78**, 903-917.

Lucchini V, Fabbri E, Marucco F *et al.* (2002) Noninvasive molecular tracking of colonizing wolf (*Canis lupus*) packs in the western Italian Alps. *Molecular Ecology*, **11**, 857-868.

Ma J, Li D, Keller M, Schmidt O, Feng X (2005) A DNA marker to identify predation of *Plutella xylostella* (Lep., Plutellidae) by *Nabis kinbergii* (Hem., Nabidae) and *Lycosa* sp (Aranaea, Lycosidae). *Journal of Applied Entomology*, **129**, 330-335.

Mambo E, Gao XQ, Cohen Y, *et al.* (2003) Electrophile and oxidant damage of mitochondrial DNA leading to rapid evolution of homoplasmic mutations. *Proceedings of the National Academy of Sciences of the USA*, **100**, 1838-1843.

Mapstone BD (1995) Scalable decision rules for environmental impact studies: effect size, Type I, and Type II errors. *Ecological Applications*, **5**, 401-410.

Margulies M, Egholm M, Altman WE, *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376-380.

Marota I, Basile C, Ubaldi M, Rollo F (2002) DNA decay rate in papyri and human remains from Egyptian archaeological sites. *American Journal of Physical Anthropology*, **117**, 310-318.

May JL, Blaber SJM (1989) Benthic and pelagic fish biomass of the upper continental slope off eastern Tasmania. *Marine Biology*, **101**, 11-25.

Merrick RL, Chumbley MK, Byrd GV (1997) Diet diversity of Steller sea lions (Eumetopias jubatus) and their population decline in Alaska: A potential relationship. *Canadian Journal of Fisheries and Aquatic Sciences*, **54**, 1342-1348.

Middleton SA, Anzenberger G, Knapp LA (2004) Denaturing gradient gel electrophoresis (DGGE) screening of clones prior to sequencing. *Molecular Ecology Notes*, **4**, 776-778.

Miller FJ, Rosenfeldt FL, Zhang CF, Linnane AW, Nagley P (2003) Precise determination of mitochondrial DNA copy number in human skeletal and cardiac muscle by a PCR-based assay: lack of change of copy number with age. *Nucleic Acids Research*, **31**, e61.

Miserez R, Pilloud T, Cheng X, *et al.* (1997) Development of a sensitive nested PCR method for the specific detection of *Mycoplasma mycoides* subsp. *mycoides* SC. *Molecular and Cellular Probes*,**11**, 103-111.

Mitchell D, Willerslev E, Hansen A (2005) Damage and repair of ancient DNA. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis*, **571**, 265-276.

Miya M, Takeshima H, Endo H *et al.* (2003) Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, **26**, 121-138.

Moore GL, Maranas CD (2000) Modeling DNA mutation and recombination for directed evolution experiments. *Journal of Theoretical Biology*, **205**, 483-503.

Morin PA, Chambers KE, Boesch C, Vigilant L (2001) Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Molecular Ecology*, **10**, 1835-1844.

Mullis K, Faloona F, Scharf S, *et al.* (1986) Specific Enzymatic Amplification of DNA Invitro - the Polymerase Chain-Reaction. *Cold Spring Harbor Symposia on Quantitative Biology*, **51**, 263-273.

Mullis KB, Faloona FA (1987) Specific Synthesis of DNA Invitro Via a Polymerase-Catalyzed Chain-Reaction. *Methods in Enzymology*, **155**, 335-350.

Murphy MA, Waits LP, Kendall KC (2003) The influence of diet on faecal DNA amplification and sex identification in brown bears (*Ursus arctos*). *Molecular Ecology*, **12**, 2261-2265.

Muyzer G (1999) DGGE/TGGE a method for identifying genes from natural ecosystems. *Current Opinion in Microbiology*, **2**, 317-322.

Myers RM, Maniatis T, Lerman LS (1987) Detection and localization of single base changes by denaturing gradiant gel electrophoresis. In: *Methods in Enzymology* (eds. Colowick SP, Kaplan NO), **155**, 501-527 Academic Press, New York.

Myers RA, Worm B (2005) Extinction, survival or recovery of large predatory fishes. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **360**, 13-20.

Noonan JP, Hofreiter M, Smith D, *et al.* (2005) Genomic sequencing of Pleistocene cave bears. *Science*, **309**, 597-600.

Norman MD, Lu CC (1997) Sex in the giant squid. *Nature*, **389**, 683-684.

Olesiuk PF (1993) Annual prey consumption by harbor seals (*Phoca vitulina*) in the Strait of Georgia, British Columbia. *Fishery Bulletin*, **91**, 491-515.

Orphan VJ, Taylor LT, Hafenbradl D, Delong EF (2000) Culture-dependent and culture-independent characterization of microbial assemblages associated with high-temperature petroleum reservoirs. *Applied and Environmental Microbiology*, **66**, 700-711.

Orr AJ, Banks AS, Mellman S, *et al.* (2004) Examination of the foraging habits of Pacific harbor seal (*Phoca vitulina richards*i) to describe their use of the Umpqua River, Oregon, and their predation on salmonids. *Fishery Bulletin*, **102**, 108-117.

Pääbo S (1989) Ancient DNA - Extraction, Characterization, Molecular-Cloning, and Enzymatic Amplification. *Proceedings of the National Academy of Sciences of the USA*, **86**, 1939-1943.

Pääbo S (1990) Amplifying ancient DNA. In: *PCR-Protocols and Applications - A Laboratory Manual* (ed. Innis MA, Gelfand, D.H., Sninsky, J.J., and White, T.J.), pp. 159-166. Academic Press, San Diego.

Pääbo S, Poinar H, Serre D, *et al.* (2004) Genetic analyses from ancient DNA. *Annual Review of Genetics*, **38**, 645-679.

Palmer C, Bik EM, Eisen MB, *et al.* (2006) Rapid quantitative profiling of complex microbial populations. *Nucleic Acids Research*, **34**, e5.

Palumbi SR (1996) Nucleic acids II: The polymerase chain reaction. In: *Molecular systematics* (eds. Hillis DM, Moritz C, Mable BK), pp 205-247. Sinauer Associates, Sunderland.

Parsons KM, Piertney SB, Middlemas SJ, Hammond PS, Armstrong JD (2005) DNA-based identification of salmonid prey species in seal faeces. *Journal of Zoology*, **266**, 275-281.

Peccoud J, Jacob C (1996) Theoretical uncertainty of measurements using quantitative polymerase chain reaction. *Biophysical Journal*, **71**, 101-108.

Phillips DL, Gregg JW (2001) Uncertainty in source partitioning using stable isotopes. *Oecologia*, **127**, 171-179.

Phillips DL, Newsome SD, Gregg JW (2005) Combining sources in stable isotope mixing models: alternative methods. *Oecologia*, **144**, 520-527.

Phillips KL, Jackson GD, Nichols PD (2001) Predation on myctophids by the squid Moroteuthis ingens around Macquarie and Heard Islands: Stomach contents and fatty acid analyses. *Marine Ecology Progress Series*, **215**, 179-189.

Phillips KL, Nichols PD, Jackson GD (2003) Size-related dietary changes observed in the squid *Moroteuthis ingens* at the Falkland Islands: stomach contents and fatty-acid analyses. *Polar Biology*, **26**, 474-485.

Pimm SL (2002) *Food Webs.* University of Chicago Press, Chicago.

Pitcher KW (1980) Food of the harbour seal, *Phoca vitulina richardsi*, in the Gulf of Alaska. *Fishery Bulletin*, **78**, 544-549.

Poinar H, Kuch M, McDonald G, Martin P, Paabo S (2003) Nuclear gene sequences from a late Pleistocene sloth coprolite. *Current Biology*, **12**, 1150-1152.

Poinar HN, Hofreiter M, Spaulding WG, *et al.* (1998) Molecular coproscopy: Dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science*, **281**, 402-406.

Poinar HN, Kuch M, Sobolik KD, *et al.* (2001) A molecular analysis of dietary diversity for three archaic Native Americans. *Proceedings of the National Academy of Sciences of the USA*, **98**, 4317-4322.

Purcell M, Mackey G, LaHood E, *et al.* (2000) Genetic identification of salmonid bone from harbor seal scat. In: *Marine Mammal Protection Act and Endangered Species Act implementation program 1999* (eds. Lopez A, DeMaster D), pp. 129-142. U.S. Department of Commerce, Seattle, WA.

Purcell M, Mackey G, LaHood E, Park L, Huber H (2004) Molecular methods for the genetic identification of salmonid prey from Pacific harbor seal (*Phoca vitulina richardsi*) scat. *Fishery Bulletin*, **102**, 213-220.

Pusch CM, Broghammer M, Nicholson GJ, *et al.* (2004) PCR-induced sequence alterations hamper the typing of prehistoric bone samples for diagnostic achondroplasia mutations. *Molecular Biology and Evolution*, **21**, 2005-2011.

Quetglas A, Alemany F, Carbonell A, Merella P, and Sanchez P, 1999. Diet of the European flying squid *Todarodes sagittatus* (Cephalopoda: Ommastrephidae) in the Balearic Sea (western Mediterranean). *Journal of the Marine Biological Association of the United Kingdom*, **79**, 479-486.

Raclot T, Groscolas R, Cherel Y (1998) Fatty acid evidence for the importance of myctophid fishes in the diet of king penguins, *Aptenodytes patagonicus*. *Marine Biology*, **132**, 523-533.

R Development Core Team (2004) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org

Reed JZ, Tollit DJ, Thompson PM, Amos W (1997) Molecular scatology: the use of molecular genetic analysis to assign species, sex and individual identity to seal faeces. *Molecular Ecology*, **6**, 225-234.

Reid K, Croxall JP, Briggs DR, Murphy EJ (2005) Antarctic ecosystem monitoring: quantifying the response of ecosystem indicators to variability in Antarctic krill. *ICES Journal of Marine Science*, **62**, 366-373.

Rollo F, Ubaldi M Ermini L, Marota I (2002) Ötzi's last meals: DNA analysis of the intestinal content of the Neolithic glacier mummy from the Alps. *Proceedings of the National Academy of Sciences of the USA*, **99**, 12594-12599.

Rosel PE, Kocher TD (2002) DNA-based identification of larval cod in stomach contents of predatory fishes. *Journal of Experimental Marine Biology and Ecology*, **267**, 75-88.

Ross HA, Lento GM, Dalebout ML, Goode M, Ewing G, McLaren P, Rodrigo AG, Lavery S, Baker CS (2003) DNA Surveillance: Web-based molecular identification of whales, dolphins, and porpoises. *Journal of Heredity*, **94**, 111-114.

Rozen S, Skaletsky HJ (2000) *Primer3 on the WWW for general users and for biologist programmers.* Humana Press, Totowa, NJ.

Saitoh K, Takagaki M, Yamashita Y (2003) Detection of Japanese flounder-specific DNA from gut contents of potential predators in the field. *Fisheries Science*, **69**, 473-477.

Saitou N, Nei M, (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406-425.

Sale PF, Cowen RK, Danilowicz BS, *et al.* (2005) Critical science gaps impede use of no-take fishery reserves. *Trends in Ecology & Evolution*, **20**, 74-80.

Santos RA, Haimovici M (1997) Food and feeding of the short-filmed squid *Illex argentinus* (Cephalopoda: Ommastrephidae) off southern Brazil. *Fisheries Research*, **33**, 139-147.

Scribner KT, Bowman TD (1998) Microsatellites identify depredated waterfowl remains from glaucous gull stomachs. *Molecular Ecology*, **7**, 1401-1405.

Sheffield VC, Cox DR, Lerman LS, Myers RM (1989) Attachment of a 40-base-pair G+C-rich sequence (G-C clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proceedings of the National Academy of Sciences of the USA*, **86**, 232-236.

Sheppard SK, Henneman ML, Memmott J, Symondson WOC (2004) Infiltration by alien predators into invertebrate food webs in Hawaii: a molecular approach. *Molecular Ecology*, **13**, 2077-2088.

Sheppard SK, Bell J, Sunderland KD, *et al.* (2005) Detection of secondary predation by PCR analyses of the gut contents of invertebrate generalist predators. *Molecular Ecology*, **14**, 4461-4468.

Sheppard SK, Harwood JD (2005) Advances in molecular ecology: tracking trophic links through predator-prey food-webs. *Functional Ecology*, **19**, 751-762.

Sinclair EH, Zeppelin TK (2002) Seasonal and spatial differences in diet in the western stock of Steller sea lions (*Eumetopias jubatus*). *Journal of Mammalogy*, **83**, 973-990.

Smith PJ, McVeagh SM, Allain V, Sanchez C (2005) DNA identification of gut contents of large pelagic fishes. *Journal of Fish Biology*, **67**, 1178-1183.

Soto K, Trites AW, Arias-Schreiber M (2006) Changes in diet and maternal attendance of a South American sea lions indicate changes in the marine environment and the abundance of prey. *Marine Ecology-Progress Series*, **312**, 277-290.

Springer AM, Estes JA, van Vliet GB, *et al.* (2003) Sequential megafaunal collapse in the North Pacific Ocean: An ongoing legacy of industrial whaling? *Proceedings of the National Academy of Sciences of the USA*, **100**, 12223-12228.

Stothard P (2000) The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques*, **28**, 1102-1104.

Sutherland RM (2000) Molecular analysis of avian diet. *PhD Thesis*, University of Oxford.

Suzuki MT, Giovannoni SJ (1996) Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Applied and Environmental Microbiology*, **62**, 625-630.

Symondson WOC (2002) Molecular identification of prey in predator diets. *Molecular Ecology*, **11**, 627-641.

Taberlet P, Griffin S, Goossens B, *et al.* (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, **24**, 3189-3194.

Taberlet P, Waits LP (1998) Non-invasive genetic sampling. *Trends in Ecology & Evolution*, **13**, 26-27.

Taberlet P, Waits LP, Luikart G (1999) Noninvasive genetic sampling: look before you leap. *Trends in Ecology & Evolution*, **14**, 323-327.

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. *Trends in Ecology and Evolution*, **18**, 70-74.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **25**, 4876-4882.

Tollit DJ, Thompson PM (1996) Seasonal and between-year variations in the diet of harbour seals in the Moray Firth, Scotland. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, **74**, 1110-1121.

Tollit DJ, Steward MJ, Thompson PM, *et al.* (1997) Species and size differences in the digestion of otoliths and beaks: Implications for estimates of pinniped diet composition. *Canadian Journal of Fisheries and Aquatic Sciences*, **54**, 105-119.

Tollit DJ, Wong M, Winship AJ, Rosen DAS, Trites AW (2003) Quantifying errors associated with using prey skeletal structures from fecal samples to determine the diet of Steller's sea lion (*Eumetopias jubatus*). *Marine Mammal Science*, **19**, 724-744.

Tollit DJ, Heaslip S, Deagle BE, *et al.* (2006) Estimating Diet Composition in Sea Lions: What Technique to Choose? In: *Sea lions of the world* (eds. Trites AW, Atkinson S, DeMaster D, *et al.*). Alaska Sea Grant College Program, University of Alaska Fairbanks.

Trebilco R (2004) Macaroni penguin (*Eudyptes chrysolophus*) foraging ecology at Heard Island: behaviour and environmental influences. *Honours Thesis*, University of Tasmania.

Trites AW, Larkin PA (1996) Changes in the abundance of Steller sea lions (*Eumetopias jubatus*) in Alaska from 1956 to 1992: how many were there? *Aquatic Mammals*, **22**, 153–166.

Trites AW (2003) Food webs in the ocean: who eats whom and how much? In: *Responsible Fisheries in the Marine Ecosystem* (eds. Sinclair M, Valdimarsson G) FAO, Rome.

van den Hoff J, Burton H, Davies R (2003) Diet of male southern elephant seals (*Mirounga leonina* L.) hauled out at Vincennes Bay, East Antarctica. *Polar Biology*, **26**, 27-31.

Vences M, Thomas M, Van der Meijden A, Chiari Y, Vieites DR (2004) Performance of 16S RRNA in DNA barcoding of amphibians. *Integrative and Comparative Biology*, **44**, 657-657.

Vestheim H, Edvardsen B, Kaartvedt S (2005) Assessing feeding of a carnivorous copepod using species-specific PCR. *Marine Biology*, **147**, 381-385.

von Wintzingerode F, Gobel UB, Stackebrandt E (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviews*, **21**, 213-229.

Walter CB, Oneill E, Kirby R (1986) Elisa as an Aid in the Identification of Fish and Molluscan Prey of Birds in Marine Ecosystems. *Journal of Experimental Marine Biology and Ecology*, **96**, 97-102.

Wandeler P, Smith S, Morin PA, Pettifor RA, Funk SM (2003) Patterns of nuclear DNA degeneration over time - a case study in historic teeth samples. *Molecular Ecology*, **12**, 1087-1093.

Ward DM, Weller R, Bateson MM (1990) 16s Ribosomal-RNA Sequences Reveal Numerous Uncultured Microorganisms in a Natural Community. *Nature*, **345**, 63-65.

Wasser SK, Houston CS, Koehler GM, Cadd GG, Fain SR (1997) Techniques for application of faecal DNA methods to field studies of Ursids. *Molecular Ecology*, **6**, 1091-1097.

Wehausen JD, Ramey RR, Epps CW (2004) Experiments in DNA extraction and PCR amplification from bighorn sheep feces: the importance of DNA extraction method. *Journal of Heredity*, **95**, 503-509.

Williams R, McEldowney A (1990) A guide to the fish otoliths from waters off the Australian Antarctic Territory, Heard and Macquarie Islands. *ANARE Research Notes*, **75**, 1-173.

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

Wilson RP (1984) An improved stomach pump for penguins and other seabirds. *Journal of Field Ornithology*, **55**, 109 -112.

Winship AJ, Trites AW (2003) Prey consumption of Steller sea lions (*Eumetopias jubatus*) off Alaska: How much prey do they require? *Fishery Bulletin*, **101**, 147-167.

Wittwer CT, Herrmann MG, Moss AA, Rasmussen RP (1997) Continuous fluorescence monitoring of rapid cycle DNA amplification. *Biotechniques*, **22**, 130-138.

Wuyts J, De Rijk P, Van de Peer Y, Winkelmans T, De Wachter R (2001) The European Large Subunit Ribosomal RNA Database. *Nucleic Acids Research*, **29**, 175-177.

Zaidi RH, Jaal Z, Hawkes NJ, Hemingway J, Symondson WOC (1999) Can multiple-copy sequences of prey DNA be detected amongst the gut contents of invertebrate predators? *Molecular Ecology*, **8**, 2081-2087.

Zimmermann K, Mannhalter JW (1998) Comparable sensitivity and specificity of nested PCR and single-stage PCR using a thermally activated DNA polymerase. *BioTechniques*, **24**, 222-224.

Appendix I: Investigation into the advantages of nested PCR and comments on its use in diet studies

Nested PCR has been used extensively for detection of specific DNA sequences in environmental microbiology and PCR diagnostics (Boon *et al*. 2002, Kanbe *et al*. 2002, Deagle *et al*. 2003). This procedure is generally considered to allow detection at a lower template concentration compared with standard PCR. A drawback of the approach is that it requires samples to be transferred to a new tube part way through the amplification, significantly increasing the chance of contamination. Recent studies have indicated the detection sensitivity of single round PCR can be comparable to results achieved with nested PCR (Zimmermann and Mannhalter 1998) and have also suggested methods to run single-tube nested PCR in order to minimize the risk of cross-contamination during transfer of PCR products (Abath *et al*. 2002 and references within). These studies discount a potentially valuable advantage of the traditional two-tube nested PCR method – the dilution of PCR inhibitors between the first and second round of amplification. The inhibition of PCR amplification can be caused by numerous compounds which act through diverse mechanisms, and dilution of many inhibitors can alleviate their negative effect (Wilson 1997). Inhibitors are most likely to be a problem when DNA to be tested is extracted from crude samples (e.g. from faecal and environmental sources) – when the composition of samples is variable, DNA concentration is low and identity of inhibitors is uncertain. Many DNA extraction methods and PCR facilitators have been suggested to remove or counteract inhibitors; however, protocol optimization can be difficult and time-consuming, making a robust PCR amplification invaluable. Here, I evaluate the performance of standard and two-tube nested PCR in the presence of the PCR inhibitory chemicals and briefly discuss the use of nested PCR in DNA-based diet studies.

As a model system, genomic DNA from the Australian fish *Arripis georgianus* was used as a target and a region of the mitochondrial 16S rRNA gene was amplified using AmpliTaq Gold polymerase. Primers developed to detect DNA from this fish species in seal faeces (Casper *et al*. unpublished data) and conserved animal primers were used in nested PCR (Table A1.1). The inhibitory chemicals investigated were EDTA and NaCl. These chemicals have both been shown to be highly inhibitory to

standard PCR at a concentration of 1 mM for EDTA (Abu Al-Soud and Rådström 2001) and 40 mM for NaCl (Abu Al-Soud and Rådström 1998).

Table A1.1 *Sequences of primers used for nested PCR amplification.*

Primer Name	Sequence (5'-3')	Use in Present Study	Product size
16S1F	gacgagaagaccct	First round nested PCR	390 bp
16SbR	ccggtctgaactcagatcacgt		
16SAgeF	gagcttcagacactcgagca	Standard PCR and second	237 bp
16S2R	cgctgttatccctatggtaact	round nested PCR	

Total cellular DNA was extracted from frozen muscle tissue using the DNAzol® reagent method (Invitrogen). The concentration of purified DNA was determined using a PicoFluor™ fluorometer (Turner Designs). Standard PCR reactions were done in a 25µl volume containing 0.4 µM of each primer, 0.14 mM dNTPs, 2.0 mM $MgCl_2$, 1X PCR Gold buffer and 0.625 units AmpliTaq Gold® (Applied Biosystems). Thermal cycling conditions were as follows: 94°C for 10 min then 40 cycles (94°C, 30s / 60°C, 30s / 72°C, 45s) followed by 72°C for 2 min. For nested PCR, in the first stage the concentration of primer was 0.2 µM, the annealing temperature was 55°C and the number of cycles was 20; other conditions were the same as in standard PCR. In the second stage 1 µl of the first amplification was used as template in the standard PCR described above. Products (8µl) were separated on a 2.0% agarose gel. Gels were stained with ethidium bromide and visualized under UV light.

Two approaches were used to look at the performance of nested and standard PCR in the presence of the inhibitors. In the first approach a concentration of inhibitor >10 times previously reported to be inhibitory to standard PCR was added the DNA template and the inhibitors were serially diluted along with the template. The resulting concentrations of the inhibitors were: 10 mM, 1mM, 0.1 mM , 10 µM, 1 µM , 0.1 µM and 10 nM for EDTA; and 400 mM, 40 mM, 4 mM , 0.4mM, 40 µM, 4 µM and 0.4 µM for NaCl. The template genomic DNA concentrations were 0.1 ng, 10 pg, 1 pg,

0.1 pg, 10 fg, 1 fg and 0.1 fg. Additional PCR reactions were also run with these DNA concentrations in the absence of inhibitors to determine the lower limits of detection for standard and nested PCR. Example gels are shown in Figure A1.1. In the second approach the concentration of the inhibitors was kept constant (4 mM EDTA, 80 mM NaCl) and the concentration of template was varied (100 fg, 50 fg, 25 fg, 12.5 fg, 2.5 fg, 1.25 fg). Each experiment was done in duplicate with negative control reactions, results are summarised in Table A1.2.

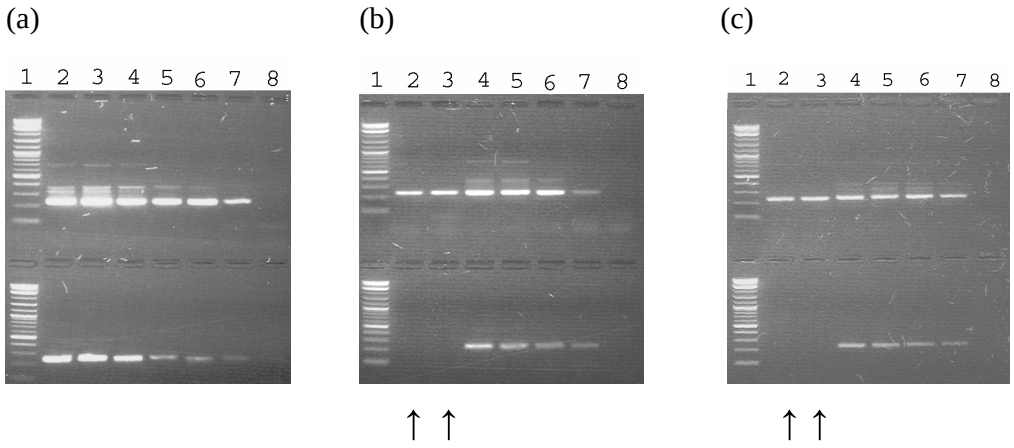


Figure A1.1 *PCR amplification products obtained by 10-fold serial dilution of template with **(a)** no PCR inhibitors **(b)** EDTA **(c)** NaCl, top half of each gel shows results from two-tube nested PCR and the bottom half of each gel the results from standard PCR. Arrows indicate amplification failure in standard but not nested PCR. For all gels **Lanes**: **1** 2-log ladder (New England Biolabs) **2-8** 0.1 ng, 10 pg, 1 pg, 0.1 pg, 10 fg, 1 fg and 0.1 fg of template DNA respectively. Concentration of inhibitors in **(b) 2-8** EDTA: 10 mM, 1mM, 0.1 mM , 10 μM, 1 μM , 0.1 μM and 10 nM **(c) 2-8** NaCl 400 mM, 40 mM, 4 mM , 0.4mM, 40 μM, 4 μM and 0.4 μM.*

Table A1.2 *Performance of nested and standard PCR under various conditions: (a) 10-fold serial dilution of DNA template and of PCR inhibitors (b) dilution of DNA with concentration of inhibitors kept constant.*[a]

(a)

PCR	DNA Template								PCR
Inhibitor	**0.1ng**	**10 pg**	**1 pg**	**100 fg**	**10 fg**	**1 fg**	**0.1 fg**	**0**	**method**
None	+,+	+,+	+,+	+,+	+,+	+,±	–,–	–,–	Nested
	+,+	+,+	+,+	+,+	+,±	±,–	–,–	–,–	Standard
EDTA	**10 mM**	**1 mM**	**10^{-1}mM**	**10^{-2}mM**	**10^{-3}mM**	**10^{-4}mM**	**10^{-5}mM**	**0**	
	+,+	+,+	+,+	+,+	+,+	±,+	–,–	–,–	Nested
	–,–	–,–	+,+	+,+	+,+	±,–	–,–	–,–	Standard
NaCl	**400 mM**	**40 mM**	**4 mM**	**0.4 mM**	**40μM**	**4μM**	**0.4 μM**	**0**	
	+,+	+,+	+,+	+,+	+,±	+,–	–,–	–,–	Nested
	–,–	–,±	+,+	+,+	+,–	±,–	–,–	–,–	Standard

(b)

PCR	DNA Template							PCR
Inhibitor	**100 fg**	**50 fg**	**25 fg**	**12.5 fg**	**2.5 fg**	**1.25 fg**	**0**	**method**
EDTA				**4 mM**				
	+,+	±,±	–,±	–,–	–,–	–,–	–,–	Nested
	–,–	–,–	–,–	NT	NT	NT	–,–	Standard
NaCl				**80 mM**				
	+,+	+,±	–,–	±,–	–,–	–,–	–,–	Nested
	–,–	–,–	–,–	NT	NT	NT	–,–	Standard

[a] Genomic DNA was from the fish *Arripis georgianus*, amount of DNA is per reaction tube. Symbols represent two independent PCR results: + strong band; ± weak band; - no band. NT: not tested.

The results are consistent with a previous study (Zimmermann and Mannhalter 1998) indicating that detection levels of standard and two-tube nested PCR are comparable when using pure templates and AmpliTaq gold polymerase (Table A1.2a). However, nested PCR did produce more reliable results at the lowest detectable level (1 fg), and bands at low DNA concentrations were generally brighter than the ones obtained in standard PCR. In the presence of high concentrations of

PCR inhibitors (10 mM and 1 mM EDTA; 400 mM NaCl) and ample DNA, the two-tube nested PCR gave positive results and the comparable conventional PCR produced negative results (Table A1.2). Dilution of template and inhibitors to lower levels resulted in positive results with both PCR approaches until the concentration of DNA became limiting.

The ability of nested PCR to work at high concentrations of inhibitors is most likely due to the dilution (1:25 in this case) which occurs during the procedure. In this scenario there is no amplification in the first round of nested PCR, but when the inhibitors are diluted during transfer to the second tube the reaction can proceed given sufficient DNA. An additional possibility is that some amplification occurs in the presence of inhibitors during the first round, allowing amplification in the second step even at low initial DNA levels. To investigate this, the performance of nested PCR was tested with a high level of inhibitors (4 mM EDTA and 80 mM NaCl) and low concentrations of template. In this case amplification failed at a DNA level approximately 25 times higher than in the absence of inhibitors (Table A1.2b). This result indicates that negligible amplification is occurring during the first round of nested PCR under these conditions; the lack of faint bands from first round PCR supports this conclusion (see Figure A1.1b, Top lanes 2 and 3). Under less inhibitory conditions it is probable some amplification would occur in the first round. There are additional mechanisms by which a nested PCR approach may reduce inhibition in PCR amplifications. One mechanism is specific to inhibitory chelating agents such as EDTA and some humic compounds commonly reported in environmental samples (Wilson 1997). These compounds are thought to inhibit PCR through binding Mg^{2+} (a necessary cofactor for DNA polymerase). Their effects can be counteracted through increased Mg^{2+} concentrations – this remedy is built into nested PCR since the amount of Mg^{2+} added is double that in standard PCR. PCR amplification can also be inhibited by excessive reconstructive polymerization of non-target DNA (Golenberg *et al*. 1996). In this case, dNTPs are used up during extension of the overlapping fragments, limiting amplification of the target at later cycles of PCR. In nested PCR, the new dNTPs added in the second round should be available exclusively for amplification of target sequences.

Nested PCR is currently applied most often where the level of target DNA is low and template DNA is from impure sources. Contrary to expectations based on

pure DNA, many of these types of DNA templates give strong bands using nested PCR and no band with standard PCR (Deagle *et al*. 2003). It seems likely that this observation and some claims of the exceptional sensitivity of nested PCR (10^4-10^5 > than standard PCR; Miserez *et al*. 1997) are due in part to dilution of inhibitors. When amplifying DNA from heterogeneous samples, the appropriate level of template dilution is often a trade-off between having enough DNA but not too many inhibitors. Since the level of target DNA is usually unknown and the type and amount of inhibitors present is also unknown, a major advantage of nested PCR is that it significantly increases the range in which a positive PCR result will be obtained compared with standard PCR. In light of this conclusion, the usefulness of single-tube nested PCR approaches (Abath *et al*. 2002) should be reconsidered – since with pure samples these methods are unlikely to be much more sensitive than standard PCR using AmpliTaq Gold polymerase, and for use with crude samples the dilution of inhibitors does not occur, negating a clear benefit of two-tube nested PCR.

The high sensitivity of two-tube nested PCR would suggest that it should be employed in DNA-based diet studies; however, the drawbacks of the approach must be considered. From a practical perspective, two-tube nested PCR requires double the amount of reagents (increasing costs considerably) and takes twice as long to run as a standard PCR. In addition, the chance of contamination occurring (i.e. obtaining false positive results) is increased – this drawback is balanced by the decreased possibility of obtaining false negative results. The trade-off is analogous in some ways to the subjective balance between Type I and Type II errors in statistical tests (Mapstone 1995). It is difficult to determine to determine the relative frequency of these PCR errors, but generally, the use of nested PCR is warranted in situations where obtaining a false positive is more acceptable than obtaining a false negative (e.g. when a PCR test is being used in an initial screen for harmful pathogens and positives results can be confirmed using another diagnostic test). When it is more acceptable to obtain a false negative than a false positive result, nested PCR should be avoided. In DNA-based diet studies, false negative results can be partially accounted for using internal positive controls (see discussion in Chapter 3); false positive results are more problematic and could have major implications on the outcome of a study. Therefore, nested PCR should be avoided in dietary studies whenever possible.

Appendix II: PCR primers

Table A2.1 *Details of all PCR primer pairs used in the thesis. The target taxa are species amplified in the various experiments; further assessment of specificity is recommended before using these primers in new studies.*

Forward Primer (5' → 3')[a]	**Reverse Primer** (5' → 3')	Target (gene: taxa)	Length (bp)	Reference
16sar-5' Fish (cctcgcctgtttaccaaaaaca)	**16Sbr-3'** (ccggtctgaactcagatcacgt)	16S mtDNA: Animal	~ 700	Palumbi (1996)[b] Chapter 2and 3
16S1F (ggacgagaagaccct)	**16S2R** (cgctgttatccctatggtaact)	16S mtDNA: Animal	~ 180 - 270	Chapter 2 and 3
ChordVf (acayaccgcccgtcac)	**ChordVr** (catratgcaaaaggta)	mtDNA: Chordata	~370	Jarman *et al.* (2004) Chapter 2
16fishF **(**agaccctatggagctttagac)	**16S2R Clamp**[c] (GC clamp + cgctgttatccctatggtaact)	16S mtDNA: Feeding trial fish	~ 260	Chapter 3
Squid28SF (cgccgaatcccgtcgcmagtaaamggcttc)	**Squid28SR** (ccaagcaacccgactctcggatcgaa)	28S nuclear: Cephalopoda	~ 280	Jarman (unpublished) Chapter 3 and 6
16fishF *KpnI* (ta***ggtacc***agaccctatggagctttagac)	**16S2R *EcoRI*** (ta***gaattc***cgctgttatccctatggtaact)	16S mtDNA: Feeding trial fish	~ 260	Chapter 4
16fishF *EcoRI* (ta***gaattc***agaccctatggagctttagac)	**16S2R *XmaI*** (ta***cccggg***cgctgttatccctatggtaact)	16S mtDNA: Feeding trial fish	~ 260	Chapter 4
16fishF *AgeI* (ta***accggt***agaccctatggagctttagac)	**16S2R *XhoI*** (ta***ctcgag***cgctgttatccctatggtaact)	16S mtDNA: Feeding trial fish	~ 260	Chapter 4
Sm.1F (acgtcaaacttccccttca)	**Sm.65 R** (ccaaccgaagacaggagaga)	16S mtDNA: Smelt	65	Chapter 4
Sal.1F (ggcagatcacgtcaaaaac)	**Sal.65R** (agacatatgggctaggggtc)	16S mtDNA: Salmon	65	Chapter 4
He.1F (accaatcacgaaaagcaggt)	**He.69R** (cgaagacgtttgtgccagta)	16S mtDNA: Herring	69	Chapter 4 and 5
	He.123R (tagggtagccccaatcctct)	16S mtDNA: Herring	123	Chapter 5
	He.184R (gcatgtagccggatcatttt)	16S mtDNA: Herring	184	Chapter 5
	He.226R (ggattgcgctgttatcccta)	16S mtDNA: Herring	226	Chapter 5
	He.304R (aatagcggctgcaccattag)	16S mtDNA: Herring	304	Chapter 5
SL.1F **(**caagtcaaccaaaacgggata)	**SL.61R** (caccccaacctaaattgctg)	16S mtDNA: Sea Lion	61	Chapter 5
	SL.91R (tcactcggaggttgttttgtt)	16S mtDNA: Sea Lion	91	Chapter 5
	SL.163R (cttgttccgttgatcaaagatt)	16S mtDNA: Sea Lion	163	Chapter 5
	SL.230R (tcgaggtcgtaaaccctgtt)	16S mtDNA: Sea Lion	230	Chapter 5
	SL.327R (gattgctccggtctgaactc)	16S mtDNA: Sea Lion	327	Chapter 5

TableA2.1 Continued…

16S1F (ggacgagaagaccct)	**16Sbr-3'** (ccggtctgaactcagatcacgt)	16S mtDNA: animal	~ 380	Chapter 5 and Appendix I
EuphMLSUF (tttattggggcgataaaaat)	**EuphMLSUR** (tcgaggtcgyaatctttcttgt)	16S mtDNA: Euphausiid	169	Chapter 6
KaMLSUF (cccacatcaaataccccta)	**KaMLSUR** (gggtcattggtggtcagaag)	16S mtDNA: *K. anderssoni*	169	Chapter 6
NotoMLSUf **(**ccctatgaagcttyagacrta)	**NotoMLSUr** (ccttgttgatawggtctctaaaa)	16S mtDNA: Notothenioidei	~ 275	Jarman *et al.* (2004) Chapter 6
AmphNSSF1 (ctgcggttaaaaggctcgtagttgaa)	**AmphNSSR1** (actgctttragcactctgatttac)	18S nuclear: Amphipoda	~ 200-375	Jarman *et al.* (2006) Chapter 6
16S1F-degenerate (gacgakaagacccta)	**16S2R-degenerate** (cgctgttatccctadrgtaact)	16S mtDNA : Animal, not birds[d]	~ 180 - 270	Chapter 6
16SAgeF (gagcttcagacactcgagca)	**16S2R** (cgctgttatccctatggtaact)	16S mtDNA: *A. georgianus*	237	Appendix I

[a] Redundancies use the code: r (ag); k (gt); d (agt); m(ac)

[b] original degenerate forward primer modified to match fish consensus

[c] unclamped primer used in first round of nested PCR for faecal DNA template. In second round amplification and clone library screening GC clamp (gggcgggggcggcgggacgggcgcggggcgcggcgggcg) added to primer for DGGE analysis

[d] more information on this primer set is provided in Table A2.2

Table A2.2 *Details of the degenerate primers used to create clone libraries from penguin faecal samples (Chapter 6). The taxa-specific primer binding sites and melting temperatures are shown. Size of amplification products are: Fish ~ 250 bp; Cephalopoda ~ 180 bp; and Euphausiid ~ 200 bp.*

Primer sequences (5' → 3')[a]	Matching taxa	Melting temperature[b]
16S1F-degenerate		
gacgakaagacccta	All	—
gacgagaagacccta	Fish	52.58
gacgagaagacccta	Cephalopoda	52.58
gacgataagacccta	Euphausiid	49.34
16S2R-degenerate		
cgctgttatccctadrgtaact	All	—
cgctgttatccctagggtaact	Fish	63.89
cgctgttatccctatggtaact	Cephalopoda	62.22
cgctgttatccctaaagtaact	Euphausiid	60.42

[a] Redundancies use the code: k (gt); r (ag); d (agt)

[b] Nearest neighbor Tm (degrees C) from Stothard (2000)

Appendix III: Real-time PCR copy number estimates used for quantification of DNA degradation

Table A3.1 *Estimated copy numbers of sea lion and herring DNA for various sizes of template. Copies per reaction for the ten sea lion faecal sample analysed in Chapter 5 are given (two replicate measurement and **mean values** are shown).*

(a) Sea lion copy number estimates:

	61 bp	**91 bp**	**163 bp**	**230 bp**	**327 bp**
1a	27067	21485	10146	1942	1109
1b	62387	23438	13784	5551	1665
	44727	**22461**	**11965**	**3746**	**1387**
2a	22490	21141	5900	969	556
2b	26204	10796	4885	2006	495
	24347	**15968**	**5393**	**1487**	**525**
3a	204777	205536	119205	33193	21066
3b	268873	181284	123862	56297	30807
	236825	**193410**	**121534**	**44745**	**25936**
4a	33024	37234	17736	6940	4303
4b	37667	24458	23030	12888	6476
	35346	**30846**	**20383**	**9914**	**5390**
5a	46202	31332	9606	2737	817
5b	41377	20882	14339	5728	1734
	43789	**26107**	**11972**	**4232**	**1276**
6a	183211	163853	61393	17999	6097
6b	175025	88659	53041	34644	11271
	179118	**126256**	**57217**	**26322**	**8684**
7a	159720	129525	57253	13268	6173
7b	174356	117391	68198	20884	9714
	167038	**123458**	**62726**	**17076**	**7944**
8a	32957	25771	14705	8408	1904
8b	26409	24241	15178	9606	3733
	29683	**25006**	**14941**	**9007**	**2818**
9a	218422	271214	173467	40689	18302
9b	179218	174226	119951	66527	21216
	198820	**222720**	**146709**	**53608**	**19759**
10a	52899	47174	33760	24900	6624
10b	64107	45893	36198	26577	12624
	58503	**46534**	**34979**	**25738**	**9624**

(a) Herring copy number estimates:

	69 bp	**123 bp**	**184 bp**	**226 bp**	**304 bp**
1a	8432	771	186	119	71
1b	7883	273	99	68	13
	8157	**522**	**143**	**93**	**42**
2a	7096	1429	372	394	208
2b	12757	560	141	85	32
	9927	**994**	**257**	**240**	**120**
3a	11925	2658	604	395	207
3b	16085	1373	769	367	148
	14005	**2016**	**687**	**381**	**178**
4a	1485	544	228	135	79
4b	1664	292	263	229	59
	1575	**418**	**246**	**182**	**69**
5a	11980	1414	318	140	58
5b	11317	491	167	157	35
	11649	**952**	**243**	**149**	**46**
6a	9041	2091	544	234	145
6b	14466	1309	667	380	95
	11754	**1700**	**605**	**307**	**120**
7a	11817	5063	1629	1656	357
7b	25360	2970	914	674	111
	18588	**4016**	**1272**	**1165**	**234**
8a	33007	10541	2178	1570	504
8b	17400	7225	2434	1163	355
	25203	**8883**	**2306**	**1366**	**429**
9a	25115	6698	1221	372	320
9b	27475	4295	1377	938	212
	26295	**5497**	**1299**	**655**	**266**
10a	736	400	161	122	84
10b	685	464	315	155	130
	711	**432**	**238**	**138**	**107**

Appendix IV: Penguin clone library results

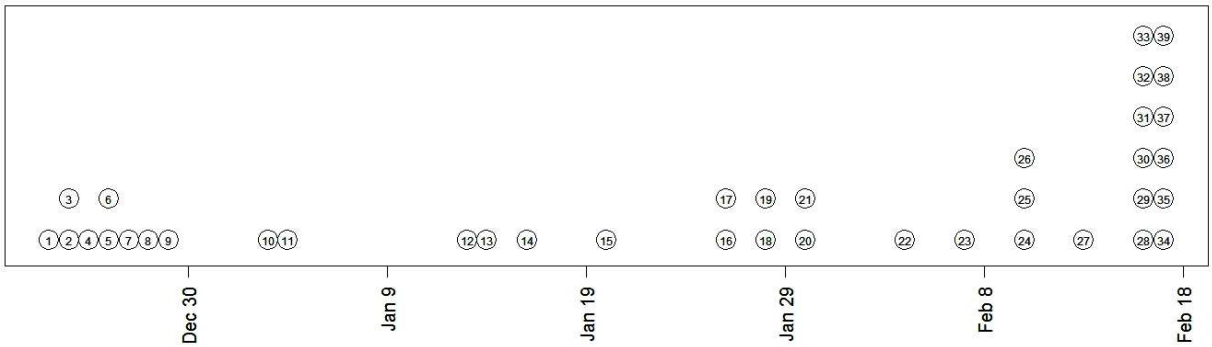


Figure A4.1 *Sample numbers for the 39 penguin faecal samples containing prey DNA. Numbers correspond to clone library results in Table A4.1 (Sample #) and presence/absence results shown in Figure 6.2.*

Table A4.1 *Results from the penguin faecal DNA clone library analysis (Chapter 6). Samples given in bold were analysed with both primers sets.*

(**A**) Conserved primers (degenerate 16S1F and 16S2R)

Clone library	Sample #	Date	# Clones - Species ID[a]
A1	**4**	Dec 25	6 - *K. anderssoni*
A2	**5**	Dec 26	6 - Acanthopterygii
A3	**12**	Jan 13	6 - *K. anderssoni*
A4	**14**	Jan 16	6 - *K. anderssoni*
A5	18	Jan 28	6 - *K. anderssoni*
A6	23	Feb 7	3 - *K. anderssoni*
			1 - *Harpagifer* sp.
			2 - *C. gunnari*
A7	24	Feb 10	6 - *K. anderssoni*
A8	27	Feb 13	5 - *K. anderssoni*
			1 - *G. antarcticus*
A9	31	Feb 15	4 - *K. anderssoni*
			2 - *E. antarctica*
A10	34	Feb 16	1 - *K. anderssoni*
			2 - *C. gunnari*
			3 - Nototheniinae sp.

(**B**) Euphausiid primer pair

Clone library	Sample #	Date	*Euphausia/ Thysanoessa*[b]
B1	3	Dec 24	6/4
B2	**4**	Dec 25	2/8
B3	**5**	Dec 26	10/0
B4	7	Dec 27	2/8
B5	8	Dec 28	9/1
B6	10	Jan 3	1/9
B7	**12**	Jan 13	0/10
B8	13	Jan 14	0/10
B9	**14**	Jan 16	0/10
B10	16	Jan 26	0/10

[a] Taxonomic identity of the six clones sequenced from each library. Further details are in Table 6.5.

[b] Identity of euphausiid sequences in ten clones sequenced from each library.