A genetics-based approach for the determination of *Phalacrocorax auritus* diet on Lake Nipissing

By

Mathieu Lecours

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science (MSc) in Chemistry

The Faculty of Graduate Studies
Laurentian University
Sudbury, Ontario, Canada

THESIS DEFENCE COMMITTEE/COMITÉ DE SOUTENANCE DE THÈSE
Laurentian Université/Université Laurentienne
Faculty of Graduate Studies/Faculté des études supérieures

Title of Thesis Titre de la thèse	A genetics-based approach for the determination of Phalacrocorax auritus diet on Lake Nipissing
Name of Candidate Nom du candidat	Lecours, Mathieu
Degree Diplôme	Master of Science
Department/Program Département/Programme	Chemical Sciences
Date of Defence Date de la soutenance	April 28, 2017

APPROVED/APPROUVÉ

Thesis Examiners/Examinateurs de thèse:

Dr. Thomas Merritt
(Supervisor/Directeur(trice) de thèse)

Dr. George Morgan
(Committee member/Membre du comité)

Dr. Tom Johnston
(Committee member/Membre du comité)

Dr. Charles Ramcharan
(Committee member/Membre du comité)

Dr. Katharine Coykendall
(External Examiner/Examinateur externe)

Approved for the Faculty of Graduate Studies
Approuvé pour la Faculté des études supérieures
Dr. David Lesbarrères
Monsieur David Lesbarrères
Dean, Faculty of Graduate Studies
Doyen, Faculté des études supérieures

ACCESSIBILITY CLAUSE AND PERMISSION TO USE

Abstract

The decline in *Sander vitreus* population on Lake Nipissing (LN) has become an increasing cause for concern for the city of North Bay and the surrounding communities. Double-crested cormorants (*Phalacrocorax auritus*) have been suggested as a possible cause for the decline of these fish stocks. Using a combination of molecular and bioinformatic techniques, I have developed a non-invasive genetics-based method for the determination of Double-crested Cormorant (DCCO) diet using the 16S ribosomal RNA gene. I observed general trends that were consistent over the overall two year-long study. I observed small differences in diet dependant on the location of the island being studied, as well as time of year. Overall, I was able to determine that DCCO consisted of the following relative abundance of prey: 43% yellow perch (*Perca flavescens*), 27% trout-perch (*Percopsis omniscomaycus*), 12% pumpkinseed (*Lepomis gibbosus*), 7% walleye (*Sander vitreus*), 4% rainbow smelt (*Osmerus mordax*), 4% unresolved, and 3% other species. The data collected from this thesis can be applied to statistical modelling analysis to possibly determine the impact DCCO have on the overall population of walleye, and the methods could be applied for long-term monitoring of overall diet.

Keywords

Genetics-based diet analysis, *Phacrocorax auritus*, Lake Nipissing, *Sander vitreus*, 16S rRNA, DNA-barcoding

Co-Authorship Statement

Chapter 1 is an introduction to the current state of the *Sander vitreus* population in LN and the possible role DCCO play. DNA-barcoding is introduced with emphasis on the advancements in next-generation sequencing and bioinformatics. Thomas J.S. Merritt (TJSM) provided critical editorial feedback as well as input on the organization of ideas.

Chapter 2 is written in the form of a manuscript for submission to the journal *Molecular Ecology*. I am the first author with co-authors Deborah Iwanowicz (DI), Gustavo Ybazeta (YB), and TJSM. DI was responsible for the next-generation sequencing, GY was responsible for the bioinformatic analysis of the data, and TJSM contributed to the overall conception of the study, as well as providing troubleshooting advice for assays.

Chapter 3 is a summary of a report presented to the Ministry of Natural Resources and Forestry. This chapter discusses all trends observed from the analyzed data. George Morgan (GM) contributed to the organization of the data. TJSM provided valuable editorial feedback and concept organization.

Chapter 4 provides general conclusions obtained from the data analysis. Suggestions for future analysis and applications of the methods are also presented.

Acknowledgements

I would like to acknowledge TJSM for his contributions to the experimental design of this project, as well as giving me the opportunity to work in the lab. I would like to acknowledge GY for his support and contribution of bioinformatic analysis and the members of the Merritt lab for their constant support over the course of this Masters degree.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

18S : 18S ribosomal RNA

16S : 16S ribosomal RNA

CO1 : Cytochrome oxidase 1

DCCO : Double-crested Cormorant

GO: Goose Island

GU: Gull Island

HW: Hardwood Islands

LN : Lake Nipissing

NGS : Next Generation Sequencing

PC : Pre-Chick

PH: Post-Hatch

WW: Wigwam Island

MNRF: Ministry of Natural Resources and Forestry

SAR: Species at Risk

Chapter 1 Lake Nipissing, Double-crested Cormorants (*Phalacrocorax auritus*), and genetics-based diet analysis

1.1 Project Outline

The decreasing stock of walleye (*Sander vitreus*) in Lake Nipissing (LN) has become an increasing cause for concern for many communities and fishing establishments on and around the lake. The correlation between the decreasing fish stocks and the increasing nesting population of Double-crested Cormorants (*Phalacrocorax auritus*) has led some to propose the bird as a possible cause for the decline in fish stocks and to further suggest that the birds be culled to promote fish stock recovery (Campbell 2014). Before drastic measures are taken, it is important to accurately infer the diet of these migratory birds. An understanding of their diet should give insight into the impact they have on the LN fish stocks. Many techniques for diet analysis have been developed, including stable-isotope analysis, otolith identification, DNA analysis, and quantitative fatty acid signature analysis (reviewed in Barrett et al 2007). Over the last decade, DNA-based diet analysis has become an increasingly popular technique, driven, in part, by advancements in DNA sequencing technologies (Deagle et al. 2007, Tollit et al. 2009, De Barba et al. 2014,). These methods can be applied to a wide variety of environmental samples including faeces, stomach content, and cough pellets; this thesis uses the last. In DCCO, cough pellets are produced roughly every 24H and are made up of undigested fish parts (Duffy and Laurensen 1983). This thesis focuses on the development of a molecular genetics-based method using DCCO cough pellets to infer their diet and the use of this method to quantify the relative abundance of various fish species in DCCO diet in LN in

2014 and 2015. Future research will be able to use this diet information to infer the possible effects of DCCO on fish populations in the lake.

1.2 Lake Nipissing and its decreasing Walleye population

Lake Nipissing is the fourth largest non- Great Lake in Ontario, and is home to 44 different species of fish (Appendix 1). The lake is located in Northern Ontario, in the Nipissing District, and various Townships and Reservations can be found along its shoreline. Understandably, the lake is an important resource for the surrounding communities, which rely heavily on the fish stocks of the lake as both a food source and a local economic resource through commercial and recreational fishing. Although there are many different species of fish that are harvested from the lake walleye (*Sander vitreus*) is the most targeted of all.

Walleye population in LN have been closely monitored by the Ministry of Natural Resources and Forestry (MNRF) for the past few decades (Ontario Ministry of Natural Resources and Forestry; 2013). This survey data suggests that over the last 20 years the walleye population has been on a steady decline, causing the MNRF to implement changes in fishing regulations. These regulations include direct measures to decrease Walleye harvesting, such as increasing the catch size for walleye, and also indirect measures, such as increasing the limit of yellow perch (*Perca flavescens*), and prolonging the Bass fishing season. Both measures are meant to deflect fishing pressure away from walleye to other species.

The exact cause of decline in walleye stocks in LN is hotly debated and has yet to be determined (Campbell 2014). The possible causes range from sport, commercial, and subsistence fisheries, to pollution, disease, and DCCO populations. To properly determine the root of the problem, it is important to objectively investigate each possible cause. DCCO populations have been on a steady incline since the populations first appeared on LN in 1990 (**Figure 1.1**), and have become an easy scapegoat when discussing the cause for walleye decrease (Campbell 2014). In order to objectively determine the exact impact these migratory birds may be having on the walleye population, it is important to accurately determine their diet.

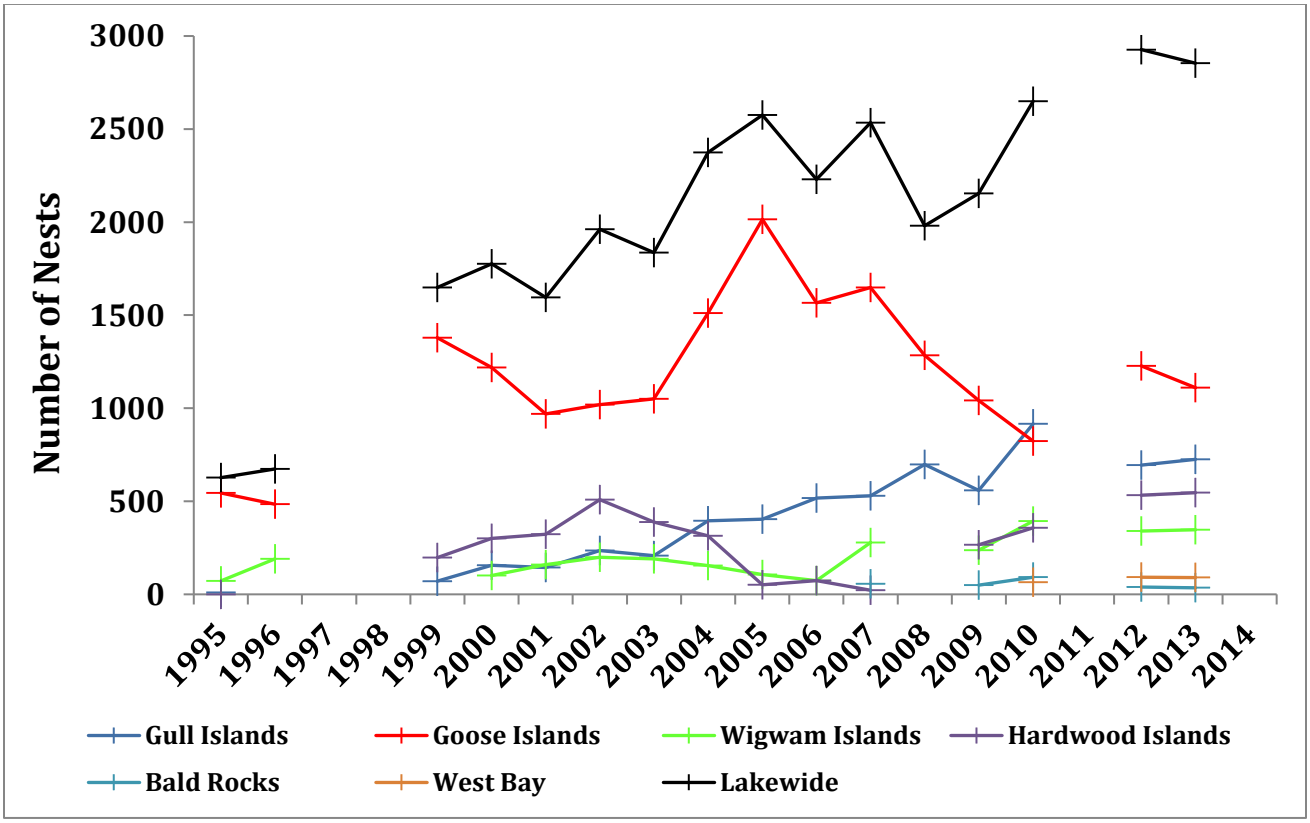


Figure 1.1 Long-term monitoring data of DCCO nest counts on LN (data provided by the Ministry of Natural Resources and Forestry`)

1.3 Cormorants and diet analysis

DCCO are migratory birds that are widely distributed across North America. DCCO are relatively large, with an average adult body length between 70 and 90 cm, and an adult wingspan between 114 and 123 cm. These birds migrate to various islands on LN during the breeding period (usually end of May), and leave the area to continue their migration once the chicks have matured (usually end of July). The birds' diet is primarily composed of fish, with small amounts of other animals including some crustaceans, and the birds' predation on small fish has been documented from testimonial accounts for hundreds of years (Hatch et al. 1999).

DCCO are opportunistic predators, usually preying on the most abundant fish in their environment, and studies indicate a relatively small slot size of prey averaging anywhere between 12-15 cm, however prey up to 35 cm in length have been observed (Craven and Lev 1987, Ross et al. 2005). DCCO are known to feed in relatively shallow water (<10 m deep) and relatively close to shore (<5 km away), by diving into the water and obtaining prey that are slowly moving or schooling fish (Hatch et al. 1999). Several other research groups have investigated DCCO diet using the regurgitated "cough pellets" produced by the DCCO, which contain undigested parts of their meal, and, as mentioned above, are produced roughly every 24 hours reflecting bird diet over that period of time (Duffy and Laurenson 1983). Prey items can sometimes be identified using specifies-specific bone morphology, such as the cleithra or otolith bones of fish (Traynor et al. 2001, Brodeur 1979). Although this bone-morphology-based method can give a general idea of diet, the

method is extremely time consuming and relies on an expert for identification of bones which can be hard to find and can be subjective (Candek and Kuntner 2015). Because of this limitation, it is important to develop additional methods of diet analysis, in order to reduce cost, time, and subjectivity.

A wide range of biochemical techniques have been used to determine diet from various environmental samples including gut contents, faeces, or pellets (Taguchi et al. 2014). The first of these often requires sacrificing the animal, but availability of cough pellets from DCCO allowed us to develop a non-invasive method for diet analysis. Other studies have shown the utility of genetics-based methods in determining diet and we were particularly interested in developing a similar method for DCCO diet analysis (Pompanon et al 2012 and Taguchi et al. 2014). In contrast to other biochemical methods, such as stable-isotope analysis, and quantitative fatty acid signature analysis, a DNA-based analysis does not require a tissue sample from the specimen (DCCO) being examined (Waldner and Traugott 2012). Although, to our knowledge, no DNA-based diet analysis has been attempted on DCCO cough pellets, the genetic techniques we used have been used to determine diet from other biological samples such as faeces and stomach contents of fish and birds (Jarman et al 2013, Carreon-Martinez et al. 2011).

1.4 DNA Barcoding

The use of set segments of DNA, termed DNA "barcodes", to identify broad suites of taxa was first popularized in the early 2000s (Hebert et al. 2003). The idea behind barcoding is to use the species-specific sequence of a common DNA fragment to identify

as many different species as possible and this method takes advantage of flanking conserved and variable sequences of the genome. Oligonucleotide primers are designed to bind to conserved regions of the gene that flank more variable regions. Pairs of these "Universal" primers bind a specific gene/gene region over a wide range of different species (Takahashi et al. 2014) and can be used to amplify and/or sequence the regions that they flank. Variable regions are selected to include appropriate levels of variability to allow determination of the taxonomic level of interest, often species. Comparing the variable regions obtained from various species allows one to taxonomically identify species in an un-biased, genetics-based, manner (Luo et al. 2011). For barcoding to be successful, researchers must identify regions that are conserved enough across their group of interest to allow universal binding of the primers that also flank regions with sufficient variability to uniquely identify taxa of interest.

DNA barcoding was initially limited in its applications due to limitations in DNA sequencing techniques. Before the introduction of Next-Generation Sequencing (NGS), Sanger sequencing was the gold standard for DNA sequencing. Sanger sequencing involves amplification and sequencing of a single sequence of DNA from a sample, often yielding 500 to 1500bp of high quality DNA sequence. The method is extremely accurate, but limited the capabilities of DNA barcoding by only being able to identify a single consensus sequence to any given sample of DNA. To identify multiple sequences from a single sample, DNA would have to be extracted, amplified, and individual sequences cloned making the resolution of the genetic mixture into single genotypes both costly and time consuming (Carreon-Martinez et al. 2011). Even in the face of this

limitation, however, the use of Sanger sequencing for DNA barcoding led to the creation of international projects dedicated to barcoding all animals (Savolainen et al. 2005). Advancements in DNA sequencing technologies has expanded the ability of DNA barcoding to allow for the identification of multiple sequences from a single environmental sample, which contains DNA from several different species (Taberlet et al. 2012). These types of sequencing technologies are known as NGS, and have significantly improved the efficiency and accuracy of barcoding.

NGS was first developed in the early 2000s as an improved method of sequencing. NGS, in contrast to Sanger sequencing, allows for hundreds of thousands of smaller segments of DNA (150 to 500 bp) to be amplified and sequenced simultaneously in parallel (Gharizadeh et al. 2006). One clear advantages of NGS (compared to traditional sequencing) is that it is possible to obtain sequence information from multiple species from a single environmental sample (Catala et al. 2015). This is contrary to traditional Sanger sequencing, where only a single consensus sequence can be obtained from any given environmental sample (Hajibabaei et al. 2011). The protocol allowing us to obtain these large sets of sequence data related to a large range of different species is known as DNA meta-barcoding.

DNA meta-barcoding can be applied to a wide range of applications, from the identification of cryptic species, to diet determination (Carew et al. 2013, Pompanon et al. 2012). This thesis is focused on using DNA meta-barcoding as a tool for diet determination of DCCO. In order to properly apply this technique, it is important to

identify a ubiquitous gene that can be amplified across a wide range of different species using a set of universal primers.

1.5 Gene Selection

There have been a number of different genes that have been proposed as possible DNA barcodes, with different genes suggested for different types of organisms. In microbial species, the 16S gene (coding for the 30S small ribosomal subunit) has been used to identify and classify bacteria (Fox et al. 1977). Many different genes have been proposed as barcodes for plants, including internal transcribed spacer 2 (ITS2), ribulose-1,5-bisphosphate carboxylase/oxygenase long chain coding gene (rbcL), and Maturase K (matK; Li et al. 2011, Yao et al. 2010). Due to the vast genetic diversity seen in plants, it has been proposed that a combination of these genes would produce the best results for genetic taxonomic identification of species (Vassou et al. 2015). For animal DNA identification, the Cytochrome Oxidase I gene (CO1) was first proposed by Heber et al, (2003) as a barcode for identifying all animal life. The use of this gene as a DNA barcode is the basis for a number of international databases such as the international Barcode of Life project (iBOL) and the Consortium for the Barcode of Life (CBOL) initiative. The CO1 gene is one of 37 genes found in the mitochondrial genome, which varies in size depending on the species, but typically the genome is around 16,000 bp (Boore 1999). A 648 bp fragment of the CO1 gene has been used as the barcode for hundreds of thousands of animal species. The use of mitochondrial DNA can have many advantages compared to the use of nuclear markers. Animal cells can contain hundreds of mitochondria per cell, each containing a single circular genome, however nuclear DNA usually only has

two copies per cell (Yang et al. 2014). When we are dealing with environmental samples, often times the DNA can be highly degraded, which makes isolation and amplification of DNA a challenge (Boyer et al. 2012). By targeting mitochondrial DNA, there is a much better chance of being able to obtain quality DNA that can lead to much more accurate identification of species. The CO1 barcode is extremely effective in its ability to accurately and consistently identify closely related animal species, however when it comes to meta-barcoding, the target sequence suffers from its lack of a highly conserved primer binding site (Deagle et al. 2014 and Luo et al 2011). It is difficult, or impossible, to design a single set of primers that will work in a wide variety of species, the gene sequence is simply too diverse, even in the "conserved" regions. With a set of "universal" primers that will work across a wide range of species, it is difficult to obtain accurate dietary information from complex samples containing multiple species (such as bird pellets). In fact, many dietary studies which have attempted to use CO1 in order to obtain dietary information from various environmental samples, (i.e. faeces, stomach content) have had to introduce various primer sets in order to amplify all possible prey species found in the sample (Lobo et al. 2013). The lack of conservation across a wide range of species in the animal kingdom has been brought forward in recent articles, thus leading to the investigation of other genes to be used as barcodes (Deagle et al. 2014).

Many studies have identified an alternative barcode to COI that can be used for DNA meta-barcoding species of the animal kingdom. The mitochondrial 16S ribosomal RNA gene (mt16SrDNA), is one of the genes that have been studied as a possible candidate for DNA barcoding (Cawthorn et al. 2012). The mt16SrDNA gene is a noncoding gene that

varies in size between 1,500-1,700 base pairs and is composed of both highly conserved and variable regions (Yang et al. 2014). Comparative analysis of the conservative regions of the CO1 and mt16SrRNA gene covering a wide range of species, demonstrated the mt16SrRNA can be a better candidate for DNA meta-barcoding (Deagle et al. 2014). The highly conserved primer binding sites increases the overall coverage of the experimental sample, increasing the accuracy of the meta-barcoding data. In contrast, the variable amplicon region allows for species level differentiation when comparing sequences from multiple species. The proposed barcode, is approximately 250 bp long and located at the 3' end of the gene, it also includes highly variable regions which make it possible to taxonomically identify closely related species (Yang et al. 2014). The ability to differentiate closely related species is extremely important in our case as many possible prey species, such as walleye and perch, are closely related. The inability to differentiate closely related species would not allow for proper identification of all DNA extracted from the environmental sample, leading to a misinterpretation of the data. The selection of the right gene, and gene region, plays a pivotal role in the optimization of DNA meta-barcoding and its application to diet determination. Following the selection of an appropriate marker gene and subsequent NGS of the environmental DNA, it is important to consider how the sequencing data is to be analyzed to allow for an unbiased and accurate representation of the results.

1.5 Bioinformatics

In order to process and analyze the large datasets produced by NGS, one must rely on bioinformatics analysis. To properly interpret the results of NGS, the data must go

through a bioinformatic pipeline of a set of data processing elements, which will allow one to properly interpret these large metabarcoding data sets (Fosso et al. 2015). Before the sequencing data can be run through any sort of analysis, the sequences must go through quality control in order to obtain a data set that can be properly interpreted. The first step in the bioinformatic pipeline is the trimming and filtering of short reads. Trimming of sequences refers to the removal of the ends sequence data, and was done in order to rid the reads of adapter sequences. Adapter sequences were fused to the original extracted DNA in order to identify the origin of the DNA. In addition to trimming the adapter sequences, trimming of reads that have obtained a low quality score was also done (Shrestha et al. 2014). The quality scores associated with most automated sequencing platforms are called Phred quality scores, or Q-scores, and relate to the probability of error in base-calling; low scores reflect poor-quality sequences that are not reliable (Ewing et al., 1998). After reads are trimmed and low scoring reads are discarded, the sequences can then be assembled and (depending on the sequencing platform being used) can paired-ends sequences were paired with a set number of minimum overlap. Due to the numerous programs available, it is important to consider the type of reads being analyzed (RNA versus DNA), as well the origin of the DNA (prokaryotic versus eukaryotic) .The final step in data clean-ups before it can be analyzed is the dereplication of the data. That is the removal of singletons (a read that is present only once), and the removal of chimeras using the UCHIME program (DNA reads originating from two different species; Fosso et al. 2015, Edgar et al. 2010). Once this final step has been achieved, the resulting file can then be subjected to further bioinformatics analyses.

One of the most widely used bioinformatics tools is the Basic Local Alignment Search Tool (BLAST). BLAST was developed as an algorithm that can be implemented in a number of different ways, including DNA and protein sequence database searches, as well as the functional analysis of newly sequenced proteins and DNA (Altschul et al. 1999). In the context of DNA meta-barcoding, BLASTs can be used as a tool to compare the large amounts of sequence data obtained from NGS to large databases of sequences (i.e. NCBI nucleotide database). This can lead to sequence identification, as well as listing closely related sequences based on overall alignment of nucleotides. The resulting output generated from performing a BLAST (BLAST file) can be imported and further analyzed by other software, that can help with the taxonomic identification of large sets of environmental data.

The resultant BLAST files obtained form a BLAST search obtained must be further analyzed in order to clearly and objectively interpret the results. Taxonomic identification is the end goal for many DNA meta-barcoding data sets, which is why there are many software programs that have been developed that are specifically designed for species level identification of sequences (eg. Taxonomic Unit Identification Tool (TUIT) and MEtaGenome Analyzer (MEGAN)). Taxonomic identification software is designed to assign classification to large sets of environmental data sets (Mitra et al. 2012). Once the data sets have been assigned a taxonomic classification at a set confidence interval, one can then pool the data to obtain semi-quantitative diet analysis related to the species from which the environmental sample was collected(Pompanon et al. 2012).

In this thesis I apply a combination of molecular genetic, and bioinformatic techniques, in an attempt to develop a genetics-based method for the determination of DCCO diet on Lake Nipissing.

Chapter 2 Genetics-based identification of DCCO diet from cough pellets

2.1 Introduction:

Drastic and dramatic changes in population sizes of endangered or commercially important species are driving increased interest in diet-focused ecological studies of these species, their predators, and their prey. Researchers and policy makers are trying to better understand food webs and the contribution of various species to a community (Chapin et al. 2010). To accomplish this, traditional diet analysis is done via morphological analyses (faeces, regurgitate, stomach lavage, etc.), observation of the feeding behavior, or stable isotope analysis (eg. Bowswer et al. 2013. Sydman et al. 1997. Camphuysen and Garthe 2004). Although these methods can be useful, they entail extensive expert training, hours of meticulous work, are subject to biases, and are often not able provide sufficient taxonomic resolution (Barrett et al. 2007). Due to these limitations, there is considerable interest in developing molecular (DNA-based) approaches that are accurate, less time consuming and cost-effective (Czernik et al. 2013 and Jarman et al. 2013).

Easy, accurate, and inexpensive DNA sequencing, and most recently advances in NGS, have led to the widespread use of molecular techniques for species identification often using common, shared, DNA sequences, called DNA barcoding. The idea of DNA barcoding was first put forth by Paul Hebert and his research group, who proposed a new system for species identification/discovery by using a short section of DNA from a

standardized region of the Genome (Hebert et al. 2003). Using these barcodes, multiple international projects have been dedicated to identifying species across the world via the DNA bacrcoding method (iBOL and CBOL). Concurrent with this broader trend, is an interest in applying these techniques to dietary identification from a suite of species within all classes of vertebrates (mammals, birds, fish, reptiles and amphibians) (Deagle et al. 2012, Boyer et al. 2013). Diet identification through DNA barcoding includes the challenges common to molecular systematics (e.g. DNA extraction, PCR amplification and sequencing biases, and analysis). Diet sample collections (e.g faeces, pellets, regurgitate, or stomach lavage) involve unique challenges. The samples contain multiple species, and degraded DNA. These last two features, complex mixtures and short DNA fragment length, both lend themselves to analysis with NGS methods, which are designed around short sequences and generate many thousands of individual sequences. Sample collection varies with species, but can be invasive (stomach lavage) or opportunistic (faeces or regurgitate collection). Invasive methods may result in better quality DNA in the diet samples, but require direct access to the organism and often include euthanasia of the specimen. Opportunistic methods allow more flexibility in accessing the diet material, are not invasive, but often result in degraded DNA.

As part of a larger project investigating fish populations and the possible effects of bird predation on these populations, we have developed a non-invasive molecular method to determine bird diet from regurgitated cough pellets that is broadly applicable to predator-prey diet studies.

Cough pellets are masses of undigested prey parts that are regurgitated by a wide range of different bird species (Higes et al. 2008, Taberlet and Fumagalli 1996). Pellet content is variable across bird species, but each pellet generally includes a variety of plant and animal material including bone, exoskeleton, scales, fur (Johnson et al. 2010). Reconstruction of the diet from a cough pellet faces four major challenges: degraded DNA, taxonomic variation in diet (prey) species, a potentially large number of individual diet items that can make up a single pellet, and complex data analysis. Pellets can include relatively recently consumed prey items, but may also include prey items that were consumed days earlier, leading to varied DNA quality (Johnson et al. 2010). Since DNA from the diet will likely be degraded from the partial digestion process, targeted gene regions (amplicons) are generally kept small (Deagle et al. 2006) to overcome this fragmentation, and avoid biasing results towards better preserved or more recently consumed items (Pääbo et al. 2004). An approximately 300 bp fragment of an appropriate gene region generally includes enough sequence variation to allow species-specific identification while being short enough to amplify from partially digested tissue, but the locus needs to include sufficient variation to match the species of interest (Wang et al. 2010). Second, the diet can also originate from multiple taxa, requiring careful choice of appropriate "universal" primers that amplify from a broad suite of taxa. No primer pair is truly universal, so pairs must be chosen to allow relatively unbiased amplification across a specific diet of interest. Many candidate loci have been identified for which "universal" primers have been developed including Cytochrome Oxidase 1 (CO1), 18S ribosomal RNA (18SrRNA), and 16S mitochondrial rRNA (16SrRNA) (Asgharian et al. 2010, Meusnier et al. 2008, Pompanon et al. 2012). Third, as previously stated an

individual pellet can contain multiple individual prey items. To ensure that all prey items are captured in the DNA analysis a large number of sequences should be obtained for each pellet. Early molecular studies of diet relied on Sanger Sequencing of cloned fragments amplified from extracted DNA. This method while not only being expensive and time consuming limits the sequencing to individual clones that severely limited the coverage of the analyses (Carreon-Martinez et al. 2011). NGS largely overcomes this limitation by inexpensively and quickly allowing the sequencing of upwards of hundreds of thousands of DNA fragments from a single sample (Ekblom and Galindo 2011). Fourth, the potentially complex nature of an individual pellets means that the recovered sequences will need to be aligned and analysed for (multiple) species identification.

In this study, we use Illumina MiSeq sequencing to investigate the fish-based diet of a Double-crested Cormorant (*Phalacrocorax auritus*) population found on Lake Nipissing. The cormorant population on Lake Nipissing has become a concern in recent years, as the possible effect they may have on the fish population has been called into question. As part of a larger study of the effects of these fish-eating birds on fish stocks in Lake Nipissing, we have developed a genetics-based method to determine adult cormorant diet by analysing the genetic make-up of regurgitated pellets collected from their nesting colonies.

2.2 Materials and methods:

Pellet collection:

A set of approximately 150 regurgitated pellets were collected between May 24-May 28, 2014 by hand near cormorant nests on Goose Island in Lake Nipissing, Ontario, Canada, and transported to the laboratory for further processing. Of the 150 pellets, a total of 20 were used for further NGS sequencing. Collected pellets were placed in individual plastic bags and stored at -80◦ C until further processing.

Pellet dissection:

Pellets were thawed to room temperature and a mucus membrane that surrounds each pellet was removed using forceps in 10mM TE buffer at pH 8.5 (10mM Tris, 10mM EDTA). The remaining organic matter (i.e. fish bones, scales, eye lenses, otoliths, tissue, crustacean) was then placed in a Magic Bullet® blender with approximately 15 ml of 10mM TE buffer, ph 8.5 and the mixture was homogenized to a slurry. Following homogenization, the contents were centrifuged at 9000 rpm for 6 mins, and 200 μL of supernatant was collected for DNA extraction.

DNA extraction:

Total DNA was extracted from the aliquot of supernatant using a DNeasy Blood & Tissue Kit (Valencia, CA) following the manufacturer's tissue extraction protocol. All samples were stored at -20ºC until PCR was performed.

DNA amplification:

A series of primer pairs were tested to determine the optimal primer set for DNA amplification and species identification via Sanger sequencing (Table 1). PCR Amplifications were done in 15 μL reactions that contained 1% of template DNA, 1% of primers, 50% of TopTaq polymerase, 1%10XCL, 20% of DNase free water (Qiagen TopTaq polymerase kit). Positive and negative controls were also used to test for PCR reaction quality and contamination. The PCR conditions consisted of an initial denaturation at 95◦ C for 15 min, followed by 33 cycles of 94◦ C for 20 s, 52◦C annealing temperature for1:30 min, 72◦C for 45 s, with a final step at 72◦C for 2 min. PCR products were visualized on 1.2% agarose gels prior to sequencing to verify amplicon presence and appropriate size. Amplified DNA from known fish samples was sent to Génome Québec Innovation Centre for Sanger Sequencing in both directions to obtain and/or confirm amplicon sequence.

Table 2.1: Primer combinations tested to obtain a universal primer pair for the amplification of vertebrate DNA

Primer Name	Gene	Sequence	Size of Amplicon (bp)
Chord_16S_F	Mitochondrial 16S	CGAGAAGACCCTRTGGAGCT	
Chord_16S_R	Mitochondrial 16S	GGATTGCGCTGTTATCCCT	550
Chord_16 s_R_Short	Mitochondrial 16S	CCTNGGTCGCCCCAAC	13-221
LCO1490 (F)	CO1	GGTCAACAAATCATAAAGATATTGG	
C1-N-1777 (R)	CO1	ACTTATATTGTTTATACGAGGGAA	229-235
F1	CO1	TCCACTAATCACAARGATATTGGTAC	
R1	CO1	GAAAATCATAATGAAGGCATGAGC	150
F2	CO1	TCAACCAACCACAAAGAGATTGGCAC	
R2	CO1	TAGACTTCTGGGTGGCCAAAGAATCA	650
F3	CO1	TCGACTAATCATAAAGATATCGGCAC	
R3	CO1	ACTTCAGGGTGACCGAATCAGAA	650
18e (F)	18S	CTGGTTGATCCTGCCAGT	
18I (R)	18S	TCTCCGGAATCGAACCCTG	350
18j (F)	18S	ACCGTGGGAAATCTAGAGCTAA	
18R (R)	18S	CTACGGAAACCTTGTTACG	400

Next-generation sequencing library prep

Extracted DNA was sent to the United States Geological Survey Leetown Science Centre (LSC) in Kearneysville, WV. After a pilot run we chose the mt16S universal primers, as the 18S primers amplified species identification poorly in our hands. Metagenomic amplicon sequencing libraries were prepared and sequenced at the USGS-LSC. As we were looking to identify fish prey from regurgitated cormorant pellets, we initially looked at both 18S and mitochondrial (mt) 16S gene regions for use with the Illumina MiSeq platform. The Iwanowicz Laboratory set the sequencing protocol, modifying standard

procedures. Amplicons for the NGS library were produced in two steps. First, 'standard' primers were used to generate a high concentration of input template, followed by less efficient 'fusion' primers that incorporate sequencing adapters. The initial amplification reaction used primers 16SF1 (5' – GAC GAK AAG ACC CTA – 3') and 16SF2 (5' – CGC TGT TAT CCC TAD RGT AAC T – 3') (Deagle et al. 2007). The thermocycle conditions consisted of an initial denaturation step of 94ºC for 3 min, followed by 35 cycles of 45 s at 94ºC, 45 s at 52ºC, and 1.5 min at 72ºC, and a final extension of 72ºC for 7 minutes. An appropriately sized amplification product was confirmed for each reaction by electrophoresis of 5 µL of the reaction product through a 1.2% agarose gel at 90V for 45 min. Polymerase chain reaction (PCR) products were cleaned with the Qiagen PCR Purification Kit (Valencia, CA) and quantified using the Qubit double stranded DNA High Sensitivity Assay Kit (ThermoFisher Scientific, Grand Island, NY). Samples were diluted in 10mM Tris buffer (pH 8.5) to a final concentration of 5 ng/µL.

We then developed custom fusion primers extending the published degenerate mt16S rDNA primers (16S1F and 16S1R) (Deagle et al. 2007) with Illumina sequence adaptors (bold), (modified 16S1F) **TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG** GAC GAK AAG ACC CTA and (modified 16S1R) **GTC TCG TGG GCTCGG AGA TGT GTA TAA GAG ACA G** CGC TGT TAT CCC TAD RGT AAC T. These primers produced a ~550 bp fragment that could be subjected to 300 bp paired –end sequencing, recovering mt16S sequence.With the normalized DNA from the initial mt16S region amplification, we followed the protocol for 16S Metagenomic Sequencing Library Preparation (CT #: 15044223 Rev. B) except for the replacement with the mt16S

primers. Libraries were diluted 1:10 with molecular grade water and quantified with the Qubit dsDNA HS Assay Kit. DNA quality and amplicon size was determined using the Agilent DNA 1000 DNA kit (Santa Clara, CA). Pooled libraries were diluted to 4 nM using 10 mM Tris pH 8.5. A final 15 pM amplicon library was created with a 6.5% PhiX control spike and sequenced using the Illumina MiSeq 2 x 300 bp paired-end technology.

Bioinformatics analyses

Bioinformatics analyses were performed in the office of Bioinformatics and Genomics at the Health Sciences North Research Institute, Sudbury, Ontario. The Illumina MiSeq paired-end sequences obtained were trimmed using Sickle ver. 1.33 (Joshi and Fass, 2011). The program Sickle was used with a parameter window of 20 base pairs, where sequences were trimmed when their quality score was lower than 20 and the sequence length was below 10 base pairs. The program Spades ver. 3.9.0 (Bankevich et. al. ,2012) was used for error correction in the trimmed sequences. This, with the overlapping paired-ending strategy, improved the amplicon assembly (Schirmer et. al., 2015). The pair endings were paired using the program PANDAseq ver. 2.9 (Masella et. al. 2012) with a minimum overlap of 20 base pairs. The resulting sequences were sorted, dereplicated and singletons were discarded using Userarch8 ver. v8.1.1756_i86linux32 (Edgar et. al. 2010). The UCHIME program (Edgar, 2011) used the chimera reference filtering system to eliminate any missed chimeras. This strategy used the database of 16S rRNA mitochondrial sequences from fish in Lake Nipissing (Appendix 1) These sequences were obtained through Sanger sequencing and downloaded from NCBI (Appendix 3). These fish sequences were curated and added to the Gold Database, which

includes 16S rRNA bacterial genes, obtained through the UCHIME site (http://drive5.com/uchime/uchime_download.html). This was due to the fact that fish and bacterial DNA were the predominant taxa following the BLAST, and the study was interested in fish as a target. This new database was used as a reference filter for the UCHIME program. In this way, chimera sequences were eliminated and did not affect the rest of the pipeline analysis. I used Robert Edgar's python scripts uc2otutab.py (http://drive5.com/python/python_scripts.tar.gz) to label consecutively the Operational Taxonomic Units (OTU) generated in these analyses and to generate tables and statistics obtained for each OTU construction per sample.

The sample taxonomic profiles were produced using BLAST. The first data was blasted against the nt database downloaded from NCBI with the parameters: blastn -db nt/ -query out.foo -evalue 0.00001 -perc_identity 95 -outfmt '6 qseqid sgi pident slen mismatch gapopen qstart qend sstart send evalue bitscore staxids' -out out_6.foo. This result was called "global blast". For the second approach, a database was created based from the sequences selected and downloaded from NCBI and those sequenced in the laboratory in order to cover most of the expected fish 16s rRNA sequences described in Lake Nipissing (Appendix 1). The sequences were curated and joined in a fasta file with headers as indicated in the NCBI web site: reference number, taxon id number, and scientific name.Once this file was in the right format, a "local" database was generated for BLAST. The commands used to generate this database were: makeblastdb -in foo.fa -parse_seqids -dbtype nucl -taxid_map foo.map. The BLAST commands were described previously, but the database used was created with the makeblastdb program. This result was called "local blast". The program MEtaGenome Analyzer (MEGAN) ver 5.11.3, build 21

(Huson et al. 2016) was designed for taxonomic classification of large sets of environmental data using the Blast output (Mitra et al. 2012). Taxonomic rank was assigned to each OTU at a given statistical confidence of percentage of significant alignment to a pre-determined reference sequence. These values were percent alignment intervals were arbitrarily assigned as being 75% for the global blast and 90% for the local blast; these were then used to evaluate the levels of uncertainty in the blast hits for each OTU. With the taxonomic names and the OTU counts we used the R module Phyloseq ver 1.6.1 (Mudie and Holmes, 2013) to quantify the reads for each taxonomic rank for each sample site. Different taxonomic ranks were evaluated and quantified.

Known fish samples

To test potential primers for amplification across fish species and DNA fragments for sequence variation to allow identification of species, we extracted DNA from tissue samples from twenty-seven different species of fish, obtained from the Ontario Ministry of Natural Resources and Forestry. DNA fragments were amplified using the various primer combinations **(Table 2.1)** and DNA sequences determined by Sanger sequencing at Génome Québec. Forward and reverse reads were trimmed, and paired using CodonCode, to obtain consensus sequences for known samples. BLAST analysis was then performed against the NCBI database; all twenty-seven sequences were successfully identified (<98% confidence).

2.3 Results and Discussion

By using a combination of molecular genetics and informed data analysis, we were able to describe and quantify the diet of the piscivorous bird, Double-crested Cormorant, using regurgitated cough pellets. The pellets are easily collected from bird habitat, and provide a snapshot of recent diet. All pellets we analyzed contained sufficient DNA of high quality for amplification and sequencing of diet contents yielding an average of 93,948 sequences/pellet (total of 20 pellets analyzed). Using an informed search of our data set against only sequences from fish species known to be potential diet items in the environment of interest allowed us to reduce the percentage of unresolved sequences almost ten fold (from 42.77% to 4.48%) and substantially improved our ability to resolve the cormorant diet to individual species.

Primer optimization

We tested eight different primer pairs from five different loci and determined that the 3' region of the 16S mitochondrial gene provided the best resolution to the species level of the known cormorant diet in this watershed. Meta-barcoding, the large-scale assessment of molecular biodiversity from an environmental DNA sample, can be applied to a range of different samples (soil, water, feaces, etc) and species (Deagle et al. 2014). In addressing the specific question of cormorant diet in Lake Nipissing, we were able to use prior knowledge of the species and location to inform this search. The Double-crested Cormorant (DCCO) feeds primarily on fish, ingesting 400-600g of fish per day (Ridgeway 2010), and our initial list of potential primers was based on previous studies examining fish species (Deagle 2014, Hebert et al. 2004 and Meusnier 2010). In our

particular case, we were specifically interested in one location and all samples were obtained from Goose Island, colonized by DCCOs in Lake Nipissing, Ontario, Canada. We focused on a single location, in order to optimize our data analysis before moving to a large-scale analysis. This lake is a popular commercial, ceremonial, and sport fishing lake, and is relatively accessible to a large human population. Because of the importance of this fishery, fish populations in Lake Nipissing are closely monitored and using this monitoring data we were able to obtain a species list of all fish known to reside in the basin, as well as tissue samples from the majority of the most common species.

The ability of our universal primers to identify and classify a wide range of species is directly related to the conserved and variable regions of the amplified sequence. PCR amplification and Sanger Sequencing of the DNA extracted from the positive controls showed the 16S ribosomal RNA gene to be the suitable candidate for our particular application. We also tested two other genes, CO1 and 18S, which have been used in previous research for the genetic identification of fish (Hebert et al. 2003, Bowser et al. 2013, Zhang and Hanner 2012). Preliminary investigation of seven fish species indicated that the 3' region of the CO1 gene included the necessary level of variation to resolve sequences to the species level. However, no primer set tested amplified all species in the basin, and thus, multiple primer sets would have to be used in order to get a good representation of genetic diet from a single environmental sample. In contrast, we were able to amplify a fragment of the 18S gene from all of our known fish samples using a single primer pair, but there was insufficient sequence variation to allow species-level resolution. In other words, CO1 was too variable, while 18S was too conserved, for

application to our system. In contrast, the Chord_16S_F and Chord_16S_R primers (**Table 2.1**) were able to amplify all 27 positive controls tested. Sanger sequencing and alignment data showed enough variation between sequences to allow for species level identification. In addition to having the ability to taxonomically identify a wide range of different species, this particular primer pair was also able to distinguish closely related species from one another, such as Walleye (*Sander vitreus*) and Yellow Perch (*Perca flavencens*), which was confirmed with Sanger Sequencing.

Next Generation Sequencing

Following the selection of an appropriate primer set, we performed NGS on 20 individual pellets. DNA was extracted from cough pellets and sent to the USGS-LSC in Kearneysville, WV, USA, for sequencing. Prior to NGS, DNA from each extraction was quantified spectrophotometrically and then checked for quality by PCR amplification; all 20 pellets yielded DNA of sufficient quantity and quality for amplification. Following sequence cleanup, the NGS analysis, resulted in 1,878, 970 reads being identified from the pellets collectively with a mean of 93,947 reads/pellet (s.d. = 40,176). The large amount of data collected allows for relatively sensitive dietary information to be extracted from these cough pellets, with a minimal amount of time, effort and money.

Species Identification

To identify the diet components in each pellet, we ran a series of sequence similarity analyses of increasing specificity. The sequences were initially compared to the entire NCBI-NT database in an undirected fashion, we then focused our analysis strictly on the

sequences identified as fish, and finally, compared our sequences to fish species found in the watershed of interest. Each run consisted of a few hours, it should be noted that not all 1.8 million sequences were run, only sequences that were identified as unique for each OTU, and they were then later quantified.

Using the BLAST approach and MEGAN analysis, we were able to identify 57.23% of the 1,878,970 sequences to the species level (70% as the percentage of significant alignment cutoff, chosen arbirtaritly for highest number of taxa identification). A total of 36 different species, representing 34 genera were identified and their relative abundances quantified (**Figure 2.1**). In addition we classified sequences only to the level of genera and were able to identify 70.52% of sequences (70% significant alignment cutoff), representing a total of 43 genera (**Figure 2.2**). This genus-based analysis substantially reduced the percentage of unresolved sequences, from 42.77 to 29.48, adding nine genera to our analysis, including *Percina* and *Catostomus* but prevented us from resolving some known diet species including *Esox lucius* (pike) and *Esox masquinongy* (muskie). The reduction in unresolved sequences was promising, but our particular question, the specific diet of the DCCO, requires species-level discrimination. These analyses include all taxa including insects that may be by-catch of cormorant feeding or gut content of consumed fish and a large percentage of sequences that represent products of the digestive tract or contamination from the environment (e.g. human, dog, cat, and bacterial sequences).

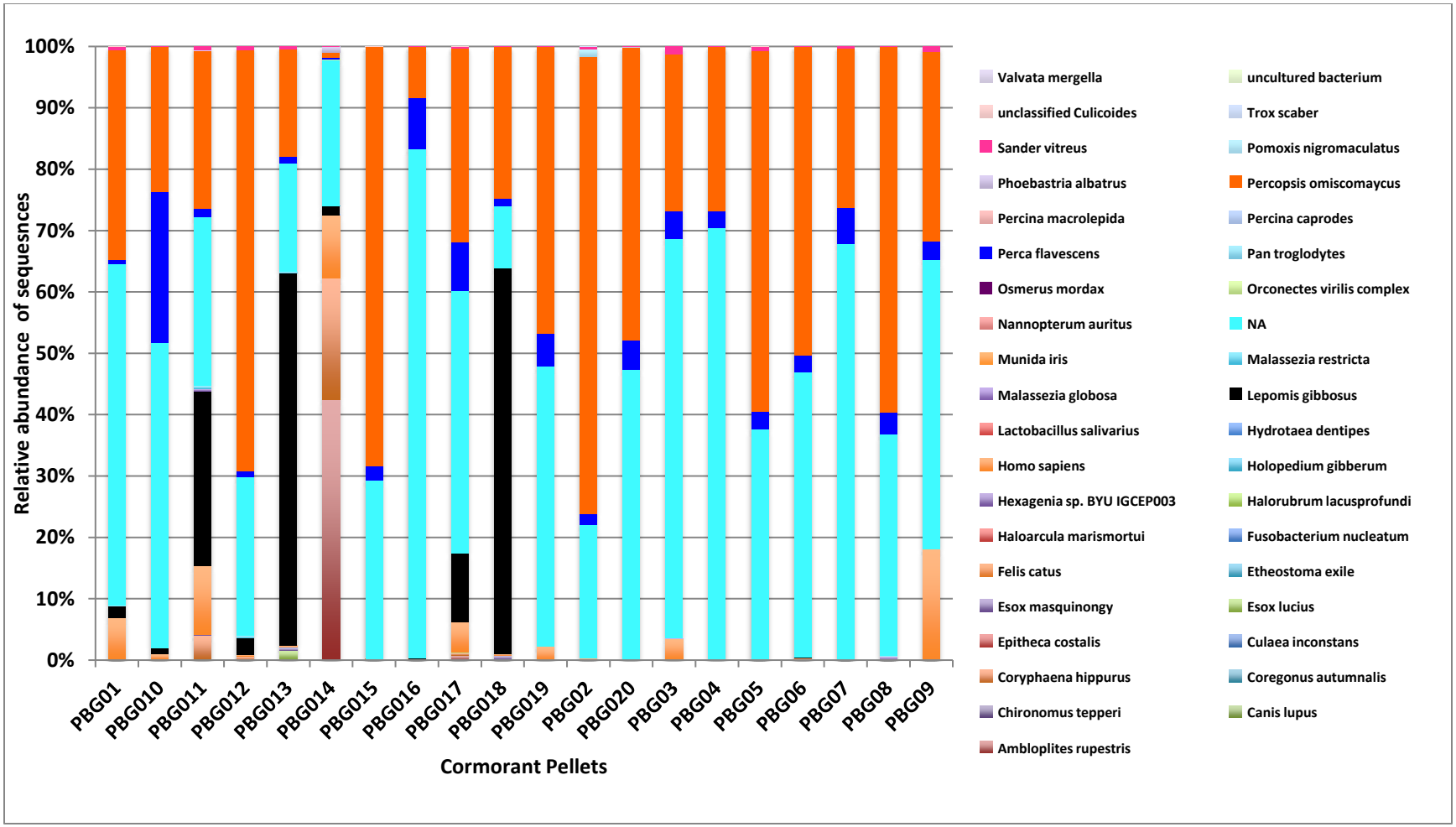


Figure 2.1 Relative abundance of reads obtained from a global BLAST analysis of the 20 cormorant pellets to species level

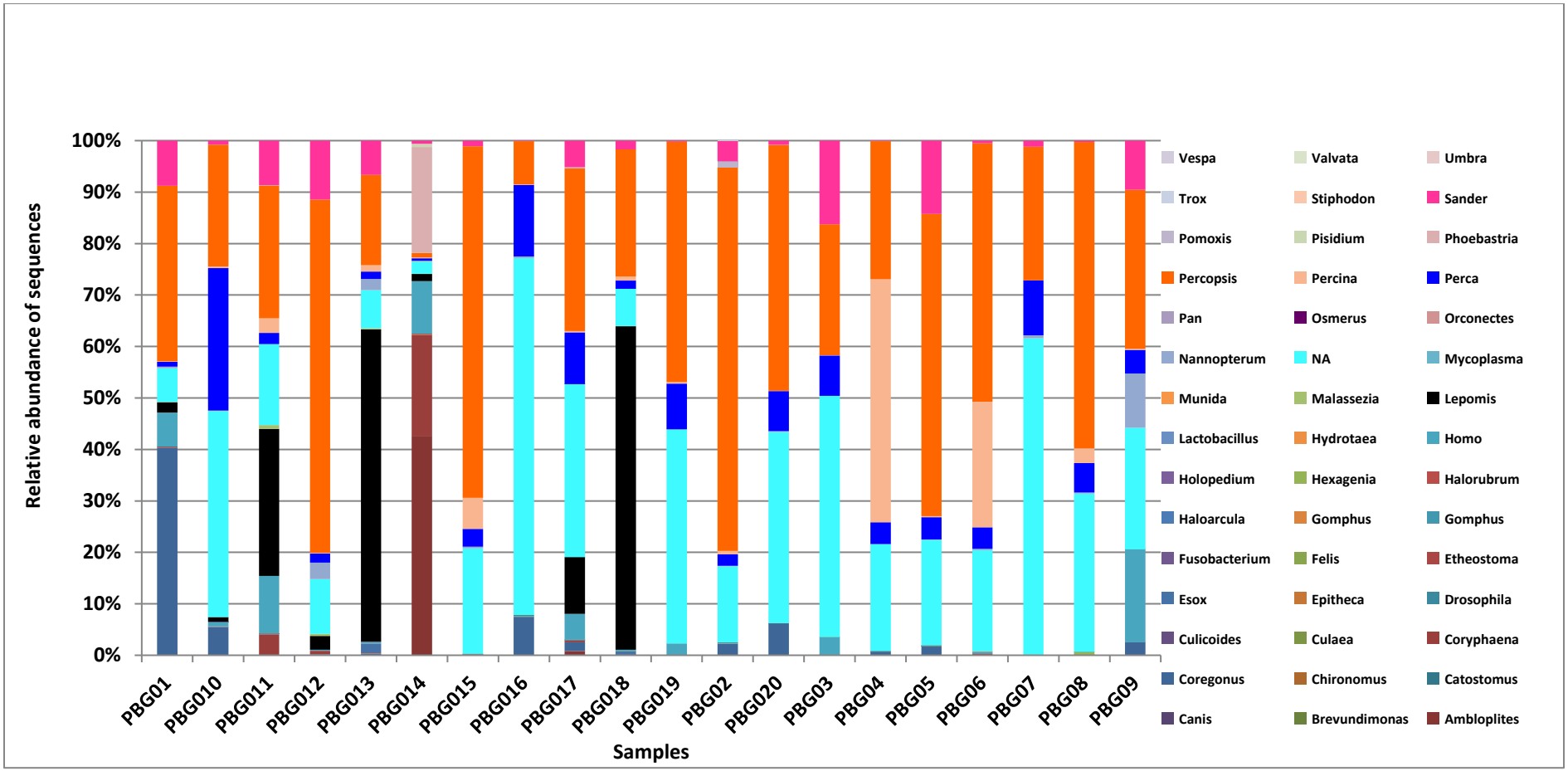


Figure 2.2 Inferred DCCO diet from relative abundances of all sequences from 20 cough pellets collected from Goose Island in 2014.

Identification of the genus was done by global BLAST using a 70% significant alignment cutoff

To focus our analysis, we restricted our description to the taxonomic group that is believed to make up the majority of the DCCO diet, ray-finned fishes (Actinopterygii). This approach to our analysis consisted of a total of 1,784,105 sequences (95% of total reads) identified as Actinopterygii, with an average of 89,205 sequences/pellet (s.d = 40,831). With this approach we identified 57.24% of the sequences to the species level with a 70% significance alignment cutoff, representing a total of 15 species of fish and 13 genera (**Figure 2.3**). In addition, if we accept classification to the level of genus (as in the initial analysis), we also see substantial improvement in our ability to identify sequences, with 69.60% of sequences classified to genus with a 70% significant alignment cutoff, representing a total of 15 species of fish and 14 genera (**Figure 2.4**).This examination of only fish species still left a substantial fraction of the sequences unresolved. Closer examination of these unresolved sequences suggested that a substantial fraction were being misidentified as closely related congeners not found in this watershed, likely reflecting the relatively short nature (approx. 550bp) of the sequences and the limited sequence divergence between these closely related species. The local species were often being identified, but at a lower score (below 70%) than we had set for this analysis. For example individual sequences were identified as either *Perca flavencens* (North American perch) or *Perca fluviatilis* (European perch), although survey data indicates that only North American perch inhabit Lake Nipissing.

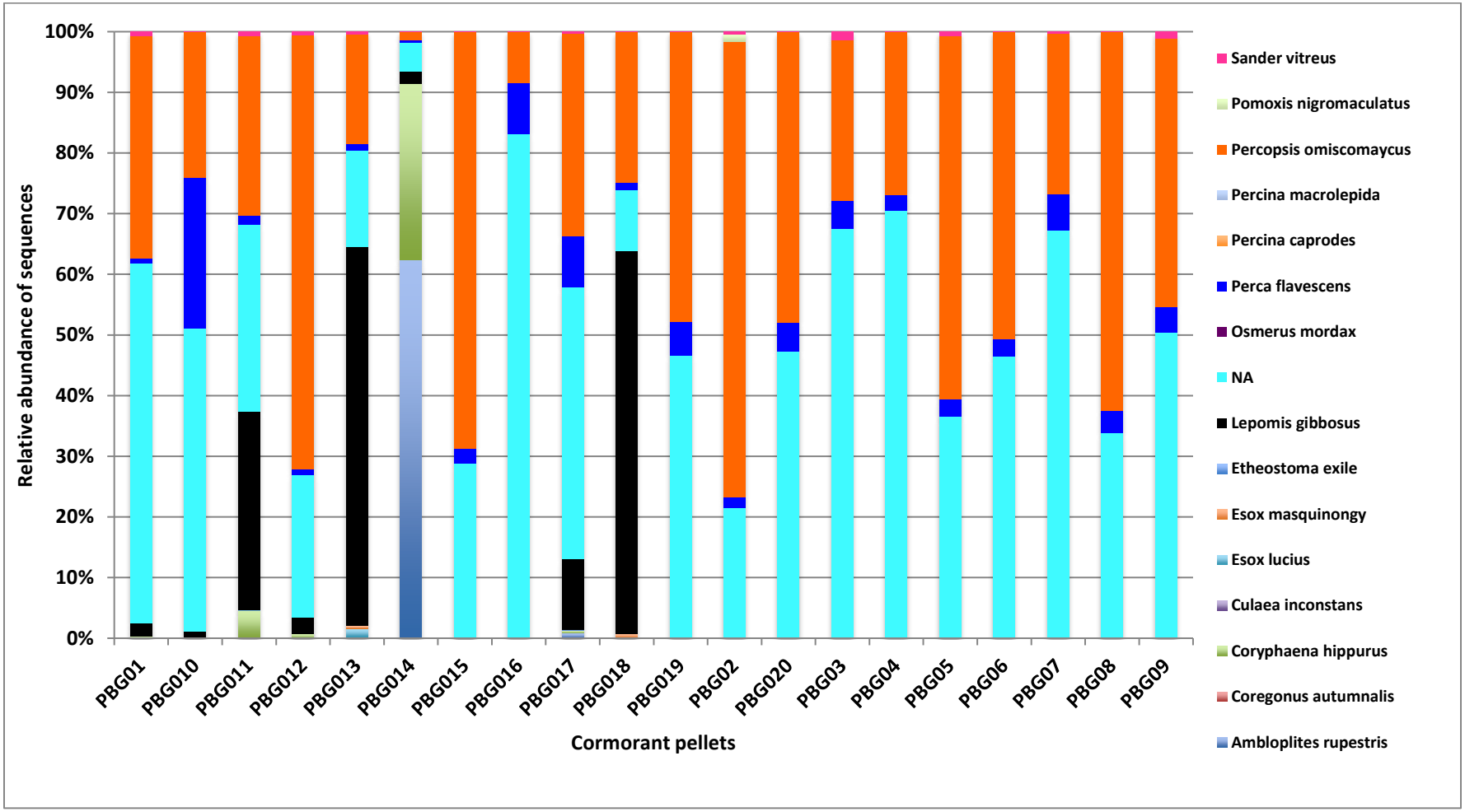


Figure 2.3 Inferred DCCO diet from 20 cough pellets collected from Goose Island in 2014. Only sequences from the Actinopterygii class were included. Sequences were identified at the species level via global BLAST, and MEGAN analysis using a 70% significant alignment cutoff.

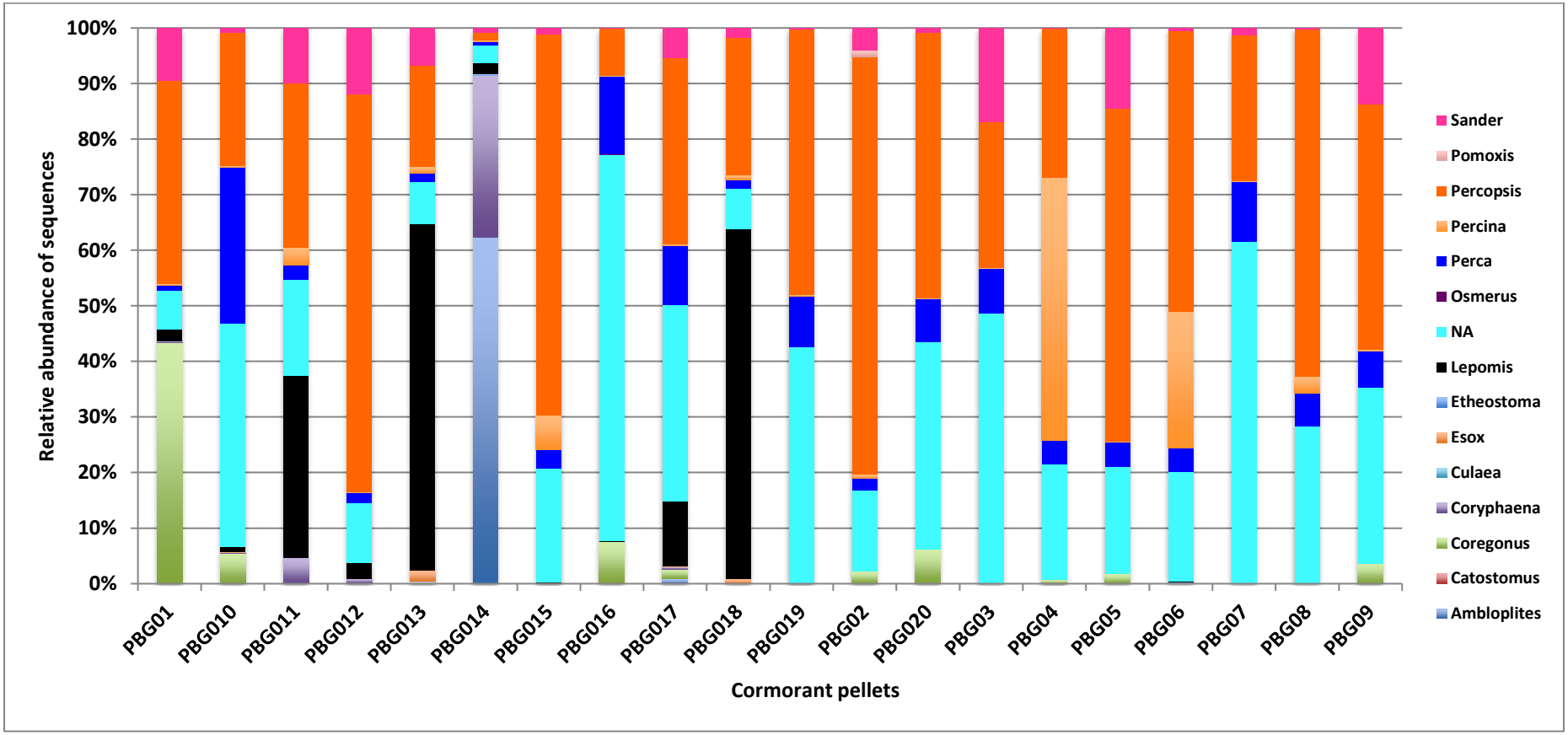


Figure 2.4 Inferred DCCO diet from 20 cough pellets collected from Goose Island in 2014. Only sequences from the Actinopterygii class were included. Sequences were identified at the genus level via global BLAST, and MEGAN analysis using a 70% significant alignment cutoff.

In light of this potentially confounding effect of closely related, but geographically improbable species, we performed an additional analysis taking into consideration both the piscivorous nature of DCCO and the fish found in Lake Nipissing (identified through monitoring data). We performed a second BLAST/MEGAN analysis, this time using a custom database instead of the NCBI-NT. Our custom database was composed of 46 different sequences, from the vast majority (87.2%) of fish species known to reside in the Lake Nipissing (41/47). We were able to compile "local" database of DNA sequences related to almost all fish found in the basin, using data from the NCBI database as well as sequence data obtained from positive controls (Appendix 3). This analysis identified a total of 1,861,745 sequences with an average of 93,087 sequences/pellet (s.d. = 40,023). A total of 95.52% were identified to the species level with a 90% significant alignment cut-off, representing a total of 15 different species from 14 genera with individual species making up 0.1 to 37% of the diet (**Figure 2.5**). These results represent a substantial improvement in the percentage of unresolved sequences reducing the number from over 45% to just under 5%. The focused analysis also allowed inclusion of a much larger number of sequences in the analysis increasing from 1,784,105 in the global analysis to 1,861,745 in the focused analysis. The increase in sequences could be related to un-resolved non-fish sequences, or could also be sequences that were unable to be identified in the Global BLAST. Overall, the technique clearly identified the diet of this species in this environment and these results highlight the utility of using species- and location-specific information in this kind of meta-barcoding ecosystem analysis.

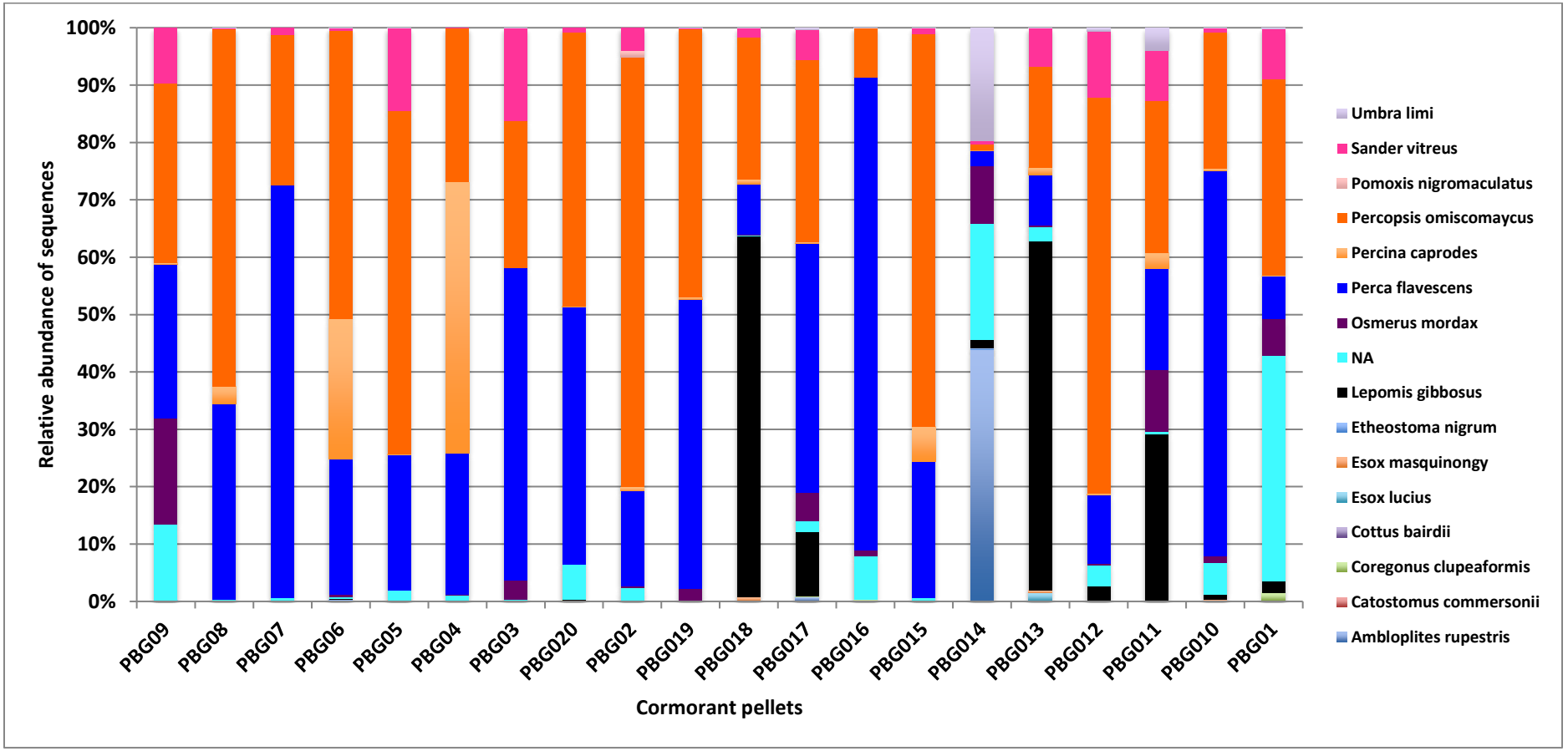


Figure 2.5 Local BLAST - Inferred DCCO diet from 20 cough pellets collected from Goose Island in 2014. Sequences were identified at the species level via local BLAST, and MEGAN analysis using a 90% significant alignment cutoff

Results Summary:

From the initial "global" BLAST/MEGAN analysis, we were able to genetically identify the constituents of the 20 cough pellets samples. The results of this analysis, which included all identified sequences, showed that DCCO diet at this location is mainly composed of the following : *Percopsis omiscomaycus* (35.84%), *Lepomis gibbosus* (9.90%), *Perca flavenscens* (5.37%), and unresolved sequences (42.77%) (**Figure 2.1**). A second analysis consisted of using the classification of Genera as the cutoff. This type of analysis was possible without losing too much accuracy (ability to identify closely related species), except for the ability to distinguish between Pike (*Esox lucius*) and Muskie (*Esox masquiqongy*). According to this analysis resulted the overall relative constituents of the diet were: *Percopsis* (35.84%), *Lepomis* (9.90%), *Perca* (7.54%), *Sander* (4.40%), and unresolved sequences (29.48%) (**Figure 2.4**).

The initial data analysis showed that *Percopsis* ,*Lepomis. Perca, and Sander* combine to account for greater than of 70% of DCCO diet. Although there seemed to be a high level of sequences that were unable to reach species level classification, our data suggests that the majority of these sequences (94.95%) are from the Actinopterygii class (fish). With this in mind, we decided to focus our analysis by reanalyzing all sequences that were identified as Actinopterygii.. The analysis yielded similar results to the initial results examining all sequence data obtained. The most abundant species were found to be *Percopsis omiscomaycus* (37.35%), *Lepomis gibbosus* (10.43%), *Perca flavenscens* (5.65%), and unresolved (42.76%) (**Figure 2.3**). Again, we performed a secondary

analysis using genera as cut-off for classification. From this the overall relative contribution of diet was composed of the following: *Percopsis* (37.75%), *Lepomis* (10.43%), *Perca* (7.52%), *Sander* (4.63%), or unresolved sequences (30.40%) (**Figure 2.4**). Our analysis of the fish sequences still resulted in a high number of unresolved sequences (43% at species level and 30% at genera) which calls into question the accuracy of the relative percentage of each species' contribution to DCCO diet. Any bias in the inclusion of sequences in the unresolved category – i.e. if one species is less likely to be unresolved – would confound our results

Our local BLAST approach, focusing analysis only on species known to be in the basin, reduced the unresolved sequences to less than 5% and identified the most abundant species to be *Perca flavenscens* (36.57%), *Percopsis omiscomaycus* (36.17%), *Lepomis gibbosus* (9.99%), *Sander vitreus* (4.44%), unresolved (4.48%), and all other species identified (8.35%) (**Figure 2.5**).

Conclusion

Our analysis shows that, with prior information of the environment and species from which a given cough pellet is collected from, one can extract useful dietary information. Although this type of information is extremely useful, there are of course downfalls to this type of approach. One of the issues related to "Universal" primers is primer bias. Primer bias refers to the preferential binding of any designed oligonucleotide primer to the gene sequence of a particular group of species, or single species. This issue has been discussed in depth in various molecular diet studies using universal primers (Elbrecht and

Leese 2016, Deagle et al 2014). Although unavoidable, we suspect the effect of primer bias in this study is minimal, as the primer binding sites are highly conserved. In addition to primer bias, the quality of DNA can also be a cause for concern when dealing with any sort of environmental sample (Shokralla et al. 2015). This issue didn't seem to affect the results, as a significant amount of DNA was extracted and sequence from each pellet tested. This could be due to the relatively short fragment size of the barcode (550 bp) that can be recovered from degraded DNA.

This non-invasive molecular technique has the potential to be used as a long-term monitoring system for the MNRF. This method has the potential to monitor not only LN over a long period of time, but the method also has the robustness to be applied to other nesting cormorant populations across the globe. Furthermore, it is possible to apply this technique to other birds producing cough pellets, such as owls or hawks.

Chapter 3 A Genetics-based assessment of DCCO diet on Lake Nipissing

3.1 Introduction

The following chapter is the summary of a two-year study of DCCO diet identified that used a novel combination of genetics-based techniques and analysis (NGS,BLAST, and MEGAN) and was presented to the North Bay District Office, Northeast Regional Operations Division of the Ministry of Natural Resources and Forestry in July 2016. Lake Nipissing has been home to an increasing nesting population of Double-crested Cormorants (DCCO), which has been closely monitored by the Ministry of Natural Resources and Forestry (MNRF) over the last twenty years (**Figure 1.1**). The increase in DCCO population has coincided with a decrease in walleye (*Sander vitreus*) stocks. Walleye is an economically important species both locally and regionally identifying the cause or causes of its decline, is a priority for the MNRF. DCCO feed on fish, including walleye, and a possible link between the decrease in walleye stocks and the increase in DCCO populations has led to speculation that the DCCO may be driving the loss of fish (Young 2014). To investigate the correlation, and possibly determine if the DCCO populations could be driving the decline in walleye, we were approached by the MNRF and tasked to develop a genetics-based technique that could be used to determine DCCO diet from their cough pellets. These cough pellets are regurgitated approximately every 24 hours by the cormorants, and contain undigested parts of their meals (eye lens, bones, scales, exoskeletons). We developed a method, using a combination of Next-Generation Sequencing of the 3' end fragment of 16S mitochondrial gene and data analysis (BLAST

and MEGAN), which allowed us to estimate the components of the pellet from genetic information collected from regurgitated DCCO pellets (see Chapter 2 for details of methods and analysis).

3.2 Lake Nipissing Pellet Collection

Over the course of two years (2014-2015), the MNRF collected regurgitated adult DCCO cough pellets from four separate nesting locations on Lake Nipissing, at two different summer time points. The pellets were collected from the Gull, Goose, Wigwam, and Hardwood Islands (**Figure 3.1**), representing approximately 95% of the overall nesting cormorant population on Lake Nipissing (**Figure 1.1**). The two time-points from which pellets were collected were Pre-Chick (PC) and Post-Hatch (PH) time-points. PC pellets were collected prior to the DCCO chick hatching (between May 24 and May 28), while PH pellets were collected during the post hatching stage of the nesting period (between June 23 and July 6). Over 2 years, a total of 1554 pellets were collected, and 176 of these pellets were randomly selected and further analyzed for diet using the method described in Chapter 2 **(Table 3.1).** In total, we were able to obtain 16S gene sequence data from all pellets samples, resulting in a total of 19,591,626 sequences (**Table 3.2**). This chapter describes the findings of the inferred diet of DCCO from cough pellets, including a Yearly, Temporal, and Spatial analysis and was initially presented to the MNRF as a final report of this DCCO diet project.

Image Source: Google earth (January 2017)

Figure 3.1 Map of lake Nipissing and the location of each respective island. 1. Goose Island, 2. Gull Island, 3. Hardwood Islands, 4. Wigwam Island. The number of pellets collected from each time-point can be seen in **Table 3.1**.

Table 3.1: Number of pellets collected from time-points at each respective nesting colony

	2014		2015	
Nesting Colony	**Pre-Chick**	**Post-Hatch**	**Pre-Chick**	**Post-Hatch**
Gull Island	20 (131)	N/A	12 (160)	12 (142)
Goose Island	20 (151)	N/A	12 (122)	12 (145)
Wigwam Island	20 (106)	N/A	12 (152)	12 (52)
Hardwood Island	20 (84)	N/A	12 (154)	12 (81)

Number of pellets collected from each nesting colony. The number of pellets that were analysed by NGS are listed, and the total number of pellets collected from each colony is listed in brackets.

Table 3.2: Total number of sequences obtained from NGS data processing

Comparison	Data Set	# of Pellets	# of Sequences	Avg. Sequence/Pellet
Yearly	2014-2015	176	19591626	111316
Yearly	2014	80	9996824	124960
Yearly	2015	96	9594802	99946
Temporal	2014-2015	176	19591626	111316
Temporal	Pre-Chick 2014	80	9996824	124960
Temporal	Pre-Chick 2015	48	1726824	35975
Temporal	Post-Chick 2015	48	7867978	163916
Spatial	Goose Island PC 2014	20	1861745	93087
Spatial	Goose Island PC 2015	12	462402	38533
Spatial	Goose Island PH 2015	12	2446571	203880
Spatial	Gull Island PC 2014	20	1857114	92856
Spatial	Gull Island PC 2015	12	407677	33973
Spatial	Gull Island PH 2015	12	889034	74086
Spatial	Hardwood Island PC 2014	20	3055412	152770
Spatial	Hardwood Island PC 2015	12	409667	34139
Spatial	Hardwood Island PH 2015	12	539563	332734
Spatial	Wigwam Island PC 2014	20	3222553	161127
Spatial	Wigwam Island PC 2015	12	447078	37256
Spatial	Wigwam Island PH 2015	12	3992810	44964

Number of sequences obtained from each time-point and location from the NGS data over a 2 year study.

3.3 Data Analysis

Cormorant diet 2014-2015 - yearly comparison:

Our initial overall analysis consisted of a total of 192 pellets. Because of technical difficulties, few pellets were collected in the 2014 PH time period and we only were able to analyze 16 pellets from Goose Island. Because of this small sample size, we did not include these samples in the overall analysis of the DCCO diet. The final analysis consisted of a total of 176 pellets representing a total of 19,591,626 DNA sequences. The resulting BLAST/MEGAN analysis of the cough pellets is shown in **Figure 3.2**, in which

the overall relative abundance of sequences is broken down to compare the amalgamated data (2014-2015), to each respective year.

Across both years, and all 176 pellets (both time points and all four locations), and 19,591,626 16S sequences, we infer that, overall, DCCO diet consisted of the following relative abundance: 43% yellow perch (*Perca flavescens*), 27% trout-perch (*Percopsis omiscomaycus*), 12% pumpkinseed (*Lepomis gibbosus*) ,7% walleye (*Sander vitreus*), 4% rainbow smelt (*Osmerus mordax*), 4% unresolved , 3% other species (primarily rock bass (*Ambloplites rupestris*), logperch (*Percina caprodes*), mudminnow (*Umbra limi*) and whitefish (*Coregonus clupeaformis*) .

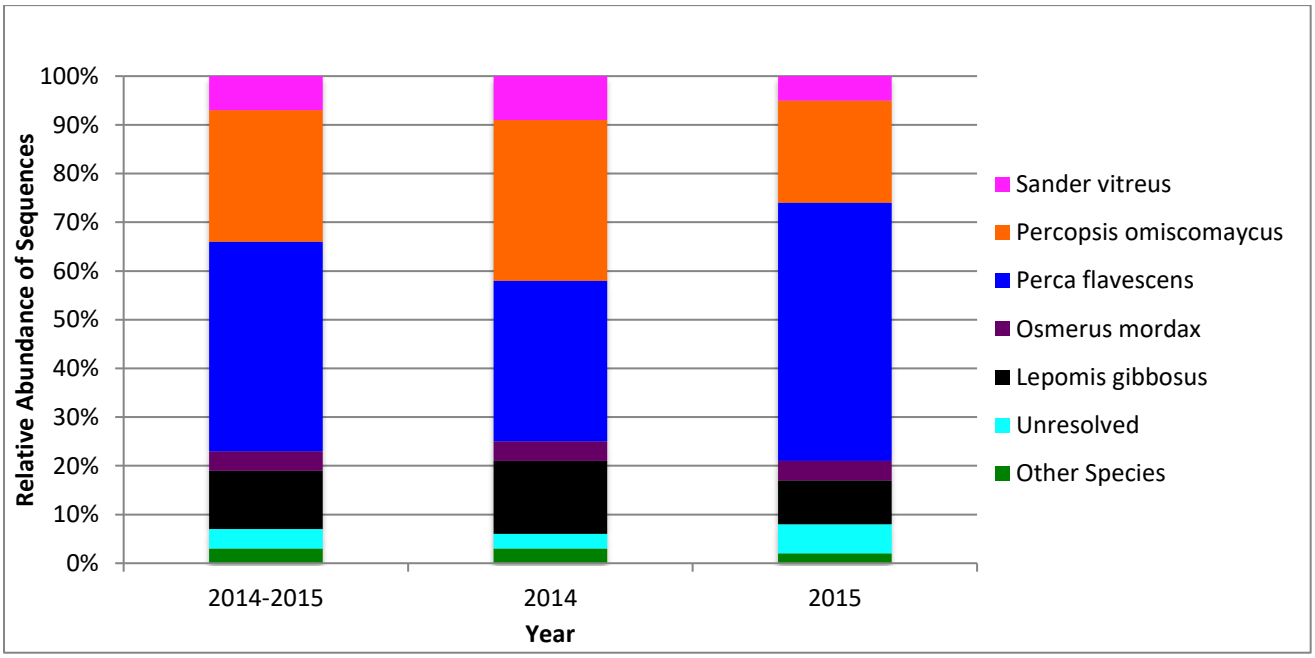


Figure 3.2 Yearly comparison of the relative abundance of sequences from the most abundant fish species obtained from NGS and local BLAST/MEGAN analysis of DCCO pellets.

We also determined diet for each year separately. The 2014 overall DCCO diet was inferred from the analysis of 80 pellets collected from four locations and only one time-point (80 PC). Local BLAST/MEGAN analysis of 9,996,824 sequences obtained from NGS resulted in the following relative abundance of sequences: 33% perch, 33% trout-perch, 15% pumpkinseed, 9% walleye, 4% rainbow smelt, 3% unresolved, and 3% other species (primarily rock bass, logperch, mudminnow, and muskie; **Figure 3.2**). The 2015 overall DCCO diet was obtained from the analysis of 96 pellets collected from four locations and two time-points (48 PC and 48 PH). Local BLAST/MEGAN analysis of the 9,594,802 sequences resulted in the following abundance of sequences: 53% perch, 21% trout-perch, 9% pumpkinseed, 6% unresolved, 5% walleye, 4% rainbow smelt, and 2% other species (mostly rock bass, whitefish, mudminnow, and log perch; **Figure 3.1**).

Although the exact percentage of relative abundance does fluctuate between years, there is a consistent hierarchy to the overall diet composition in both years with yellow perch, trout-perch, and pumpkinseed accounting for over 80% of all sequences identified. Walleye, the primary economic species of interest, accounted for less than 10% of the diet in both years. In addition to the yearly comparison of the overall DCCO, it was also important to compare PC and PH, the two time points within each year, to see if any clear change in diet could be related to the nesting period of the colony.

Cormorant diet 2014-2015 – Temporal comparison

In order to determine if there were large differences in diet depending on time period (PC versus PH), we grouped pellets by respective time-points and compared relative diet abundances (**Figure 3.3**). This analysis included the same 176 cough pellets, but separated each time-point into its own respective diet analysis.

The 2014 PC pellet collection consisted of a total of 80 samples (20/site). Local BLAST/MEGAN analysis of 9,996,824 sequences obtained from NGS resulted in the following relative abundance: 33% perch, 33% trout-perch, 15% pumpkinseed, 9% walleye, 4% rainbow smelt, 3% unresolved, and 3% other species (primarily logperch, rock bass, mudminnow, and muskie).

The 2015 PC pellet collection consisted of a total of 48 samples (12/site). Local BLAST/MEGAN analysis of 1,726,824 sequences resulted in the following relative abundance: 44% perch, 24% trout-perch, 12% pumpkinseed, 8% walleye, 7% rainbow smelt, 3% unresolved, and 2% other species (primarily logperch, rock bass, and mudminnow).

The 2015 PH pellet collection consisted of a total of 48 samples (12/site). Local BLAST/MEGAN analysis of 7,867,978 sequences resulted in the following relative abundance: 55% perch, 20% trout-perch, 8% pumpkinseed, 7% unresolved, 4% rainbow smelt, 4% walleye, and 2 % other species (primarily rock bass, whitefish, and mudminnow).

Similar to the yearly comparison, this temporal analysis resulted in an identical hierarchy of fish abundance across all time-points: perch, trout-perch, pumpkinseed, walleye, rainbow Smelt, unresolved, and other species, with slight variations in the actual relative abundances for all species across the time-points. This variation between time points was particularly apparent in perch, which showed the highest amount of variation amongst time-points varying between 33% and 55% relative abundance.

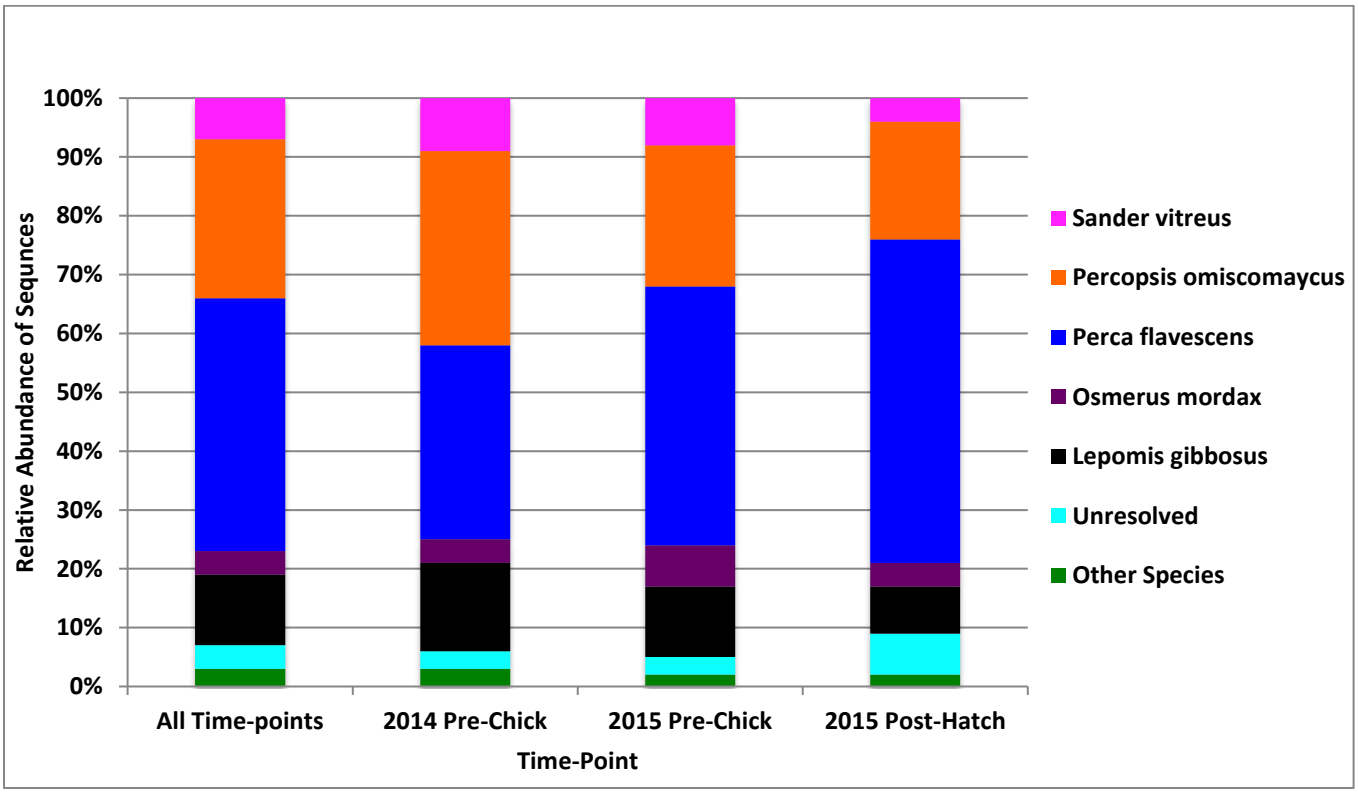


Figure 3.3 Temporal comparison of the relative abundance of sequences from the most abundance fish species obtained from NGS and bioinformatic analysis of DCCO pellets.

Comparison of the yearly and within year (PC and PH) relative abundances shows the general trends across Lake Nipissing. In calculating these trends, we are extrapolating

from four sites to the entire lake, but the four nesting islands from which the cough pellets were collected account for approximately 95% of the overall population (**Figure 1.1**) making this a reasonable extrapolation. The nesting islands are each unique, with different numbers of birds, different geographical locations, and possibly different fish prey availability. Given these differences, we also examined our data grouping the pellets by both nesting location, year, and within year time-period.

Cormorant Diet 2014-2015 – Spatial comparison

We examined location and time point by separating the NGS 16S data and local BLAST/MEGAN results by the respective time-points and locations. A summary of the numbers of pellets and sequences is in Table 3. Relative abundance of sequences is shown in **Figure 3.4**.

Goose Islands:

The relative abundance of the sequences for all 20 pellets collected from Goose Island (GO) 2014 PC time-point were found to be: 38% trout-perch, 32% perch, 9% pumpkinseed, 8% other species (primarily rock bass and logperch), 5% walleye, 5% unresolved, and 3% rainbow smelt. The relative abundance of the sequence identity for all 12 pellets collected from GO 2015 PC time-point were found to be: 51% perch, 29% trout-perch, 11% walleye, 6% pumpkinseed, and 2% unresolved. The relative abundance of fish species for the 12 GO 2015 PH time-point pellets was found to be: 51% perch, 24% trout-perch, 14% unresolved, 8% other species, and 3% rainbow smelt. Goose Island

reflected the overall trends in the data, but did show the lowest relative abundance of walleye during the 2015 PH time-point (0%).

Gull Island:

The relative abundance of sequences for all 20 pellets collected from Gull Island (GU) 2014 PC time-point was found to be: 24% perch, 24% pumpkinseed, 16 % trout-perch, 16% other species (primarily logperch and rock bass), 12% walleye, and 8% unresolved. The relative abundance of sequences for al 12 pellets collected from GU 2015 PC time-point was found to be: 31% perch, 26% trout-perch, 17% pumpkinseed, 15% rainbow smelt, 5% other species (primarily rock bass, logperch, and black crappie), 3% walleye, and 3% unresolved. The pellets collected from the GU 2015 PH time-point (12) were analyzed using the same technique. The analysis of the 12 pellets resulted in the following relative abundance of sequences: 70% perch, 11% trout-perch, 10% unresolved, 5% walleye, 3% other species (primarily whitefish), and 1% pumpkinseed. Gull Island showed similar trends to the overall analysis, however the lowest relative abundance of walleye for the 2015 PC time-point was observed on Gull Island (3%).

Wigwam Islands:

The relative abundance of the sequence identity for all 20 pellets collected from Wigwam Islands (WW) 2014 PC time-point was found to be: 40% perch, 39% trout-perch, 8% pumpkinseed, 6% walleye, 4% rainbow smelt, and 3% other species (primarily logperch and muskie). The relative abundance of the sequence identity for all 12 pellets collected

from WW 2015 PC time-point was found to be: 49% perch, 14% rainbow smelt, 13% walleye, 8% pumpkinseed, 8% other species, 6% trout-perch 2% unresolved. The relative abundance of the sequence identity for all 12 pellets collected from WW 2015 PC time-point was found to be: 58% perch, 18% pumpkinseed, 9% trout-perch, 5% walleye, 5% rainbow smelt, 4% unresolved, and 1% other species (primarily rock bass). Data collected from Wigwam Islands showed similar trends to the overall analysis. Wigwam Island also displayed the highest relative abundance for perch for the 2014 PC time-point and 2015 PC time-point at 49% and 58% respectively.

Hardwood Islands:

The relative abundance of sequence identity for all 20 pellets collected from Hardwood Island (HW) 2014 PC time-point was found to be: 30% trout-perch, 30% perch, 18% pumpkinseed, 16% walleye, 4% other species (primarily logperch, muskie, and mudminnow) and 3% rainbow smelt. The relative abundance of the sequence identity for all 12 pellets collected from HW 2015 PC time-point was found to be: 31% trout-perch, 30% perch, 11% pumpkinseed, 9% walleye, 7% rainbow smelt, 7% other species (primarily logperch) 5% unresolved. The relative abundance of sequence identity for all 12 pellets collected from HW 2015 PH time-point was found to be: 22% walleye, 21% unresolved, 18% rainbow-smelt, 16% pumpkinseed, 13% trout-perch, and 10% other species (primarily mudminnow and sculpin). The trends observed during the 2015 PC time-point reflected the overall trends. However, data collected from the 2014 PC, and 2015 PH contained the highest relative abundance of walleye at 16% and 22%,

respectively. The 2015 PH time-point also consisted of the highest level of unresolved sequences (21%) compared to all other time points.

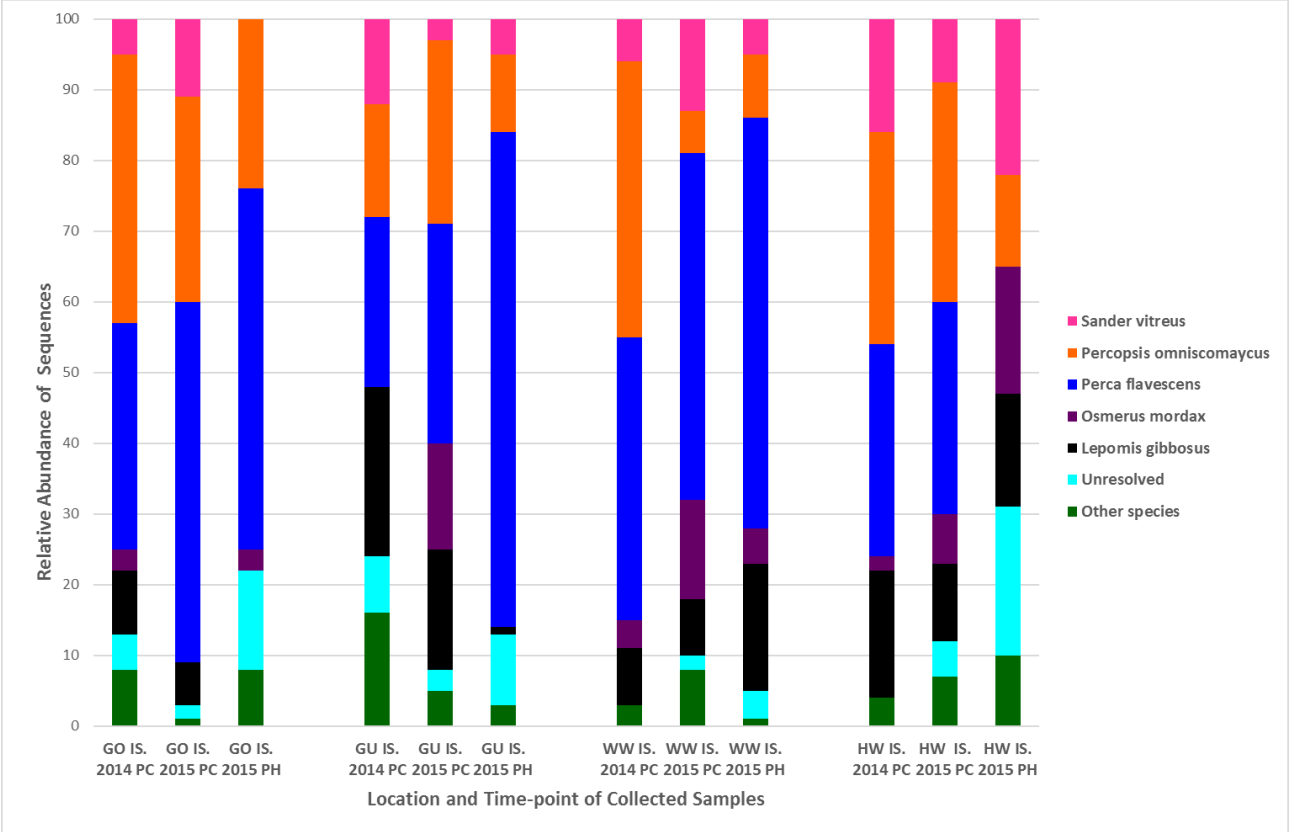


Figure 3.4 Spatial comparison of the relative abundance of sequences from the most abundant fish species obtained from NGS and bioinformatic analysis of DCCO pellets.

<u>Spatial comparison summary</u>

Overall, the time-points and locations had a similar hierarchy of fish abundance for the vast majority of the locations and time-points investigated (1.perch 2. trout-perch 3. pumpkinseed 4.walleye 5. rainbow trout 6.unresolved 7. other species) with relatively limited variation in the exact relative abundance from one time-point to the next for each respective species. This being said, certain time-points/locations deviated from this

general trend, and could warrant further investigation. The HW location is particularly interesting with respect to the deviations from the general hierarchy, as all three time-points investigated had trout-perch as the number one contributor of diet, and not perch like most other locations. In addition to this, the HW 2015 PH time-point not only demonstrated trout-perch as its number one contributor of diet, but also showed elevated frequency of walleye, rainbow smelt and pumpkinseed in the diet. Further study is necessary to determine if these differences in diet with respect to location is consistent over time or is an isolated anomaly. It should also be noted that while we were able to amplify and sequence the 16S gene fragment from all pellets analyzed, there were substantial differences in the average number of sequences per pellet both between pellets, time points, and locations. The difference in average number of sequences per pellet could be related to the degradation of the DNA (older pellets consisted of lower quality DNA). Pellets collected in 2014 had the least amount of variation, with the highest average sequences/pellets being 161,127 (Hardwood Islands PC), and the lowest corresponding to 92,855 average sequences/pellet (Gull Islands PC). There was more variation across the 2015 pellets with the highest average sequences/pellets being 332,734 (Wigwam Islands PH) and the lowest being 33,937 (Gull Island PC). Although the number of pellets analyzed certainly plays a role in the level of variation (20 pellets/location in 2014 versus 12 pellets/location in 2015), it is possible that there was variation in degradation in the sample DNA collected between pellets that could, for example, reflect environmental differences between the years.

The number of sequences recovered from a given environmental sample, gives a good indication of the overall integrity of the DNA that was initially extracted from the sample. All pellets analyzed from 2014 (80 total) resulted in > 20,000 average reads/pellet, however 18 of the total 96 pellets (18.75%) analyzed in 2015 resulted in an average reads/pellet < 20,000. The lowest average reads/pellets collected was from a pellet collected from Gull Island, during the 2015 PH period. The DNA extractions, NGS, BLAST, and MEGAN analysis of this pellet resulted in a total of only 510 sequences. Analysis of this sample, while yielding a very low number of sequences, does still show Perch as the most abundant species (63%), but indicates unresolved sequences as second on the hierarchy of contribution to diet (22%). This percentage of unresolved sequences is unusually high, as the relative abundance of unresolved species usually around four percent. This type of pellet to pellet analysis is important not only to get an idea of which samples may have been subjected to degradation, but can also give an indication into some kinds of sampling biases. Further investigation of the possible causes of DNA degradation may give some insight on how to limit that degradation and reduce the pellet-to-pellet variation.

3.4 Overall cormorant diet 2014-2015 – Distribution by pellet

To further investigate how similar or unique each pellet was we also quantified the amount of variation seen across all the individual pellets. To make this comparison we determined the relative abundance of species in each sample and plotted the contribution of each species to the overall diet (**Figure 3.5**). This plot shows that the most of the species considered to be prey (Yellow perch, Walleye, Rainbow smelt, Rock Bass,

Unresolved, Other species) for the DCCO are found to be contributors of less than 10% of the overall diet in 70% or more of the samples with a few pellets showing much higher relative abundance. In contrast to these species, perch and trout-perch are more evenly distributed throughout all pellets, indicating that the relatively high contribution of these species to the overall diet is a general feature across the entire sample and not driven by a few pellets with very high relative abundance.

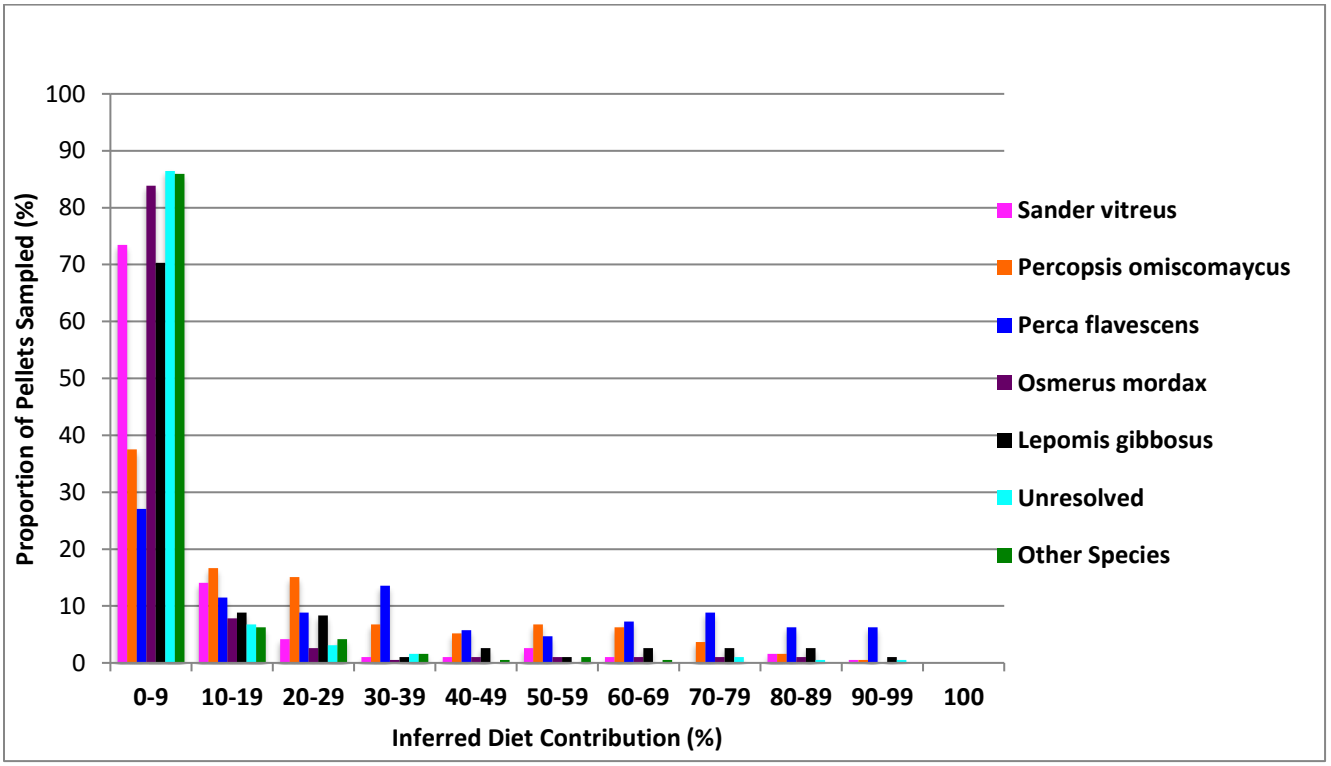


Figure 3.5 Distribution of overall diet among the most abundant fish species identified by NGS and bioinformatic analysis of DCCO pellets.

3.5 Conclusions

Using the novel combination of Next Generation Sequencing of the mitochondrial 16S fragment and local BLAST analysis, described in Chapter 2, we were able to determine the relative abundance of known fish species in the DCCO nesting in LN using a

genetics-based technique as requested by the MNRF. This novel approach shows promise for broad applicability in its ability to infer DCCO diet in an objective, relatively inexpensive, and non-invasive manner. Perhaps most interestingly, the methods we have developed are easily applied to any predatory bird that produces a cough pellets, a list that includes hawks, eagles, herons, gulls and a variety of shorebirds. The DCCO data does suggest slight changes in percentage of diet contribution over locations and time-points, however the exact hierarchy of prey contributors remains quite consistent throughout (1.perch 2.trout-perch 3.pumpkinseed 4. walleye 5. rainbow smelt 6. unresolved 7. other species). It is interesting to note that the perch population has been identified as the most abundant species in LN according to MNRF broad-scale monitoring data reflects the number one prey of DCCO (Appendix 2). In addition to this, trout-perch and walleye are also found in the top 4 most abundant species in LN which is reflected in the top 5 prey species identified in the diet. This said, it should be noted that there our analysis did not identify spottail shiner as a major prey species, even though it is the second most abundant fish species in LN according to MNRF data. Although our analysis does not directly indicate the effect DCCO diet can have on the LN Walleye population, this sort of approach, combined with food-web models, could determine the exact ramifications these migratory birds may have on the walleye stocks (if any) long-term.

Chapter 4 Conclusions and Future Directions

In this last chapter I address some issues for possible further examination and future research that grow out of the work in my thesis but are beyond the scope of the work presented here.

4.1 Potential biases

The use of genetics techniques and bioinformatic analysis is an improvement over previous methods of DCCO diet analysis allowing simpler, accurate, determination of the diet in a non-invasive manner. With this said, there are potential biases that could possibly skew the results I obtained. So-called "primer bias" is one of the main potential biases we encounter with this genetics-based approach.. Such preferential binding could lead to biases in the results (Acinas et al 2005). It is possible this issue is present in our particular data with respect to the *Notropis spp.*, which was the second most abundant fish species (spottail shiner) in the LN basin, however the fish species was rarely found as prey in DCCO diet. It is difficult to say whether or not this is related to primer bias, or the fact that these species are generally smaller and are not considered normal prey size for DCCO. One of the ways to investigate this primer bias would be by preparing a sample containing known proportions (or concentrations) of each individual fish from the basin, and subjecting this sample to NGS analysis. By analysing the sequence data set generated from this "known" sample, one could determine the effect of primer bias, e.g. by asking the question "Does the data set reflect the known proportions?". If primer bias occurs consistently, one could apply a form of correction coefficient in order to get a better representation of the data, e.g. if one species is consistently underrepresented by 15%, its

observed value could be "corrected" by an increase in 15% (and proportional decrease of other species). These types of studies have been done in other animals, where a species was fed a known diet (Deagle et al. 2005). Although the prey species included in the fed diet were all amplified in this study, the relative abundance of the proportion of species was not always reflective of the fed diet, thus it is important to further investigate the possible role primer bias may play in the analysis (Pompanon et al. 2012). Another way to investigate potential primer bias in a 16S primer pair would be by performing the same analysis on a different gene, essentially replicating the study with an independent gene. If the second gene shows similar results, one can reasonably affirm that there is limited, if any, primer bias.

Other possible biases that could be addressed in future work is related to the sample size of the data. The lake-wide diet is inferred from data collected at four separate nesting locations, at two different time points. Although the overall trends seen when comparing years, or seasons together, are very consistent, diet collected from certain time-points (such as Wigwam PH 2015) show clear differences from the overall trends. In order to investigate the possible effect sample size has on the overall diet, additional data should be collected from time-points using a larger subset of samples. This analysis would aid in determining whether certain deviations from the diet are in fact real, and not a result of small sample size bias. These data would also provide information related to between-pellet variation. One of the Islands that would benefit from this analysis is Wigwam Island, which showed a lack of perch in diet during the 2015 PH time-point. By applying

our proposed protocol to a larger set of pellets, one could determine whether the lack of perch in diet at this particular location (Wigwam) is an artefact related to sample size.

Finally, it is important to consider rate of degradation when considering potential biases. Samples were all collected at similar time-points, and fresh pellets were targeted during collection periods. This being said, the exact time between regurgitation and pellet collected is unknown. Some pellets may have been exposed to the elements for an extended period of time, leading to an increased chance of degradation to the available prey DNA in the pellets. This degradation, especially different amounts of degradation between samples, could affect the efficiency of the analysis, as much of the DNA could become too damaged to be amplified during NGS. Further, it is possible that different types of tissues, or different species, could degrade at different rates. "Fragile" fish species could degrade faster than more robust fish species – possibly reflecting differences in scale cover or scale type, for example. If one species is degraded, digested, faster than another it could make up disproportionately less of the pellet. To test this, we would have to feed birds a known diet, collect pellets and repeat the sequencing experiment described above. Such an experiment would be a logistical challenge, but could shed interesting light on the accuracy of this pellet/genetic-based approach.

4.2 Statistical modelling

Our protocol was used to determine DCCO diet on LN however the exact impact these cormorants have on the walleye population of the lake is yet to be determined. These data

sets could be applied to statistical models, which would allow one to estimate the total biomass of walleye DCCO are responsible for removing from the lake every year. Applying statistical modelling is important in determining the effect the cormorants are having on the walleye fish stocks of LN. By estimating the total amount of biomass of walleye being removed by cormorants, one could determine the impact the DCCO population is having on the fish stocks, and whether or not policies should be introduced to maintain walleye sustainability. In addition to the short-term effects the cormorants may have on the lake, modelling could be applied to long-term monitoring of cormorant diet.

Statistical modelling has the potential to determine the exact impact DCCO are having on LN walleye stocks. If it is determined that the birds do not pose a threat to the sustainability of the walleye, it is important that samples be continually collected and analysed in order to assure that no abnormal changes in diet have occurred. This sort of vigilant diet monitoring can be an efficient tool in ecosystem monitoring where any significant variations in DCCO diet could be indicators of other changes occurring in the ecosystem.

4.3 Future/further applications

In addition to cormorants, the novel combination of bioinformatic and molecular techniques described in this thesis could also be applied to other predator-prey systems. The production of cough pellets is not unique to cormorants, other birds such as hawks, owls, and herons also produce cough pellets. Although the abundance of pellets would

not be as evident with these particular species compared to the DCCO, it is possible that diet of these other birds could be monitored.

One of the biggest advantages to the proposed technique is its non-invasiveness. Although the majority of predator species do not produce cough pellets, this technique could possibly be applied to other samples, such as faeces, in order to broaden the possible species this technique could be applied to. Developing this sort of protocol could be important in monitoring diet for species at risk, where invasive methods for diet determination are not an option. One could apply this protocol to other fish feeding birds found on LN, such as the heron. These birds do produce pellets, and feed mainly on fish, thus it would be interesting to see if the diet inferred from heron is similar to that collected from DCCO.

It is clear that the protocol described in this thesis, in addition to being able to determine DCCO diet, is suited for a wide-range of applications. These combinations of data analysis and molecular techniques are becoming an important tool in conservation, and can allow for accurate, objective, and non-invasive analysis of samples.

APPENDIX 1

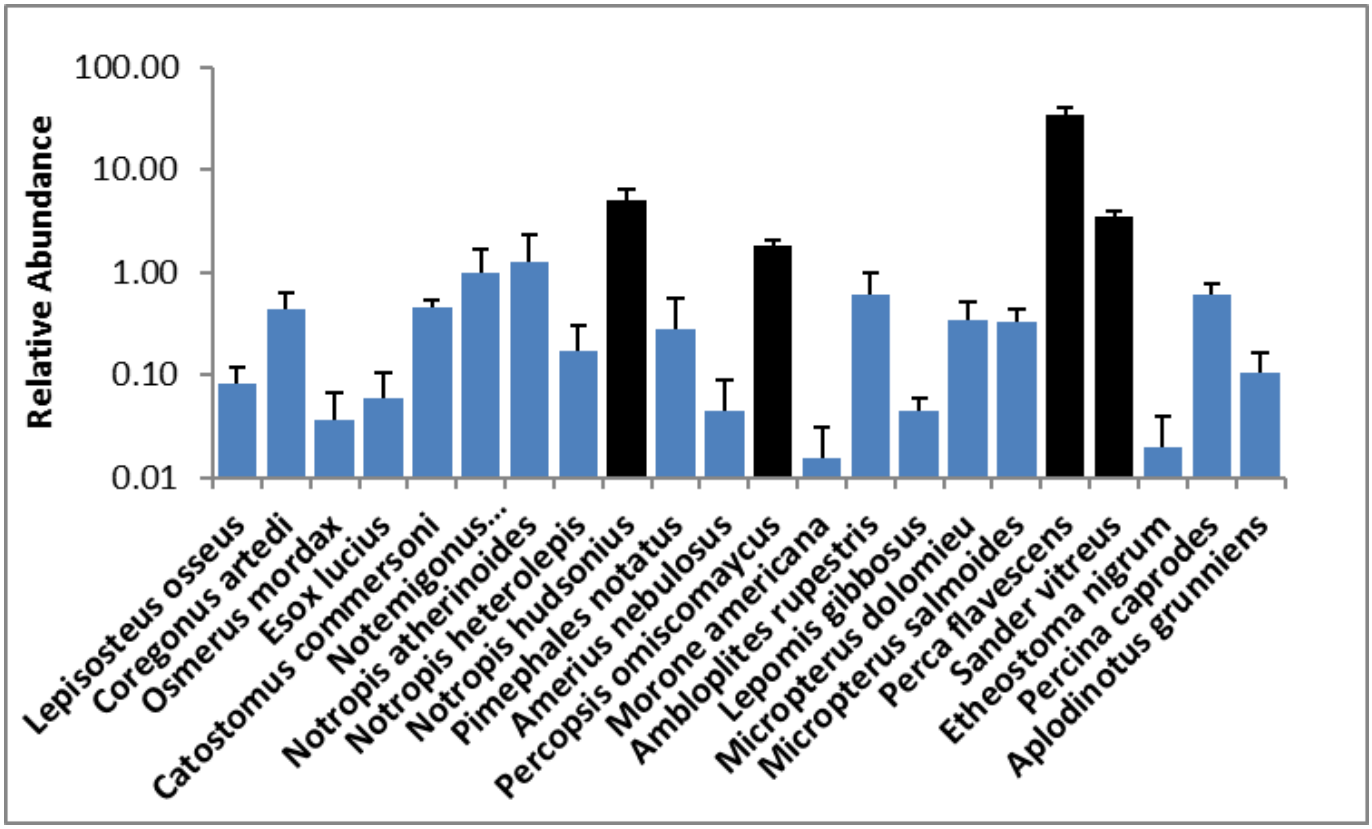
Lake Nipissing Fish Species List (Last updated June 2013)

FISH SPECIES CODE	FISH SPECIES REPORT CODE	FISH COMMON NAME	FISH SCIENTIFIC NAME	COMMENTS
012	NBLam	Northern Brook Lamprey	*Ichthyomyzon fossor*	SAR: Special concern
013	SiLam	Silver Lamprey	*Ichthyomyzon unicuspis*	SAR: Special concern
031	LAStu	Lake Sturgeon	*Acipenser fulvescens*	SAR: Threatened
041	LnGar	Longnose Gar	*Lepisosteus osseus*	
051	Bowfi	Bowfin	*Amia calva*	
080	BkTro	Brook Trout	*Salvelinus fontinalis*	Incidental catches/occasional migrant from Duchesney Creek
091	LaWhi	Lake Whitefish	*Coregonus clupeaformis*	
121	RaSme	Rainbow Smelt	*Osmerus mordax*	Introduced
131	NoPik	Northern Pike	*Esox lucius*	
132	Muske	Muskellunge	*Esox masquinongy*	
141	CeMud	Central Mudminnow	*Umbra limi*	
163	WhSuc	White Sucker	*Catostomus commersoni*	
172	GrRed	Greater Redhorse	*Moxostoma valenciennesi*	
183	FsDac	Finescale Dace	*Chrosomus neogaeus*	
185	LaChu	Lake Chub	*Couesius plumbeus*	
186	CoCar	Common Carp	*Cyprinus carpio*	Introduced – incidental catches on west end of lake and Callander Bay
194	GoShi	Golden Shiner	*Notemigonus crysoleucas*	
196	EmShi	Emerald Shiner	*Notropis atherinoides*	Introduced
198	CoShi	Common Shiner	*Luxilus cornutus*	
200	BnShi	Blacknose Shiner	*Notropis heterolepis*	
201	SpShi	Spottail Shiner	*Notropis hudsonius*	
204	SaShi	Sand Shiner	*Notropis stramineus*	

FISH SPECIES CODE	FISH SPECIES REPORT CODE	FISH COMMON NAME	FISH SCIENTIFIC NAME	COMMENTS
206	MiShi	Mimic Shiner	*Notropis volucellus*	
208	BnMin	Bluntnose Minnow	*Pimephales notatus*	
209	FhMin	Fathead Minnow	*Pimephales promelas*	
211	LnDac	Longnose Dace	*Rhinichthys cataractae*	
212	CrChu	Creek Chub	*Semotilus atromaculatus*	
213	Fallf	Fallfish	*Semotilus corporalis*	
233	BrBul	Brown Bullhead	*Ameiurus nebulosus*	
271	Burbo	Burbot	*Lota lota*	
282	ThSti	Threespine Stickleback	*Gasterosteus aculeatus*	
283	NiSti	Ninespine Stickleback	*Pungitius pungitius*	
291	TrPer	Trout-perch	*Percopsis omiscomaycus*	
302	WhBas	White Bass	*Morone chrysops*	
311	RoBas	Rock Bass	*Ambloplites rupestris*	
313	Pumpk	Pumpkinseed	*Lepomis gibbosus*	
316	SmBas	Smallmouth Bass	*Micropterus dolomieu*	
317	LmBas	Largemouth Bass	*Micropterus salmoides*	
319	BlCra	Black Crappie	*Pomoxis nigromaculatus*	Introduced – incidental catches reported to MNR
331	YePer	Yellow Perch	*Perca flavescens*	
334	Walle	Walleye	*Sander vitreus*	
338	IoDar	Iowa Darter	*Etheostoma exile*	
341	JoDar	Johnny Darter	*Etheostoma nigrum*	
342	Logpe	Logperch	*Percina caprodes*	
346	TeDar	Tessellated Darter	*Etheostoma olmstedi*	
371	FwDru	Freshwater Drum	*Aplodinotus grunniens*	
381	MoScu	Mottled Sculpin	*Cottus bairdii*	

APPENDIX 2

Broad-scale Monitoring Data 2009 and 2016 provided by the MNRF



Appendix 3

List of species and sequences used for local BLAST

NCBI ID	Taxonomic #	Species Name	16S rRNA target	Comments
AF004954	41871	Acipenser fulvescens	Mitochondrial partial sequence	
AY742515	109273	Ambloplites rupestris	Mitochondrial partial sequence	
JX899750	27778	Ameiurus nebulosus	Mitochondrial partial sequence	
NC_004742	7924	Amia calva	Mitochondrial complete genome	
AY520093	225389	Aplodinotus grunniens	Mitochondrial complete sequence	
NC_008647	7971	Catostomus commersonii	Mitochondrial complete genome	
NC_020762	59861	Coregonus clupeaformis	Mitochondrial complete genome	
KP013090	147208	Cottus bairdii	Mitochondrial complete genome	
N/A	147208	Cottus_bairdii	Mitochondrial partial sequence	This research
AF038470	67539	Couesius plumbeus	Mitochondrial complete sequence	
KJ511883	7962	Cyprinus carpio	Mitochondrial complete genome	
AP004103	8010	Esox lucius	Mitochondrial complete genome	
AF262308	126735	Esox masquinongy	Mitochondrial partial sequence	
EF120839	54337	Etheostoma nigrum	Mitochondrial partial sequence	
AP002944	69293	Gasterosteus aculeatus	Mitochondrial complete genome	
NC_025552	245073	Ichthyomyzon fossor	Mitochondrial complete genome	
NC_025553	30308	Ichthyomyzon unicuspis	Mitochondrial complete genome	
JF912024	34771	Lepisosteus osseus	Mitochondrial complete sequence	
NC_028284	270329	Lepomis gibbosus	Mitochondrial complete genome	
KC844053	69944	Lota lota	Mitochondrial complete genome	
N/A	33539	Luxilus cornutus	Mitochondrial partial sequence	This research
AB378750	147949	Micropterus dolomieu	Mitochondrial complete genome	
NC_008106	27706	Micropterus salmoides	Mitochondrial complete genome	
HQ731434	46259	Morone chrysops	Mitochondrial partial sequence	
N/A	154820	Moxostoma carinatum	Mitochondrial partial sequence	This research
N/A	154817	Moxostoma_anisurum	Mitochondrial partial sequence	This research
AB127393	28800	Notemigonus crysoleucas	Mitochondrial complete genome	
AF038486	67550	Notropis atherinoides	Mitochondrial partial sequence	
LU000001	254296	Notropis hudsonius	Mitochondrial partial sequence	
NRU09475	33540	Notropis rubellus	Mitochondrial partial sequence	
DQ536429	28795	Notropis stramineus	Mitochondrial complete genome	
AY216553	28797	Notropis volucellus	Mitochondrial complete sequence	
NC_015246	8014	Osmerus mordax	Mitochondrial complete genome	
NC_019572	8167	Perca flavescens	Mitochondrial complete genome	
EF120840	54317	Percina caprodes	Mitochondrial partial sequence	
AF049741	81381	Percopsis omiscomaycus	Mitochondrial partial sequence	
AY216556	51138	Pimephales notatus	Mitochondrial complete sequence	
KT278765	90988	Pimephales promelas	Mitochondrial complete genome	
KP013112	8182	Pomoxis nigromaculatus	Mitochondrial complete genome	
NC_011571	134920	Pungitius pungitius	Mitochondrial complete genome	

NCBI ID	Taxonomic #	Species Name	16S rRNA target	Comments
EU811082	340988	Rhinichthys cataractae	Mitochondrial partial sequence	
NC_000860	8038	Salvelinus fontinalis	Mitochondrial complete genome	
NC_028285	283036	Sander vitreus	Mitochondrial complete genome	
AF023199	67558	Semotilus atromaculatus	Mitochondrial partial sequence	
NC_028282	75935	Umbra limi	Mitochondrial complete genome	

References

Acinas, S. G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., & Polz, M. F. (2005). PCR-Induced Sequence Artifacts and Bias: Insights from Comparison of Two 16S rRNA Clone Libraries Constructed from the Same Sample . *Applied and Environmental Microbiology, 71*(12), 8966–8969. http://doi.org/10.1128/AEM.71.12.8966-8969.2005

Altschul, S.F., Gish, W., Miller W., Myers, E.W., Lipman, D.J. "Basic local alignment search tool." *Journal of Molecular Biology*. (1990):215(3):403-410.

Asgharian, H., Sahafi, H.H., Ardalan A.A., Shearriz S., Elahi E. "Cytochrome c oxidase subunit 1 barcode data of fish of the Nayband National Park in the Persian Gulf and analysis using meta-data flag several cryptic species." *Molecular Ecology Resources.* (2010):11, 461–472

Bankevich A., Nurk S., Antipov D., Gurevich A., Dvorkin M., Kulikov A. S., Lesin V., Nikolenko S., Pham S., Prjibelski A., Pyshkin A., Sirotkin A., Vyahhi N., Tesler G.,Applications to Single-Cell Sequencing." *Journal of Computational Biology*. (2012):19(5): 455-477. doi:10.1089/cmb.2012.0021.

Barrett, R.T., Camphuysen, C. J., Anker-Nilssen, T., Chardine, J.W., Furness, R.W., Garth, S.,Hüppop, O., Leopold, M.F., Montevecchi, W.A., Veit, R.R. "Diet studies of seabirds: a review and recommendations." *ICES Journal of Marine Science*. (2007):64:1675–1691.

Boyer S[1], Brown SD, Collins RA, Cruickshank RH, Lefort MC, Malumbres-Olarte J, Wratten SD. **"**Sliding window analyses for optimal selection of mini-barcodes, and application to 454-pyrosequencing for specimen identification from degraded DNA.**"** *PLoS One*. (2012):7(5):e38215. doi: 10.1371/journal.pone.0038215.

Boyer, S., Wratten, S.D., Holyoake, A., Abdelkrim, J., Cruickshank, R.H. "Using next-generation sequencing to analyse the diet of a highly endangered land snail (*Powelliphanta augusta*) feeding on endemic earthworms." *PLoS One*. (2013):8(9):e75962. doi: 10.1371/journal.pone.0075962.

Bowser, A.K., Diamond, A.W.,Addison, J.A. "From Puffins to Plankton: A DNA-Based Analysis of a Seabird Food Chain in the Northern Gulf of Maine." *PLoS One*. (2013):8(12): 383152

Brodeur, R. D. "Guide to otoliths of some Northwest Atlantic Fishes." *Massachusetts: National Marine Fisheries Service*. (1979):Available at: http://www.nefsc.noaa.gov/publications/series/whlrd/whlrd7936.pdf

Campbell, L.J. "MNR looks at 10-year recovery for Lake Nipissing Walleye catch limits

remain unchanged, size restriction change" *The North Bay Nugget*. (March 26, 2014). Retrieved from : http://www.northbaynipissing.com/news-story/4431713-mnr-looks-at-10-year-recovery-for-lake-nipissing-walleye-catch-limits-remain-unchanged-size-restric/

Camphuysen, C. J., and Garthe, S. "Recording foraging seabirds at sea: standardized recording and coding of foraging behavior and multi-species foraging associations." *Atlantic Seabirds*. (2004):6:1 – 32.

Candek, K., Kuntner, M. "DNA baracoding gap: reliable species identification over morphological and geographical scales." *Molecular Ecology Resources*. (2015) 15(2): 268-77

Carew, M.E., Pettigrove, V.J., Metzeling, L., Hoffmann, A.A. "Environmental monitoring using next generation sequencing: rapid identification of macroinvertebrate bioindicator species." *Frontiers in Zoology*. (2013):10(1):45. doi: 10.1186/1742-9994-10-45.

Carreon-Martinez, L., Johnson, T.B., Ludsin, S.A., Heath, D.D. "Utilization of stomach content DNA to determine diet diversity in piscivorous fishes." *Journal of Fish Biology*. (2011):78(4):1170-1182. doi: 10.1111/j.1095-8649.2011.02925.x.

Català, S., Pérez-Sierra, A., Abad-Campos, P. "The use of genus-specific amplicon pyrosequencing to assess phytophthora species diversity using eDNA from soil and water in Northern Spain." *PLoS One.* (2015):10(3). doi: 10.1371/journal.pone.0119311

Cawthorn, D.M., Steinman, H.A., Witthuhn, R.C. "Evaluation of the 16S and 12S rRNA genes as universal markers for the identification of commercial fish species in South Africa". *Gene*. (2012) 431(1):40-8.

Chapin, F.S., Carpenter, S.R., Kofinas G.P., Folke, C., Abel, N, Clark W.C., Olsson, P., Smith, D.M., Walker, B., Young, O.R., Berkes, F., Biggs, R., Grove, J.M., Naylor, R.L., Pinkerton, E., Steffen, W., Swanson, F.J. "Ecosystem stewardship: sustainability strategies for a rapidly changing planet." *Trends Ecological Evolution*. (2010):25(4):241-9.

Craven, S.R., Lev, E. "Double-Crested Cormorants in the Apostle Islands, Wisconsin, USA: Population Trends, Food Habits, and Fishery Depredations." *Colonial Waterbirds*. (1987):10(1): 64-71.

Czernik, M., Taberlet, P., Swislocka, M., Czajkowska, M., Duda, N., Ratkiewicz, M. "Fast and efficient DNA-based method for winter diet analysis from stools of three cervids: moose, red deer, and roe deer." *Acta Theriologica*. (2013):58(4):379-386. doi: http://doi.org/10.1007/s13364-013-0146-9

De Barba, M., Miquel, C., Boyer, F., Mercier, C., Rioux, D., Coissac, E., Taberlet, P. "DNA metabarcoding multiplexing and validation of data accuracy for diet assessment:

application to omnivorous diet." *Molecular Ecological Resources*. (2014): 14(2):306-23. doi: 10.1111/1755-0998.12188
Deagle,B.E., Tollit, D.J., Jarman, S.N., Hindell, M.A., Trites, A.W., Gales, N.J. "Molecular scatology as a tool to study diet: analysis of prey DNA in scats from captive Stellar sea lions". *Molecular Ecology.* (2005): 14(6): 1832-42

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

Deagle, B.E., Jarman, S.N., Coissac, E., Pompanon, F., Taberlet, P. "DNA metabarcoding and the cytochrome *c* oxidase subunit I marker: not a perfect match." *Biology Letters*. (2014):10(9)

Deagle, B.E., Gales, N.J., Evans, K., Jarman, S.N., Robinson, S., Trebilco, R., Hindell, M.A. "Studying seabird diet through genetic analysis of faeces: a case study on macaroni penguins (Eudyptes chrysolophus)." *PLoS One*. (2007):2(9):e831

Duffy, D.C., Laurenson, L.J.B. "Pellets of cape cormorants as indicators of diet." *The Condor: Ornithological Applications*. (1982): 85:305-307

Edgar, R.C. "Search and Clustering orders of magnitude faster than BLAST." *Bioinformatics*. (2010):26(19), 2460-2461. Doi:10.1093/bioinformatics/btq461

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., Knight, R. "UCHIME improves sensitivity and speed of chimera detection." *Bioinformatics*. (2011):Doi:10.1093/bioinformatics/brt381

Ekblom, R., Galindo, J. "Applications of next generation sequencing in molecular ecology of non-model organisms." *Heredity.* (2011): 107(1), 1-15

Elbrecht, V., Leese, F., Hajibabaei, M. "Can DNA-Based Ecosystem Assessments Quantify Species Abundance? Testing Primer Bias and Biomass—Sequence Relationships with an Innovative Metabarcoding Protocol." *PLoS One*.(2015):10(7)

Ewing, B., Hillier, L., Wendl, M.C., Green, P. "Base-calling of automated sequencer traces using phred. I. Accuracy assessment." *Genome Research*. (1998):8(3):175-185

Fosso B, Santamaria M, Marzano M, Alonso-Alemany D, Valiente G, Donvito G, Monaco A, Notarangelo P, Pesole G. "BioMaS: a modular pipeline for Bioinformatic analysis of Metagenomic AmpliconS." *BMC Bioinformatics*. (2015):16:203. doi: 10.1186/s12859-015-0595-z.

Fox, G.E., Magrum, L.J., Balch, W.E., Wolfe, R.S., Woese, C.R. "Classification of methanogenic bacteria by 16S ribosomal RNA characterization."

Proceedings of the National Academy of Sciences USA. (1977):74: 4537–454

Geller, J., Meyer, C., Parker, M., Hawk ,H. "Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys." *Molecular Ecology Resources.* (2013):13(5):851-61.

Gharizadeh, B., Herman, Z.S., Eason, R.G., Jejelowo, O., Pourmand, N. "Large-scale pyrosequencing of synthetic DNA: a comparison with results from Sanger dideoxy sequencing." *Electrophoresis*. (2006):27(15):3042-3047

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A., Baird, D.J. "Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos." *PLoS One*. (2011): 6(4). doi: 10.1371/journal.pone.0017497.

Hatch, J. J., and D. V. Weseloh. (1999). Double-crested Cormorant (*Phalacrocorax auritus*). *In* The Birds of North America, No. 441 (A. Poole and F. Gill, eds.). The Birds of North America Online, Ithaca, New York

Hebert, P.D., Cywinska, A., Ball, S.L., deWaard, J.R. "Biological identifications through DNA barcodes." *Proceeding of the Royal Society B: Biological Sciences*. (2003): 270(1512):313-21.

Higes, M., Martín-Hernández,R., Garrido-Bailón, E., Botías, C., García-Palencia, P., Meana, A. "Regurgitated pellets of Merops apiaster as fomites of infective Nosema ceranae (Microsporidia) spores." *Environmental Microbiology*. (2008):10(5):1374-9.

Huson, D., Beler, S., Flade, I., G'orska, A., El-Hadidi, M., Mitras,S., Rusheweyh, H. and Rewati, T.D. "MEGAN Community Edition – Interactive exploration and analysis of large – scale microbiome sequencing data." *PLoS Comput Biol*, (2016):12(6):e1004957

Jarman, S. N., McInnes, J. C., Faux, C., Polanowski, A. M., Marthick, J., Deagle, B. E., Emmerson, L. "Adélie Penguin Population Diet Monitoring by Analysis of Food DNA in Scats." *PLoS ONE*. (2013) :*8*(12), e82227. http://doi.org/10.1371/journal.pone.0082227

Johnson, J.H., Ross, R.M., McCullough, R.D., Mathers, A. "A comparative analysis of double-crested cormorant diets from stomachs and pellets from two Lake Ontario colonies." *Journal of Freshwater Ecology*. (2010):25(4): 669-672

Joshi NA, Fass JN. "Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]." (2011). Available at https://github.com/najoshi/sickle.

Li, F.W., Kuo, L.Y., Rothfels, C.J., Ebihara, A., Chiou, W.L., Windham, M.D., Pryer, K.M. "rbcL and matK earn two thumbs up as the core DNA barcode for ferns." *PLoS One*. (2011):6(10). doi: 10.1371/journal.pone.0026597.

Lobo, J., Costa, P.M., Teixeira, M.A., Ferreira, M.S., Costa, M.H., Costa, F.O. "Enhanced primers for amplification of DNA barcodes froma broad range of marine metazoans." *BMC Ecology*. (2013):13:34. doi: 10.1186/1472-6785-13-34

Luo, A., Zhang, A., Ho, S.Y., Xu, W., Zhang, Y., Shi, W., Cameron, S.L., Zhu, C. "Potential efficacy of mitochondrial genes for animal DNA barcoding: a case study using eutherian mammals." *BMC Genomics*. (2011):12:84. doi: 10.1186/1471-2164-12-84.

Masella, A.P., Bartram, A.K., Truszkowski, J.M., Brown, D.G., and Neufeld, J.D. "PANDAseq: paired-end assembler for illumina sequences." *BMC Bioinformatics*. (2012): 13:31 doi: 10.1186/1471-205-13-31

McGeoch, M.A., Butchart, S.H.M., Spear, D., Marais, E., Kleynhans, E.J., Chanson, J., Hoffman, M. "Global indicators of biological invasion: species numbers, biodiversity impact and policy responses." *Diversity Distribributions*. (2010):16(1): 95–108

McMurdie, P.J. and Holmes, S. "phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data." (2013): http://dx.doi.org/10.1371/journal.pone.0061217

Meusnier, I., Singer, G.A.C., Landry, J-F., Hickey, D.A., Hebert P.D.N., Hajibabaei, M. A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics*. (2010):9: 214.

Mitra, S., Stärk, M., Huson, D.H. "Analysis of 16S rRNA environmental sequences using MEGAN." *BMC Genomics*. (2012):12 (Suppl 3):S17. doi:10.1186/1471-2164-12-S3-S17.

Ontario Ministry of Natural Resources and Fire. "5 Critical Facts about Walleye in Lake Nipissing" (2013) Retrieved from Government of Ontario website : https://dr6j45jk9xcmk.cloudfront.net/documents/2628/stdprod-110492.pdf

Pompanon, F., Deagle, B.E., Symondson, W.O.C., Brown, D.S., Jarman, S.N., Taberlet, P. "Who is eating what: Diet assessment using next generation sequencing." *Molecular Ecology*. (2012):21(8), 1931-19

Ridgeway, M.S. "A Review of Estimates of Daily Energy Expenditure and Food Intake in Cormorants (*Phalacrocorax* spp.). *Journal of Great Lakes Research.*(2010): 36(1):93-99.

Ross, R.M., Johnson, J.H., Adams, C.M. "Use of Fish-Otoloth-Length Regressions to Infer Size of Double-Crested Cormorant Prey Fish from Recovered Otoliths in Lake Ontario." *Northeastern Naturalist*. (2005): 12(2):133-140. http://dx.doi.org/10.1656/1092-6194(2005)012[0133:UOFRTI]2.0.CO;2

Savolainen, V., Cowan, R.S., Vogler, A.P., Roderick, G.K., Lane, R. "Towards writing the encyclopedia of life: an introduction to DNA barcoding." *Philosophical Transactions of the Royal Society of London B: Biological Sciences.* (2005): 29;360(1462):1805-11.

Schirmer, M., Ijaz, U.Z., D'Amore , R., Hall, N., Sloan, W.T., Quince, C. "Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform." *Nucleic Acid Research*.(2015): 43(6): e37.doi: 10.1093/nar/gku1341

Shokralla, S., Hellberg, R.S., Handy, S.M., King, I., Hajibabaei M. "A DNA Mini-Barcoding System for Authentication of Processed Fish Products." *Scientific Reports*. (2015):25:15894

Shrestha RK, Lubinsky B, Bansode VB, Moinz MB, McCormack GP, Travers SA. "QTrim: a novel tool for the quality trimming of sequence reads generated using the Roche/454 sequencing platform." *BMC Bioinformatics*. (2014):12:33. doi: 10.1186/1471-2105-15-33.

Sydeman, W. J., Hobson, K. A., Pyle, P., and McLaren, E. B. "Trophic relationships among seabirds in central California: combined stable isotope and conventional dietary approach." *Condor*. (1997): 99: 327–336.

Taberlet, P., Fumagalli, L. "Owl pellets as a source of DNA for genetic studies of small mammals." *Molecular Ecology*. (1996):5(2):301-5.

Taberlet, P., Prud'Homme, S.M., Campione, E., Roy, J., Miquel, C., Shehzad, W., Gielly, L., Rioux, D., Choler, P., Clément, J.C., Melodelima, C., Pompanon, F., Coissac, E. "Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies." *Molecular Ecology*. (2012):21(8):1816-1820. doi: 10.1111/j.1365-294X.2011.05317.x.

Taguchi, T., Miura, Y., Krueger, D., Sugiura, S. "Utilizing stomach content and faecal DNA analysis techniques to assess the feeding behaviour of largemouth bass Micropterus salmoides and bluegill Lepomis macrochirus." *Journal of Fish Biology.* (2014):84(5):1271-1288. doi: 10.1111/jfb.12341.

Takahashi S, Tomita J, Nishioka K, Hisada T, Nishijima M "Development of a Prokaryotic Universal Primer for Simultaneous Analysis of Bacteria and Archaea Using Next-Generation Sequencing." *PLoS ONE*. (2014):9(8): e105592. doi:10.1371/journal.pone.0105592.

Tollit, D.J., Schulze, A.D., Trites, A.W., Olesiuk, P.F., Crockford, S.J., Gelatt, T.S., Ream, R.R., Miller, K.M. “Development and application of DNA techniques for validating and improving pinniped diet estimates.” *Ecological Applications*.(2009): 19(4):889-905.

Traynor, D., Moerke, A., and Greil, R. “Identification of Michigan fishes using cleithra.” *Great Lakes Fishery Commission. Misc. Publ.* (2010).

van der Valk, A.G. “From Formation to Ecosystem: Tansley's Response to Clements' Climax.” *Journal of the history of biology*. (2014): 47(2): 293-321

Vassou, S.L., Kusuma, G., Parani, M. “DNA barcoding for species identification from dried and powdered plant parts: a case study with authentication of the raw drug market samples of *Sida cordifolia*.” *Gene*. (2015): 559(1):86-93. doi: 10.1016/j.gene.2015.01.025

Wang, J.-F., Jiang, L.-Y., Qiao, G.-X. “Use of a mitochondrial COI sequence to identify species of the subtribe Aphidina (Hemiptera, Aphididae).” *ZooKeys*. (2011):122:1–17.

Waldner, T., Traugott, M. “DNA-based analysis of regurgitates: a noninvasive approach to examine the diet of invertebrate consumers.” *Molecular Ecology Resources*. (2012): 12(4):669-675. doi: 10.1111/j.1755-0998.2012.03135.x -75.

Yang, L., Tan, Z., Wang, D., Xue, L., Guan, M.X., Huang, T., Li, R. “Species identification through mitochondrial rRNA genetic analysis.” *Scientific Reports*. (2014):4:4089. doi: 10.1038/srep04089.

Yao, H., Song, J., Liu, C., Luo, K., Han, J., Li, Y., Pang, X., Xu, H., Zhu, Y., Xiao, P., Chen, S. “Use of ITS2 region as the universal DNA barcode for plants and animals.” *PLoS One*. (2010):5(10). doi: 10.1371/journal.pone.0013102.

Young, G. (2014, March 25). MNR: Catch limits remain at two walleye. *The North Bay Nugget*. Retrieved from http://www.nugget.ca

Zhang, J., & Hanner, R. “Molecular Approach to the Identification of Fish in the South China Sea.” *PLoS ONE*. (2012):*7*(2). doi:http://doi.org/10.1371/journal.pone.0030621