Parentage Analysis of Commercial Lumpfish (*Cyclopterus lumpus*) Brood Stock in Newfoundland.

by

Ashley Skillings

A Thesis

presented to

The University of Guelph

In partial fulfilment of requirements
for the degree of

Master of Science

in

Integrative Biology

Guelph, Ontario, Canada

ABSTRACT

PARENTAGE ANALYSIS OF COMMERICAL LUMPFISH (CYCLOPTERUS LUMPUS)

BROOD STOCK IN NEWFOUNDLAND.

Ashley Skillings
University of Guelph, 2021

Advisor(s):
Dr. Elizabeth Grace Boulding

Lumpfish (*Cyclopterus lumpus*) are widely used in the Northern Atlantic as a biological control for the destructive parasite *Lepeophtheirus salmonis* (salmon louse) in the commercial Atlantic salmon sea cages. Previous studies on lumpfish suggest that there is a genetic component associated with salmon louse consumption. My goal was to create the foundation for a new selective breeding program for sea lice eating efficiency derived from Canadian lumpfish populations. My main objective was to create a panel of 96 SNPs for parentage analysis of hatchery-reared lumpfish founded from Canadian wild-collected lumpfish. The *Colony* software showed that first year of breeding (2015) only produced 4 full-sib families. The *Cervus* parentage analysis showed that inbreeding was present and was therefore unable to determine unique parents among the brood stock. Estimated Breeding Values for natural prey foraging trait were calculated but were not significant due to the small number of available families.

ACKNOWLEDGEMENTS

I would like to thank Dr. Elizabeth Boulding for giving me this opportunity and guiding me throughout this process. I would also like to thank my committee members Dr. Angela Canovas and Dr. Roy Danzmann for their expertise. A special thanks goes out to Jessica Roy for providing the stomach content data, my thesis would not be complete without it. Another special thanks goes out to Isaac Framst for helping with all of those DNA extractions. I would like to thank all the other members of the Boulding Lab including but not limited to Camden, Ben and Meaghan. Thanks to Cooke Aquaculture Inc. for in kind support for NSERC Strategic Project Grant (STPGP 506655 – 17 to E. G. Boulding and L. R. Schaeffer) and for providing the sea cage lumpfish samples. I would also like to thank the Danny Boyce and staff of the Joe Brown Aquatic Research Building at Memorial University for providing the brood stock samples. Lastly, I would like to thank all of my family and friends who supported me throughout this process!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

1 Chapter 1: General Introduction

1.1 Sea lice in salmon sea cages

The salmon louse (*Lepeophtheirus salmonis)* is a destructive parasite that negatively affects the salmon aquaculture industry in Canada (Hayward *et al.,* 2011). The sea lice attach to the gills of the salmon and feeds off the mucus, flesh and blood of the salmon (Hayward *et al.,* 2011). If infection rates are too high, the salmon will die if not treated (Hayward *et al.,* 2011). Emamectin benzoate was the most popular chemical treatment used in the salmon aquaculture industry to control sea lice infections (Igboeli *et al.,* 2012). Emamectin benzoate was an effective method for controlling all life stages of sea lice (Igboeli *et al.,* 2012). Chemicals were distributed either orally with medicated foods containing Emamectin benzoate or topically (chemicals released in the sea cages: e.g., deltamethrin). The Emamectin benzoate that was distributed through the food got ingested by the sea lice and blocked nerve transmission which ultimately killed the sea lice (Igboeli *et al.,* 2012). However, in recent years the sea lice are becoming increasingly resistant to chemical treatments (Imsland *et al.,* 2016). Furthermore, there is more public pressure to stop the use of chemicals in the salmon sea cages to ensure better salmon meat quality (Imsland *et al.,* 2014). Given the chemical resistance and public pressure, there is a need for a more sustainable way to control sea lice infections.

One solution is to use lumpfish (*Cyclopterus lumpus*) as a biological control for the sea lice as the lumpfish clean and eat the sea lice off the salmon (Imsland *et al.,* 2016b). Lumpfish are native to the cold waters of the North Atlantic Ocean, the same

environment as farmed Atlantic salmon (Eliasen *et al.,* 2017). Population genetic studies have shown that there are two distinct subpopulations of lumpfish: North American and European. Previous studies have shown that lumpfish are an effective alternative control for sea lice. Imsland *et al.* (2014) showed that there were approximately 95% less mature female salmon lice found in sea cages that had lumpfish compared to sea cages without lumpfish. However, using lumpfish as a biological control for salmon lice does have certain disadvantages. Lumpfish are opportunistic feeders; therefore, during the zooplankton bloom, lumpfish are more likely to eat the zooplankton rather than the salmon lice (Eliasen *et al.,* 2017). Similarly, larger lumpfish are able to eat the salmon pellets, therefore some lumpfish just eat the salmon pellets instead of forging for the salmon lice (Eliasen *et al.,* 2017, Imsland *et al.,* 2014, Imsland *et al*., 2016b). However, smaller lumpfish are not able to feed on the salmon pellets which make them more likely to forge for salmon lice (Eliasen *et al.,* 2017, Imsland *et al.,* 2014, Imsland *et al*., 2016b). In Canada, salmon producers have shown that if the lumpfish are stocked at the optimal density of 10-15% of lumpfish to salmon, lumpfish can keep the sea lice infestations under control and there is no need for further treatments (Marbase Cleanerfish Ltd., 2020).

The demand for lumpfish has increased significantly in the past few years, exceeding 30 million lumpfish juveniles needed world-wide in 2016 for delousing purposes (Powell *et al.,* 2018). However, to achieve that volume of lumpfish per year sustainably, lumpfish have to be produced at a commercial scale (Powell *et al.,* 2018). In recent years, wild lumpfish populations are declining in Norway and Canada due to

over harvesting spawning stock (Pampoulie *et al.,* 2014). In fact, lumpfish were added to the near-threatened IUCN red list in 2015 (Lorance *et al.,* 2015).

All lumpfish used for the commercial salmon industry in Canada are produced at the Dr. Joe Brown Aquatic Research Building at Memorial University (Newfoundland). The Dr. Brown Aquatic Research Building (JBARB) cultures domestic brood stock and collects wild brood stock each spring to meet the demands of commercial lumpfish (Monk, 2019). The lumpfish production in captivity project started at the JBARB in 2014. In 2019, only 13% of the lumpfish produced at the hatchery was from hatchery-reared lumpfish; the other 87% of the lumpfish were produced from wild females (Monk, 2019).

1.2 Selective Breeding Practices

Traditional selective breeding practices follow the response to selection principal, where the animals with superior performance for a particular trait are selected for breeding, the animals with poor performance for that trait are culled (i.e., not bred). The theory behind selective breeding, is that by breeding the animals with superior performance for multiple generations, the population performance for the particular trait improves after each generation. Selective breeding requires a measurable trait that can be recorded for the candidate breeding stock. The rate of response to the selective breeding given a selection differential is dependent on how heritable the trait is. Heritability in a broad sense is defined as the proportion of the phenotypic variance that is due to genetic variation. The more heritable a trait is the quicker the population will reach the desired superior performance mean for that trait. To determine which animals are superior with respect to a particular trait(s), estimated breeding values (EBVs) for

the selected traits are commonly calculated using the best linear unbiased prediction (BLUP) method (Henderson, 1975; Fernando and Grossman, 1989). EBVs are estimates of an individual's genetic merit for a particular trait or set of traits that could be passed on to its offspring. BLUP is used to calculate the random effects by fitting both fixed and random effects to a mixed model equation. Fixed effects are effects that have few levels, that normally do not change with a different subset of samples, usually individual fall into a category. Random effects are effects that have numerous levels, that change with a different subset of samples, and usually follow normal distribution. For animal breeding, the additive genetic effect is defined as a random effect with a variance that can be contributed to the pedigree or family relationships (Henderson, 1975). Selective breeding and BLUP analysis have been used for selectively breeding livestock and crops for decades and as a result have seen increased production for the selected traits. Prime examples of increased production due to selective breeding in the livestock industry can be seen in the increased meat quantity/bird in broiler chickens from 1957 to 2001(Havenstein et al., 2003).

In theory, the response to selection principle could be used to selectively breed lumpfish for sea lice eating efficiency. Superior lumpfish that eat the most sea lice would be selected for whereas as the inferior lumpfish that eat other readily available food items would be culled (i.e., not bred). After generations of selective breeding, the whole commercial lumpfish population should be more efficient at eating sea lice than the starting population, mirroring a classic response to selection scenario. However, to implement a selective breeding program, individuals and a pedigree needs to be known,

which is where the parentage analysis using Single Nucleotide Polymorphisms (SNPS) is helpful.

1.3 Parentage analysis using Single Nucleotide Polymorphisms

An accurate pedigree of a defined population is essential for any selective breeding program. However, to produce an accurate pedigree, identification of each individual within the breeding population is paramount. For traditional agricultural livestock species tagging individuals at a young age is simple and is a common practice for both breeding companies and farmers (Yue and Xia, 2014). In contrast, identification tagging of individuals in aquatic species such as fish is not feasible given that most fish larvae are minuscule (Yue and Xia, 2014). To solve the issue of unfeasible tagging at a young age, some commercial fish facilities rear different families in separate tanks until the fish can be properly tagged (Vanderputte and Haffray, 2014). However, rearing fish in separate tanks creates different environmental conditions for each family which could falsely enlarge the heritability of a particular trait in question (Vanderputte and Haffray, 2014). Furthermore, the need of separate tanks for each family limits the number of families and crosses that can be made because it depends on the number of tanks available. Lastly, for separate rearing tanks to work, pre-existing family information needs to be known (Vanderputte and Haffray, 2014).

A current method for identifying fish families and ultimately constructing a pedigree is to perform a parentage analysis using DNA markers such as SNPs (Lui and Cordes, 2004, Premachandra *et al.,* 2018). SNPs are variations within the genome at a particular nucleotide where there are least two base variants at the specific nucleotide

site (Neilson *et al.,* 2000). SNPs are abundant throughout the genome of most species; for example, SNPs account for 90% of the genetic variation within the human genome (Brumfield *et al.,* 2003). Most SNPs are stable within the genome and have low mutation rates of 10^{-8}- 10^{-9}, which makes SNPs the perfect DNA marker for parentage analysis (Nielsen *et al.*, 2000). In contrast, microsatellites, the traditional DNA marker for parentage analysis have high mutation rates that can reach 10^{-2} from generation to generation and can have large numbers of null allele outside the population that the microsatellites are designed for (Lui and Cordes, 2004).

Although SNPs are plentiful in the genome, not all SNPs are adequate for parentage analysis. The most powerful SNPs for parentage analysis have minor allele frequencies of around 0.5 (Anderson and Garza, 2006). It is recommended that SNPs with minor allele frequencies greater than 0.2 should be used for parentage analysis to ensure correct parentage assignments while using fewer SNPs (Anderson and Garza, 2006). In fact, Dussault and Boulding (2017) concluded that a minimum of 60 SNPs were needed to achieve 98% accuracy in the studied population when the average minor allele frequency = 0.30. Furthermore, linked SNPs, such as SNPs close together on the same chromosome, should not be used for parentage analysis because most parentage analysis software assumes independent assortment between SNPs (Jones and Ardren, 2003). Likewise, SNPs with null alleles should be avoided because null alleles cannot be genotyped correctly which could cause inaccurate parent assignments. For example, if an individual had a genotype of A/ - (null), the genotype results would show the individual as A/A (Jones and Ardren, 2003). Lastly, biallelic

SNPs (only two different nucleotide alleles) are best used for parentage analysis due to simplicity (Lui and Cordes, 2004).

Once the desired individuals have been genotyped based on the defined SNPs, the actual parentage analysis can be performed. Parentage analyses are based on the genetic concept of Mendelian segregation, that each parent contributes one allele at each locus to the offspring (Yue and Xia, 2014). There are two common parentage analysis methods, the exclusion method and the likelihood method. The exclusion method uses the Mendelian principles of inheritance which makes it the simplest method to compute (Jones and Ardren, 2003). Basically, the exclusion method uses incompatibilities within the SNP alleles to disprove parentage; for example, if the offspring has no alleles in common with the parent in question at a particular SNP locus, then the parent in question can be excluded as a possible parent of the offspring (Jones and Ardren, 2003). Jones and Ardren (2003) argued that the exclusion method does not take into account genotyping error and/ or mutations which can cause false exclusions. In contrast, the likelihood method calculates a likelihood ratio at each SNP locus, then sums all of the likelihood ratios to create the final LOD (likelihood) ratio (Marshall *et al*., 1998). The likelihood ratio is defined as the probability of the parent in question being the true parent of the offspring/ the probability of the parent in question being unrelated to the offspring (Marshall et al., 1998, Jones and Ardren, 2003). The parent(s) in question that have highest LOD ratio are deemed the probable parents of the offspring (Jones and Ardren, 2003). If the LOD ratio is below 1, the parent in question is deemed an improbable parent of the offspring (Jones and Ardren, 2003). Likewise, if the LOD

ratio is equal to zero, the parentage remains unassigned (Jones and Ardren, 2003). The likelihood method takes into account genotyping errors. It is common to assume a genotype error of 0.01 when using SNPs (Morrissey and Wilson, 2005).

The use of SNPs for parentage analysis and genome-wide selection is a common practice in the agricultural industry. SNP genotyping has been widely used in the beef industry for parentage analysis since the early 2000's (Heaton *et al.,* 2002). Molecular parentage analysis for beef cattle is more reliable than traditional identification methods because the DNA can be collected throughout the different stages of production (live yearling to the carcass) which ensures that the correct animal is assigned to the right carcass phenotypes. Likewise, most beef production systems use a multi-sire mating system, therefore SNP parentage analysis is beneficial to accurately determine the superior bulls in the mating system (Heaton *et al.,* 2002). Similarly, parentage testing using SNPs is becoming a common practice in the sheep industry to produce accurate pedigrees in hopes of better genetic improvement (Heaton *et al.,* 2014). Even though the agricultural industry has been using SNPs for parentage analysis successfully for nearly two decades, SNPs are not as readily used in the aquaculture industry but are becoming more readily used in recent years.

Besides parentage analysis, SNPs can be used as markers for QTLs (quantitative trait locus). The use of SNPs for QTLs markers, is a way to predict the phenotype given the genotype the animal in question has. Therefore, instead of waiting for performance data, young animals can be genotyped and their future performance can be predicted by their genotype. Genome-wide selection using SNP markers for

QTLs increases genetic improvement because animals can be genotyped earlier in life which then decreases the generation interval (Berry *et al.,* 2011). SNP QTL markers are also useful for traits that have a low heritability and/or require phenotypic data that is hard to measure (e.g., fertility in dairy cows) (Berry *et al.,* 2011).

1.3.1 Parentage analysis software

Once the samples are genotyped, parentage can then be determined using parentage analysis software. There are two parentage analysis software commonly used by ecologists, which will be used in this study, *Cervus* and *Colony* (Kalinowski et al., 2007; Jones and Wang, 2010).

i) Cervus

Cervus is one of the most widely used parentage analysis software for population genetics (Kalinowski et al., 2007). It uses the likelihood method for parentage analysis to assign the most probable parent to the offspring. This particular computer program assumes that i) the species analyzed is diploid and ii) all the SNPs analyzed are independently assorted autosomal SNPs. The parentage analysis is based on the allele frequencies of the particular SNP and number of potential parents of the offspring. *Cervus* can compute paternity testing, maternity testing, paired parent testing where the sexes of the parents are known and paired parent testing where the parents' sexes are unknown.

Cervus assumes that all candidate parents are equally likely to have produced the offspring in question prior to the analysis. Therefore, the user cannot specify

more likely candidate parents for a particular offspring. Furthermore, *Cervus* only considers the likelihood of parent-offspring relationships, which works well for single bearing species (one offspring per birth). However, *Cervus* has limitations for multi-bearing species (multiple offspring per birth) such as fish because in these populations there are not only parent-offspring relationships but also kinship relationships among the individuals which *Cervus* cannot compute.

ii) Colony

Colony is another parentage analysis software that uses the maximum likelihood method to assign relationships among the individuals (Jones and Wang, 2010). Unlike *Cervus*, *Colony* can compute both parent-offspring relationships and kinship relationships among the population(s) of interest (Wang, 2021*).* In particular, *Colony* can assign both full- sib and half sib relationships, which makes it useful for mating systems of multi-bearing species like fish. It can also be used for parentage analysis for diploid, haplo-diploid species, dioecious and monoecious species. Furthermore, *Colony* can use codominant or dominant markers for the parentage analysis. *Colony* also allows the user to specify the mating system of the population of interest, the user can specify whether the males and females are monogamous or polygamous and if the mating system has inbreeding or no inbreeding. Similarly, unlike *Cervus*, *Colony* allows the user to input known parent-offspring relationships and even known parent kinship relationships.

1.4 Objectives and Significance

The first objective of this study was to develop a panel of 96 SNPs for parentage analysis of commercial lumpfish. If the panel 96 SNPs developed in this study is successful in determining parentage, then it can be used for parentage analysis on a commercial scale at the Joe Brown Aquatic Research Building at Memorial University (JBARB) to aid in their breeding program. The second objective is to use simulations to compare the accuracy of the 96 SNPs for parentage under different parameters. The third objective of this study is to use the 96 SNPs to determine the pedigree (family relationships) of the commercial brood stock population at the JBARB facility. The fourth objective of the study is to use the pedigree defined in objective #2 and lumpfish deployed in the sea cages with stomach content data to determine if sea lice eating efficiency is heritable. If the sea lice eating trait is heritable, estimated breeding values for the sea lice eating efficiency trait can be calculated, which would further aid in selection decisions. This study will lay the foundation for the JBARB facility to start a successful breeding program. Not only would the constructed pedigree be useful to help make better breeding decisions but, the 96 SNP panel developed could be used to genotype future brood stock at the JBARB facility. Furthermore, if sea lice eating is a heritable trait, selective breeding practices could be implemented to develop a strain of lumpfish that are more effective at eating sea lice. Genetically superior sea lice-eating lumpfish, would not only allow for fewer sea lice infections, but would also mean that fewer lumpfish per sea cage would be needed to control infections. Basically, the product of a selective breeding program for sea lice-eating lumpfish would allow for

better quality and more quantity of salmon which, means more profits for producers. Similarly, with less lumpfish needed per cage, there would be a reduction in sea lice treatment expenses, which would also help the bottom line for producers.

2 Chapter 2: Data Chapter

2.1 Abstract

Lumpfish (*Cyclopterus lumpus*) are widely used in Newfoundland as a biological control for the destructive parasite *Lepeophtheirus salmonis* (salmon louse) in the commercial salmon sea cages. Lumpfish are opportunistic feeders; previous studies have shown that there is a genetic component associated with the number of salmon lice lumpfish eats, suggesting that lumpfish could be selectively bred to eat more sea lice. The purpose of this study was to create a panel of 96 SNPs for parentage analysis of commercial lumpfish in Canada and determine the conditions under which the panel was effective. Using the designed 96 SNP panel, the pedigree of the pre-existing commercial lumpfish brood stock produced at Memorial University Newfoundland was determined using *Cervus* and *Colony* software. The final objective was to correlate morphological stomach content data of lumpfish deployed in the salmon sea cages to different families. The second principal component from stomach content abundance data which contrasted natural prey with food pellets, was used as the phenotypic trait; Estimated Breeding Values (EBVs) were then calculated using a linear model. The *Cervus* simulation results showed that the 96 SNPs panel was effective at determining parentage under optimal conditions with no first- or second-degree relatives among the

candidate parents, sexes of parents known, and under low rates of inbreeding. The *Cervus* empirical parentage analysis showed that first-degree relatives were mated and was therefore unable to determine single probable sire-dam-offspring trios among the brood stock. Only 49/231 offspring analyzed could be assigned to one of the four 2015YC families. The *Colony* empirical results showed that first year of breeding (2015) only produced four full-sibling families. EBVs for the natural food forging trait were calculated but were not significant due to the small sample size provided.

2.2 Introduction

The salmon louse (*Lepeophtheirus salmonis*) is one of the most prevalent ectoparasites affecting the salmon aquaculture industry (Hayward *et al.,* 2011). High infection rates of sea lice negatively affect the salmon meat quality and quantity (lbs of meat/ fish) which causes an economical loss for producers. In 2016, Newfoundland and Labrador produced 25,411 tonnes of salmon (valued at $263 million CDN) (Marbase Cleanerfish Ltd., 2020). However, it is estimated that sea lice infections cost around $150 CAD per tonne (Marbase Cleanerfish Ltd., 2020). That means in 2016, roughly $3.8 million CAD was spent on sea lice infection control in Newfoundland, that is more than 1% of gross revenue. Traditionally, sea lice infections were controlled and treated with chemicals. However, sea lice are becoming increasingly resistant to the chemical treatments and there is public pressure to stop the use of chemicals for environmental and health reasons (Imsland et al., 2016). Now the salmon aquaculture industry uses cleaner fish as a biological control for sea lice infections such as the lumpfish

(*Cyclopterus lumpus*). However, lumpfish are opportunistic feeders, therefore they eat whichever prey are abundant. Studies have shown that lumpfish are more likely to eat zooplankton during a zooplankton bloom and larger lumpfish are more likely just to eat the salmon pellets instead of feeding off the sea lice (Eliasen *et al.,* 2017, Imsland *et al.,* 2014, Imsland *et al*., 2016b).

Previous studies have suggested that there is a genetic component to the amount of sea lice a lumpfish eats (Imsland et al.,2016a). If sea lice foraging is a heritable trait, then lumpfish could be selectively bred to be more efficient at delousing salmon. With lumpfish that are more effective at eating sea lice present in the sea cages, there would be less sea lice infections meaning more profits for producers. Furthermore, less lumpfish would be needed per cage which would also reduce the expense of the sea lice treatment.

Selective breeding for biological control has been investigated in other species, primarily in insects. Strauch et al. (2004) studied the potential for selectively breeding nematodes (*Heterorhabditis bacteriophora*) for increased desiccation stress resistance, which are used as a biological control to eat grubs in turf grass. Similarly, Lommen et al (2019) studied the potential of selectively breeding for wingless ladybird beetles, which are used as a biological control for aphids in greenhouses. Besides insects, there have been a few studies that look at selectively breeding livestock to become better agents of biological control, such as domestic goats. Waldron et al. (2009) proposed that goats could be selectively bred to be used as a biological control for juniper plants which disrupts grazing production in pastures.

Selective breeding is not as widely used in the aquaculture industry as it is in the agricultural industry. In the aquaculture industry it was estimated that only 10% of the global aquaculture industry relied on selective breeding programs in 2012 (Gjedrem et al., 2012). However, selectively bred aquaculture species on average have higher annual genetic gains compared to mammal livestock species given their shorter generation interval. Therefore, if selective breeding was implemented in the aquaculture industry, production could increase substantially (Gjedrem et al., 2012). However, selective breeding requires a known pedigree, which is rarely known in aquatic species. The traditional way to determine parentage in aquatic species is to use separate family rearing tanks, so individual tagging is not needed. However, separate family tanks have limitations including variation of environmental conditions between tanks, the number of families are limited to the number of tanks and pre-existing family information needs to be known (Vanderputte and Haffray, 2014).

An increasingly common method to determine parentage and families in aquatic species is to perform a parentage analysis using DNA markers such as SNPs (Lui and Cordes, 2004, Premachandra et al., 2018). Parentage analysis using SNPs has already been successfully implemented in economically important aquatic species such as salmon and yellowtail kingfish (Beacham et al., 2017, Premachandra *et al.*, 2018).

In this study the objectives are to: i) develop a panel of 96 SNPs for parentage analysis of commercial lumpfish, ii) to use simulations to compare the accuracy of the 96 SNPs for parentage under different parameters, iii) to estimate parentage of the lumpfish brood stock at the Dr. Brown Aquatic Research Building at Memorial University

(JBARB) using the 96 SNP panel and iv) to use the estimated pedigree and stomach content data from lumpfish deployed in the salmon cages to determine if sea lice eating efficiency is heritable. If so, EBVs for the sea lice eating efficiency trait will be calculated using a linear model. This study will provide the foundation to start a successful lumpfish breeding program in Canada. This is the first study in Canada to develop a panel of SNPs for parentage analysis for the Canadian lumpfish population and is the first study to discuss selective breeding applications for sea lice eating efficiency in Canadian lumpfish.

2.3 Materials and Methods

2.3.1 Brood stock samples

The brood stock samples that were genotyped came from the Dr. Brown Aquatic Research Building at Memorial University (JBARB). As of 2021, all commercial lumpfish produced in Canada are hatched at the JBARB facility. The brood stock samples were sent to the University of Guelph as fin clips stored in 95% ethanol. The quantity and year class (the year in which the fry hatched from the egg) of the sampled brood stock were as follows: 23 fin-clips from 2015YC, 98 fin-clips from 2016YC and 82 fin-clips from 2017YC (Table 1). The sexes of the brood stock samples were unknown.

2.3.2 Samples from the salmon sea cages

Along with the brood stock there were 84 fin clips from lumpfish that were deployed in the salmon sea cages owned by Cooke Aquaculture Inc. in Olive Cove Newfoundland and caught in 2018 (Table 1). The sampled sea cages were the standard industrial size sea cages in Newfoundland with a circumference of 45 meters and a final

depth of 34 meters. The Olive Cove lumpfish were hatched at the JBARB facility and it is presumed that the 2016 YC brood stock are the parents (but not confirmed). These lumpfish from the salmon sea cages had corresponding stomach samples which were used for the morphological stomach content analysis. The morphological stomach content analysis was performed at the University of Guelph by Jessica Roy, MSc., as part of her Masters' thesis. The stomach contents were categorized into 10 prey classes consisting of i) sea lice, ii) shrimp, iii) krill, iv) ctenophore/jellyfish, v) benthic invertebrates, vi) salmon pellets, vii) lumpfish pellets, viii) caprellid amphipod sp., ix) other amphipods, x) other and unknown. Each prey class was counted and weighed; the prey weights then were used for the principal component analysis (PCA). The PCA was used as the multi-variate statistic to correlate the different semi-quantitative prey class variables. The PCA results done by Jessica Roy, showed that the PC2 scores included lumpfish stomachs that contained krill, caprellid spp., amphipods, benthic invertebrates, and ctenophore/jellyfish that were less likely to contain lumpfish pellets and/or salmon pellets. This indicated that there was variation of diet preferences among the sampled lumpfish. The PC2 scores were then used as the observable trait (y) in the best linear unbiased predictions (BLUP) analysis to calculate EBVs.

2.3.3 Choosing SNPs for parent analysis

To perform the parentage analysis using SNPs as genetic markers, 96 SNPs from the lumpfish genome had to be selected. The lumpfish genome used was from a Norwegian lumpfish at AquaGen (Norway) which contained 5735 scaffolds with a total of 553, 370,415 base pairs

(https://figshare.com/articles/dataset/Lumpfish_Cyclopterus_lumpus_draft_genome_assembly/7301546). In addition, Dr. Boulding's lab (University of Guelph) commissioned 10X coverage whole genome sequencing for 9 North American lumpfish from Trinity Bay, Newfoundland at the Toronto Sick Kid's Hospital. Tim Martin Knutsen, a Research Scientist at AquaGen in Norway, called SNPs from the 9 whole genome sequences (North American lumpfish) and an unknown number of whole genome sequences from Norwegian lumpfish using the Norwegian lumpfish genome file. The final SNP file contained roughly 12 million SNPs across the genome. Both the lumpfish genome and the lumpfish SNP vcf file (Variant Call Format file) were uploaded to Public Galaxy. Galaxy (url: https://usegalaxy.org/) is an open-sourced platform where files can be uploaded to be filtered, converted or analyzed using different pre-programmed functions. First, the vcf distance function in Galaxy was used to add the number of base pairs to the closest variant (another SNP) to all the SNPs within the SNP vcf file. The SNPs were then filtered by the number of base pairs to the closest variant, the closest variant had to be greater than 100 base pairs away so that there were no other SNPs within the flanking region of the candidate SNP. The SNPs were then filtered by both allele frequency and allele count using the vcf filter function. When the SNPs were called at AquaGen, the allele frequency (AF) in the SNP vcf file was calculated with the 9 North American lumpfish and the unknown number of Norwegian lumpfish (AquaGen did not provide the number of Norwegian lumpfish used); whereas the allele count (AC) and the total number of alleles (AN) were only calculated using the 9 North American lumpfish. Therefore, the AF did not equal AC (allele count)/AN (total number of alleles).

The allele frequency (AF) was filtered between 0.35-0.65 and the allele count (AC) for the alternate allele was filtered between 6 and 11, the total number of alleles was 18. To account for any false SNPs within the North American population, the SNPs were further filtered by: North American AF (AC/AN) - AF (combined North American and European AF) had to be between +/-0.1 for the SNP to be a candidate for the parentage analysis. Therefore, the candidate SNPs had to be possible SNPs for both the North American population and the European population. The SNPs were further filtered by quality (QUAL); QUAL had to be greater than 250. SNPs were then randomly chosen from the filtered SNP file (AC, AF, and QUAL) using the random sample function within Galaxy to get a random distribution of SNPs across the genome. Primers (100 bp) were designed for the randomly sampled SNPs using the design flanking sequence function. The designed primers were scanned for Short Tandem Repeats, ranging from mononucleotides repeats to hexa-nucleotide repeats using the STR detection function. The SNPs with primers that contained short tandem repeats were discarded because the short tandem repeats could have interfered with the SNP genotyping and primer binding. After this stage of filtering the possible SNPs for parentage analysis was 124 SNPs with an average allele frequency of 0.466. From the list of 124 SNPs, 96 were chosen so that each SNP was on a unique contig across the genome. The 96 SNPs and primers were then uploaded to the Fluidigm's D3 Assay Design for ordering.

2.3.4 Preparing and genotyping samples

All the lumpfish fin-clip samples were stored in 96% ethanol until DNA extractions were performed. The DNA was extracted from each fin-clip using the 96 well Qiagen

DNeasy Blood and Tissue DNA extraction kit following the defined protocol. The DNA concentrations were then checked using the Qubit machine. The DNA concentration ranged from 11.0 ng/µl- 50.0 ng/µl. The DNA was then diluted to DNA concentrations ranging from 16.5 -31 ng/µl, depending on the sample because the required DNA concentration for genotyping was roughly 16.5 ng/µl. The DNA samples were genotyped using the Fluidigm 96 SNP assay which was performed according to the manufacturer's protocol at the Princess Margaret Genomics Center, Toronto Ontario by Dr. Gurbaksh Basi. The Fluidigm SNP assay uses a 96x96 integrated fluid circuit (IFC) with enough space for 96 samples and 96 SNP primers. The loaded IFC was then placed in the Biomark machine, where 96 PCR reactions were performed consecutively on all 96 samples loaded on the assay (Appendix 1) (Fluidigm 96.96 SNP Type Assay Protocol).

2.3.5 Parentage analysis

2.3.5.1 Cervus

The parentage analysis software *Cervus* 3.0.7 was used to determine the parentage of the 84 sea cage stomach samples from the candidate parents, the brood stock samples. *Cervus* calculates the parentage using LOD scores. The genotype file with all the brood stock samples and the sea cage stomach samples was uploaded to *Cervus*' allele frequency analysis to calculate the allele frequency data file. The allele frequency data file was then uploaded to the simulation of parentage analysis where the sexes of the parent pairs were unknown. The simulation parameters were as follows:

- Offspring simulated: 10,000
- Candidate parents: 20
- Proportion sampled: 0.95
- Proportion loci typed: 0.99
- Proportion loci mistyped: 0.01
- Minimum typed loci: 90

The simulation software calculates the critical value for the LOD scores based on the allele frequencies. The critical value is the threshold where if the candidate parent has an LOD score above the critical value parentage can be assigned at a define confidence level (95%). The genotype file, the allele frequency data file and the simulation file were then uploaded to the parentage analysis where the sexes of the parent pairs were unknown to calculate the parentage analysis for the brood stock samples. There was no previous knowledge of relationships among the brood stock. The following parentage analyses were performed:

i) The 2015YC samples were defined as parents and the 2016YC samples were defined as offspring. This parent analysis was performed to see if there were any parent-offspring relationships among the 2015YC and the 2016YC samples. However, it would be unlikely that the 2015YC were the parents of 2016YC samples because lumpfish take 2 years to mature. That being said, this analysis was performed just to see what kind of results would be produced.

ii) The 2015YC and 2016YC were defined as parents and 2017YC along with sea cage samples were defined as the offspring. This parent analysis was performed to see any parent-offspring relationships between the 2015YC /2016YC and the

2017YC/sea cage samples. The 2015YC and 2016YC were defined as parents because they were the presumed parents of the sea cage samples caught at Olive Cove Newfoundland in 2018. The 2017YC were grouped with the sea cage samples because it was presumed that the sea cage samples were around 1 year of age and since they were caught in 2018 it would make them part of the 2017YC hatch.

2.3.5.1.1 Cervus simulations

Simulations were done in *Cervus* to see how well the chosen 96 SNPs were at assigning parentage under a variety of different parameters and conditions. This was done to determine under which conditions the 96 SNP panel can be used to determine parentage and which conditions the 96 SNP panel is not appropriate for parentage analysis. The first condition tested was small vs large family sizes, small or large family size. The second condition tested was different proportions (0.2,0.5,1.0) of relatives among the candidate parents among both the small and large family. The third condition tested was different inbreeding matings (full sib, half-sib and 1st cousin mating) performed for both the large and small family size. The allele frequency file that was created for the empirical data was used for all the simulations. Simulations were grouped into i) parent pairs (sexes known) and ii) parent pairs (sexes unknown). Within the parent pairs (sexes known), the following simulations were done:

i) Large family size

- Number of offspring: 10,000
- Number of candidate mothers: 50
- Proportion of candidate mothers sampled: 0.95
- Number of candidate fathers: 50

- Proportion of candidate fathers sampled: 0.95
- Number of parent pairs: 2500
- Proportion of loci typed: 0.99
- Proportion of loci mistyped: 0.01
- Error rate in likelihood calculations: 0.01
- Minimum number of typed loci: 0.01

ii) Small family size
- Number of offspring: 500
- Number of candidate mothers: 5
- Proportion of candidate mothers sampled: 0.95
- Number of candidate fathers: 5
- Proportion of candidate fathers sampled: 0.95
- Number of parent pairs: 25
- Proportion of loci typed: 0.99
- Proportion of loci mistyped: 0.01
- Error rate in likelihood calculations: 0.01
- Minimum number of typed loci: 0.01

iii) Large family size with relatedness analysis with 3 different simulations (1) 0.2 or 2) 0.5 or 3) 1.0)
- Same parameters as the i) Large family size
- Proportion of related candidate females: i) 0.2 or ii) 0.5 or iii) 1.0
- Related to: Mother
- Relatedness: 0.5
- Proportion of related candidate males: i) 0.2 or ii) 0.5 or iii) 1.0
- Related to: Father
- Relatedness: 0.5

iv) Large family size with inbreeding with 3 different inbreeding matings (1) full sib mating, 2) half-sib mating, and 3)1st cousins mating
- Same parameters as the i) Large family size
- Relatedness of true parents: i)0.5, ii)0.25, iii)0.125
- Rate of inbreeding in population: 0.021 (2.1% assuming Ne is 24 and rate of inbreeding is equal to 1 / (2 * Ne).)

v) Small family size with relatedness analysis with 3 different simulations (1) 0.2 or 2) 0.5 or 3) 1.0)
- Same parameters as the ii) Small family size
- Proportion of related candidate females: i) 0.2 or ii) 0.5 or iii) 1.0
- Related to: Mother
- Relatedness: 0.5
- Proportion of related candidate males: i) 0.2 or ii) 0.5 or iii) 1.0

- Related to: Father
- Relatedness: 0.5

vi) Small family size with inbreeding with 3 different inbreeding matings (1) full sib mating, 2) half-sib mating, and 3)1st cousins mating
    - Same parameters as the ii) Small family size
    - Relatedness of true parents: i)0.5, ii)0.25, iii)0.125
    - Rate of inbreeding in population: 0.021 (2.1% assuming Ne is 24 and rate of inbreeding is equal to 1 / (2 * Ne).)

Within the parent pairs (sexes unknown) the following simulations were performed:

a) Large family size
    - Number of offspring: 10,000
    - Number of candidate parents: 50
    - Proportion of candidate parents sampled: 0.95
    - Number of parent pairs: 4950
    - Proportion of loci typed: 0.99
    - Proportion of loci midtyped:0.01
    - Error rate in likelihood calculations: 0.01
    - Minimum number of typed loci: 90
b) Small family size
    - Number of offspring: 500
    - Number of candidate parents: 10
    - Proportion of candidate parents sampled: 0.95
    - Number of parent pairs: 45
    - Proportion of loci typed: 0.99
    - Proportion of loci midtyped:0.01
    - Error rate in likelihood calculations: 0.01
    - Minimum number of typed loci: 90
c) Large family size with relatedness analysis with 3 different simulations (1) 0.2 or 2) 0.5 or 3) 1.0)
    - Same parameters as the a) Large family size
    - Proportion of related candidate parents: i) 0.2 or ii) 0.5 or iii) 1.0
    - Related to: Offspring
    - Relatedness: 0.25

d) Large family size with inbreeding
e) Small family size with relatedness analysis with 3 different simulations (1) 0.2 or 2) 0.5 or 3) 1.0)
    - Same parameters as the b) Small family size
    - Proportion of related candidate parents: i) 0.2 or ii) 0.5 or iii) 1.0
    - Related to: Offspring

- Relatedness: 0.25

f) Small family size with inbreeding with 3 different inbreeding matings (1) full sib mating, 2) half-sib mating, and 3)1st cousins mating
    - Same parameters as the b) Small family size
    - Relatedness of true parents: i)0.5, ii)0.25, iii)0.125
    - Rate of inbreeding in population: 0.021 (2.1% assuming Ne is 24 and rate of inbreeding is equal to 1 / (2 * Ne).)

In addition, another set of simulations were done using 200 females and 10 males under different parameters. Two hundred females and twenty males were chosen based on the information from the Environmental Preview Report for Marystown Marbase Cleanerfish Hatchery Project, where it was specified that 200 females and 10 males are needed to produce 2 million lumpfish (Marbase Cleanerfish Ltd., 2020). This set of simulations were used to show the practical application of using the 96 SNPs for parentage analysis at a commercial scale. The following set of simulations were performed:

a) 200 females mated with 10 males
    - Number of offspring: 10,000
    - Number of candidate mothers: 200
    - Proportion of candidate mothers sampled: 0.95
    - Number of candidate fathers: 10
    - Proportion of candidate fathers sampled: 0.95
    - Number of parent pairs: 2000
    - Proportion of loci typed: 0.99
    - Proportion of loci mistyped: 0.01
    - Error rate in likelihood calculations: 0.01
    - Minimum number of typed loci: 0.01

b) 200 females mated with 10 males with relatedness analysis with 0.2 proportion full sib relatedness (2 different simulations: i) 0.2 proportion female only or ii) 0.2 proportion both females and males).
    - Same parameters as the a) 200 females mated with 10 males
    - Proportion of related candidate females: 0.2 (both i) and ii))
    - Related to: Mother

- Relatedness: 0.5 (full siblings)
- Proportion of related candidate males: i) 0 or ii) 0.2
- Related to: Father
- Relatedness: 0.5 (full siblings)

c) 200 females mated with 10 males with relatedness analysis with 0.5 proportion full sib relatedness (2 different simulations: i) 0.5 proportion female only or ii) 0.5 proportion both females and males).
    - Same parameters as the a) 200 females mated with 10 males
    - Proportion of related candidate females: 0.5 (both i) and ii))
    - Related to: Mother
    - Relatedness: 0.5 (full siblings)
    - Proportion of related candidate males: i) 0 or ii) 0.5
    - Related to: Father
    - Relatedness: 0.5 (full siblings)

d) 200 females mated with 10 males with relatedness analysis with 0.2 proportion half sib relatedness (2 different simulations: i) 0.2 proportion female only or ii) 0.2 proportion both females and males).
    - Same parameters as the a) 200 females mated with 10 males
    - Proportion of related candidate females: 0.2 (both i) and ii))
    - Related to: Mother
    - Relatedness: 0.25 (half siblings)
    - Proportion of related candidate males: i) 0 or ii) 0.2
    - Related to: Father
    - Relatedness: 0.25 (half siblings)

e) 200 females mated with 10 males with relatedness analysis with 0.5 proportion half sib relatedness (2 different simulations: i) 0.5 proportion female only or ii) 0.5 proportion both females and males).
    - Same parameters as the a) 200 females mated with 10 males
    - Proportion of related candidate females: 0.5 (both i) and ii))
    - Related to: Mother
    - Relatedness: 0.25 (half siblings)
    - Proportion of related candidate males: i) 0 or ii) 0.5
    - Related to: Father
    - Relatedness: 0.25 (half siblings)

The simulation output contained the summary statistics of the critical LOD score at the defined confidence level, the assignment number of assignments and the

assignment rate for single parent (sexes unknown), parent pair (sexes unknown), father alone (sexes known), mother alone (sexes known) and parent pair (sexes known) depending on the simulation. The simulation output also provides LOD distributions for parent pairs which contained the Mean LOD for True parent pair (both - sexes unknown and known, Mean LOD non-parent pair (true parent pair sampled), Mean LOD non-parent pair (true parent pair sampled), Mean LOD non-parent pair (true mother unsampled, sexes known),Mean LOD non-parent pair (true father unsampled, sexes known), Mean LOD non-parent pair (one true parent, sexes unknown) ,Mean LOD non-parent pair (neither true parent sampled).

2.3.5.2 Colony

Colony, another parentage analysis software was also used in this study. Unlike *Cervus*, *Colony* can compute kinship analysis along with parentage analysis. Kinship analysis was useful for analyzing the relationships among the 2015 YC, 2016 YC and 2017 YC (presumed parents of the sea cage stomach sample) where their parents were not part of the sample size. The parameters used in *Colony* were as followed:

- Mating System I: Female polygamy, Male polygamy
- Mating System II: Without Inbreeding, Without Cloning
- Species: Dioecious, Diploid
- Length of Run: Long
- Analysis Method: Full – Likelihood Method
- Likelihood Precision: High
- Run Specifications: Update Allele Frequency: No, Sibship Scaling: No, Number of Runs: 3, Random Seed Number: 1234 (default)
- Sibship Prior: No Prior

For this study *Colony* was used to find any kinship relationships among the different year classes. The following kinship analyses were performed within *Colony*:

i) The 2015YC and 2016YC were defined as parents and the 2017YC/ sea cage samples were defined as the offspring.
ii) The 2015YC, 2016YC, 2017YC and the sea cage samples were all defined as offspring.

The probability threshold for assignment was 0.85, therefore any pair (offspring-1 and offspring-2) that had a probability greater than 0.85 was deemed siblings and any pair with a probability less than 0.85 was deemed non-siblings.

2.3.6 Calculating Estimated Breeding Values for the natural food forging trait

To calculate Estimated Breeding Values (EBV) for the natural food forging trait (i.e., eating natural prey items instead of pellets measured by PC2), a linear model of the fixed and random effects was made. The natural food forging trait was used as the trait instead of sea lice eating efficiency because the morphological stomach content analysis did not show any sea lice present. It is presumed that if the fish eat more natural foods instead of pellets, those fish would also eat more sea lice given sea lice is a natural prey item. The linear model that was created is as follows:

$$y = \mu + c_i + a_j + l_q + s_p + e_{ijqpd} \quad 2.1$$

y = PC2 score (lumpfish that eat more wild prey items compared to salmon/lumpfish pellet). This was determined by the stomach content analysis on the lumpfish samples caught in Olive Cove Newfoundland in 2018.

μ = the population mean of the PCA dimension 2 value

c_i = fixed cage effect, i levels (i = 7, because there was 7 different cages)

a_j = random animal effects, j levels (j = the number of lumpfish)

s_p= fixed sex effect, p levels (p=2, 1 for males, 2 for females)

l_q = fixed covariate length effects, q levels (the standard length of the lumpfish in mm)

e_{ijp} = random residual effect

Expectations and Variances

$E(y) = \mu + c_i$ (Expected phenotype= population mean + cage effects)

$E(a_j) = 0$ (Expected Animal Effect= 0)

$E(e_{ijp}) = 0$ (Expected Error Effect= 0)

$\sigma^2_e = \sigma^2_e$ (error variance)

$\sigma^2_a = \sigma^2_a$ (animal variance)

$\sigma^2_l = \sigma^2_l$ (length variance)

Assumptions and Limitations of the Linear Model

- All of the lumpfish caught in the sea cages are from the same year class (same age)
- All the progeny lumpfish were collected from the same location (Olive Cove Newfoundland)
- All of the lumpfish sampled were healthy individuals
- The depth that the lumpfish were collected had no effect on the lumpfish diet
- Stocking density of the salmon and lumpfish was constant between sea cages
- All sea cages were at the same water temperature
- All sea cages had the same chance of sea lice infection

To calculate the heritability and the EBVs the ASREML-R software was used. The ASREML-R software uses the restricted maximum likelihood (REML) approach to fit the data to the defined linear model. REML assumes a normal distribution for the random effects and does not require a balanced design unlike the traditional one-sided ANOVA.

2.3.6.1 ASREML-R

To use the ASREML-R package, asreml was loaded into a R-Studio session, using the library(asreml) function. The dataset csv file which contained the empirical data was then uploaded to the R-session. The final dataset only contained 50 sea cage stomach samples, that were assigned to a particular family and had PC2 data. The pedigree that was created using the parentage analysis results was also uploaded to the R-session. The pedigree csv file was formatted with the columns as ID, Sire, Dam. The pedigree also used phantom parents because the parents of the 84 sea cage stomach samples were not part of the brood stock samples.

The inverse A matrix was created using the ASREML-R's ainverse() function which calculates the A matrix (i.e. the relationship matrix) from the pedigree file and then inverts the A matrix. The dataset was then fitted to the mixed model equation in ASREML-R using the asreml(fixed = y ~ fixed effects, random=~vm(random effects), data=dataset, na.action = na.method(x="include")). The vm() is known variance structure function which was used because there was a known inverse A matrix from the pedigree . The na.action includes observations with missing data. The empirical linear model was set up in the asreml() function as the following:

asreml(fixed= PC2 scores ~ cage + sex + mean center squared total length ,
random=~vm(finclip.ID, A inverse matrix), data=dataset, na.action =
na.method(x="include")) 2.3

The variances of the random effects and the error effect were computed with the summary(model)$varcomp. If the dataset had more samples then the heritability of the natural food forging trait could be estimated by $\sigma^2_a / \sigma^2_{phenotypic}$. However, with so few samples, heritability could not be accurately estimated. The fixed effects were computed with the summary(model)$ coef.fixed. The random effects were computed with the summary(model)$ coef.random. The random effects calculated were the EBVs for the natural food forging trait for each animal present in the dataset.

2.4 Results

2.4.1 SNPs for parentage analysis

The final 96 SNPs for this study were chosen, most of which followed the defined criteria (see methods) (Appendix 2). The only criterion that the 96 SNPs did not follow was the criterion of alternate allele frequency between 0.35-0.65. The average alternative allele frequency for the 96 empirical SNPs was 0. 49283. However, the average alternate allele frequency ranged from 0.0996- 0.9074. Thirty two of the 96 SNPs were not in Hardy Weinberg equilibrium, while the other 64 SNPs were in Hardy Weinberg equilibrium.

2.4.2 Colony kinship analysis

The *Colony* results when all year classes and sea cage samples were defined as offspring showed that there were some full sib kinship relationships within the different year classes. Within the 2015YC, the *Colony* results showed that there were 4 full-sib

families (hereby referred to as 2015YC family #1, 2015YC family #2, 2015YC family #3 and 2015YC family #4). Family #1 contained 6 individuals, family #2 contained 8 individuals, family #3 contained 4 individuals and family #4 contained 2 individuals (Table 2). Within the 2016YC, the *Colony* results showed that there were 6 full-sib families, where the family sizes ranged from 3 individuals to 25 individuals (Table 3). Within the 2017YC/sea cage samples, the results showed that there were 25 full-sib families, where the family sizes ranged from 2 individuals to 12 individuals (Table 4). The *Colony* results showed that there were no conclusive half-sib relationships among any of the samples. When the 2015YC and 2016YC were defined as parents and the 2017YC /sea cage samples were defined as offspring there were no conclusive parent-offspring relationships.

2.4.3 Cervus simulations

2.4.3.1 sexes of parents are unknown with large family size

When the sexes of the parents are unknown with a large family size, all simulations performed had a pair parent assignment rate of 95% (simulations a), c1-3), d1-3))(Table 5). When relatives among the candidate parents were included the critical LOD score increased as the proportion of the candidate parents were related increased (Figure 1) . The lower the critical LOD score is in value, the more statistical power the SNPs have for parentage. That means as the proportion of relative increase among the parents, the 96SNPs become less effective at assigning correct parents. The critical LOD results show that simulation a) large family size with no relatives had the best parameters to assign parentage because it had the lowest critical LOD score, meaning

that more true parents can be assigned to a given offspring at the defined confidence level (95%). Whereas, simulation c-3 (1.0 relatives) had the worst parameters to assign parentage because it had the highest critical LOD score, meaning that less true parents can be assigned to a given offspring at the defined confidence level (95%). All large family simulations with relatives among the candidate parent (simulations c-1)-c-3)) had similar Parent Pair: True Parent Mean LOD (SD) to simulation a)(no relatives)(Table 5)

However, when relatives among the candidate parent were included the non-parent pair (true parent pair sampled) mean LOD scores were positive and similar in value to the Parent Pair: True Parent Mean LOD scores compared to simulation a) large family (Table 5). The non-parent pair (true parent pair sampled) mean LOD scores shows that as the proportion of relatives increased, it became harder for simulations to distinguish the non-true parents from the true parents given that the non-parent pair (true parent pair sampled) mean LOD scores were similar in value to the Parent Pair: True Parent Mean LOD scores (Table 5).

When inbreeding was included in the simulations (simulations d-1)-d-3)), the critical LOD scores were similar in value to the simulation a) large family size critical LOD score (-7.50)(Figure 1). The critical LOD results show that simulation d-3) 1st cousin mating had the best parameters to assign parentage because it had the lowest critical LOD score, meaning that more true parents can be assigned to a given offspring at the defined confidence level (95%). Whereas, simulation d-1) full sib mating had the worst parameters to assign parentage because it had the highest critical LOD score,

meaning that less true parents can be assigned to a given offspring at the defined confidence level (95%)(Figure 1).

When inbreeding was included, the Parent Pair: True Parent Mean LOD scores were similar in value to that of the simulation a) large family size with no inbreeding. The non-parent pair (true parent pair sampled) mean LOD scores shows that when inbreeding was included, all simulations were able to distinguish the non-true parents from the true parents given the non-parent pair (true parent pair sampled) mean LOD scores were equal to zero.

2.4.3.2 sexes of parents are unknown with small family size

When the sexes of the parents are unknown with a small family size, the pair parent assignment rate ranged from 92% to 97% (simulations b), e-1)- e-3), f-1)- f-3)). When relatives among the candidate parents were included the critical LOD score increased as the proportion of the candidate parents were related increased (Figure 1). The lower the critical LOD score is in value, the more statistical power the SNPs have for parentage. That means as the proportion of relative increase among the parents, the 96SNPs become less effective at assigning correct parents. The critical LOD results show that simulation b) small family size with no relatives had the best parameters to assign parentage because it had the lowest critical LOD score, meaning that more true parents can be assigned to a given offspring at the defined confidence level (95%). Whereas, simulation e-3 (1.0 relatives) had the worst parameters to assign parentage because it had the highest critical LOD score, meaning that less true parents can be assigned to a given offspring at the defined confidence level (95%).

All small family simulations with relatives among the candidate parent (simulations e-1)-e-3)) had similar Parent Pair: True Parent Mean LOD (SD)(Table 5). The non-parent pair (true parent pair sampled) mean LOD scores shows that when there were relatives among the candidate parents, all simulations (e-1)-e-3)) were able to distinguish the non-true parents from the true parents given the non-parent pair (true parent pair sampled) mean LOD scores were equal to zero.

When inbreeding was included in the simulations (simulations f-1)- f-3)), the critical LOD scores were similar in value to the simulation b) small family size critical LOD score (-14.00)(Figure 1). The critical LOD results show that simulation f-3) 1st cousin mating had the best parameters to assign parentage because it had the lowest critical LOD score, meaning that more true parents can be assigned to a given offspring at the defined confidence level (95%). Whereas, simulations f-1) full sib mating, f-2) half sib mating and b) small family size had the worst parameters to assign parentage because they had the highest critical LOD score, meaning that less true parents can be assigned to a given offspring at the defined confidence level (95%).

When inbreeding was included, the Parent Pair: True Parent Mean LOD scores were similar in value to that of the simulation b) small family size with no inbreeding (Table 5). Likewise, all simulations f-1)- f-3) had a Parent Pair: Non-parent pair (true parent pair sampled) Mean LOD score of 0(0), just like simulation b) small family size with no inbreeding (Table 5). The non-parent pair (true parent pair sampled) mean LOD scores shows that when inbreeding was included, all simulations were able to

distinguish the non-true parents from the true parents given the non-parent pair (true parent pair sampled) mean LOD scores were equal to zero.

2.4.3.3 sexes of the parents are known with a large family size

When the sexes of the parents are known with a large family size, the pair parent assignment rate ranged from 92% to 95% (simulations i), iii-1)- iii-3), iv-1)- iv-3)). When relatives among the candidate parents were included the critical LOD score increased as the proportion of the candidate parents were related increased (Figure 2). The lower the critical LOD score is in value, the more statistical power the SNPs have for parentage. That means as the proportion of relative increase among the parents, the 96SNPs become less effective at assigning correct parents. The critical LOD results show that simulation i) large family size with no relatives had the best parameters to assign parentage because it had the lowest critical LOD score, meaning that more true parents can be assigned to a given offspring at the defined confidence level (95%). Whereas, simulation iii-3 (1.0 relatives) had the worst parameters to assign parentage because it had the highest critical LOD score, meaning that less true parents can be assigned to a given offspring at the defined confidence level (95%).

All small family simulations with relatives among the candidate parent (simulations iii-1)-iii-3)) had similar Parent Pair: True Parent Mean LOD (Table 6). However, when relatives among the candidate parent were included the non-parent pair (true parent pair sampled) mean LOD scores were positive and similar in value to the Parent Pair: True Parent Mean LOD scores compared to simulation i) large family with no relatives. The non-parent pair (true parent pair sampled) mean LOD scores shows

that as the proportion of relatives increased, it became harder for simulations to distinguish the non-true parents from the true parents given that the non-parent pair (true parent pair sampled) mean LOD scores were similar in value to the Parent Pair: True Parent Mean LOD scores.

When inbreeding was included in the simulations (simulations iv-1)- iv-3)), the critical LOD scores were the same value as the simulation i) large family size with no breeding critical LOD score (-7.50) (Figure 2). The critical LOD score results show that all simulations had the same ability to assign parentage at the defined confidence level (95%), given the same critical LOD scores. When inbreeding was included, the Parent Pair: True Parent Mean LOD scores were similar in value to that of the simulation i) large family size (35.87(5.55)). Likewise, all simulations iv-1)- iv-3) had a Parent Pair: Non-parent pair (true parent pair sampled) Mean LOD score of 0(0), just like simulation i) large family size. The non-parent pair (true parent pair sampled) mean LOD scores shows that when inbreeding was included, all simulations were able to distinguish the non-true parents from the true parents given the non-parent pair (true parent pair sampled) mean LOD scores were equal to zero.

2.4.3.4 sexes of the parents are known with a small family size

When the sexes of the parents are known with a small family size, the pair parent assignment rate ranged from 92% to 97% (simulations ii), v-1)- v-3), vi-1)- vi-3)). When relatives among the candidate parents were included the critical LOD score increased as the proportion of the candidate parents were related increased. The lower the critical LOD score is in value, the more statistical power the SNPs have for parentage. That

means as the proportion of relative increase among the parents, the 96SNPs become less effective at assigning correct parents. The critical LOD results show that simulation ii) small family size with no relatives had the best parameters to assign parentage because it had the lowest critical LOD score, meaning that more true parents can be assigned to a given offspring at the defined confidence level (95%). Whereas, simulation v-3 (1.0 relatives) had the worst parameters to assign parentage because it had the highest critical LOD score, meaning that less true parents can be assigned to a given offspring at the defined confidence level (95%).

All small family simulations with relatives among the candidate parent (simulations v-1)-v-3)) had similar Parent Pair: True Parent Mean LOD (SD). Likewise, simulations v-1) and v-3) had a Parent Pair: Non-parent pair (true parent pair sampled) Mean LOD score of 0(0), which was the same Parent Pair: Non-parent pair (true parent pair sampled) Mean LOD score of simulation ii) small family size. Whereas, simulation v-2) 0.5 relatives had a Parent Pair: Non-parent pair (true parent pair sampled) Mean LOD score of 31.19(0). The non-parent pair (true parent pair sampled) mean LOD scores shows that when relatives among candidates were included, all simulations expect for simulation v-2) 0.5 relatives were able to distinguish the non-true parents from the true parents given the non-parent pair (true parent pair sampled) mean LOD scores were equal to zero. The non-parent pair (true parent pair sampled) mean LOD score in simulation v-2) 0.5 relatives showed that it was hard to distinguish the non-parents from the true parents.

When inbreeding was included in the simulations (simulations vi-1)- vi-3)), the critical LOD scores were similar in value to the simulation ii) small family size with no inbreeding critical LOD score (-24.00)(Figure 2). The critical LOD results show that simulation ii) small family size with no inbreeding had the best parameters to assign parentage because it had the lowest critical LOD score, meaning that more true parents can be assigned to a given offspring at the defined confidence level (95%). Whereas, simulation vi-1) full sib mating and simulation vi-3) 1st cousin mating had the worst parameters to assign parentage because they had the highest critical LOD score, meaning that less true parents can be assigned to a given offspring at the defined confidence level (95%).

When inbreeding was included, the Parent Pair: True Parent Mean LOD scores were similar in value to that of the simulation b) small family size with no inbreeding (Table 6). Likewise, all simulations vi-1)- vi-3) had a Parent Pair: Non-parent pair (true parent pair sampled) Mean LOD score of 0(0), just like simulation ii) small family size. The non-parent pair (true parent pair sampled) mean LOD scores shows that when relatives among candidates were included, all simulations were able to distinguish the non-true parents from the true parents given the non-parent pair (true parent pair sampled) mean LOD scores were equal to zero.

2.4.3.5 200 females and 10 males under different parameters

When the practical application simulations with 200 females and 10 males were performed the pair parent assignment rate ranged from 93%-96%. When a high proportion of full sib relatives among the candidate parents (simulations b-i),b-ii), c-i),c-

ii)) were included the critical LOD score increased as the proportion of the candidate parents were related increased (Table 7). The lower the critical LOD score is in value, the more statistical power the SNPs have for parentage. That means as the proportion of relative increase among the parents, the 96SNPs become less effective at assigning correct parents. The critical LOD results show that simulation a) 200 females with 10 males had the best parameters to assign parentage because it had the lowest critical LOD score, meaning that more true parents can be assigned to a given offspring at the defined confidence level (95%) (Table 6). Whereas, simulation c-ii (0.5 of females are full sib relatives and 0.5 males are full sib relatives) had the worst parameters to assign parentage because it had the highest critical LOD score, meaning that less true parents can be assigned to a given offspring at the defined confidence level (95%). All simulations with full sib relatives among the candidate parent (simulations b-i), b-ii), c-i),c-ii)) had similar Parent Pair: True Parent Mean LOD (SD) compared to simulation a) (Table 7).

Likewise, when a proportion of full sib relatives among the candidate parents (simulations b-i),b-ii), c-i),c-ii)) were included, the Mean LOD Non-parent pair (true parent sampled) score increased as the proportion of the candidate parents were related increased (Table 7). The non-parent pair (true parent pair sampled) mean LOD scores shows that as the proportion of full sib relatives increased, it became harder for simulations to distinguish the non-true parents from the true parents given that the non-parent pair (true parent pair sampled) mean LOD scores were similar in value to the Parent Pair: True Parent Mean LOD scores (Table 7).

When half sib relatives among the candidate parents (simulations d-i),d-ii), e-i),e-ii)) were included the critical LOD score increased as the proportion of the candidate parents were related increased. However, the increase of the critical LOD scores among the half sib simulations was not as extreme as the increase of the critical LOD score among the full sib simulations (simulations b-i), b-ii), c-i), c-ii)) (Figure 3). The lower the critical LOD score is in value, the more statistical power the SNPs have for parentage. That means as the proportion of relative increase among the parents, the 96SNPs become less effective at assigning correct parents. The critical LOD results show that simulation a) 200 females with 10 males had the best parameters to assign parentage because it had the lowest critical LOD score, meaning that more true parents can be assigned to a given offspring at the defined confidence level (95%). Whereas, simulation e-ii (0.5 of females are half sib relatives and 0.5 males are half sib relatives) had the worst parameters to assign parentage because it had the highest critical LOD score, meaning that less true parents can be assigned to a given offspring at the defined confidence level (95%).

All simulations with half sib relatives among the candidate parent (simulations d-i), d-ii), e-i),e-ii)) had similar Parent Pair: True Parent Mean LOD (SD) compared to simulation a). However, when a high proportion of half sib relatives among the candidate parent were included the non-parent pair (true parent pair sampled) mean LOD scores were positive and similar in value to the Parent Pair: True Parent Mean LOD scores compared to simulation a). The non-parent pair (true parent pair sampled) mean LOD scores shows that as the proportion of half sib relatives increased, it became

harder for simulations to distinguish the non-true parents from the true parents given that the non-parent pair (true parent pair sampled) mean LOD scores were similar in value to the Parent Pair: True Parent Mean LOD scores, expect for simulation d-1).

2.4.4 Cervus parentage analysis

The *Cervus* results were unable to provide two single probable parents (e.g., a dam and a sire) for any candidate offspring at the defined confidence level. This was due to the fact that all candidate offspring had multiple LOD Trio scores that were positive and similar in value. For example, the *Cervus* results for the 2017YCFID001 offspring had 5 top Trio LOD scores of 1.93E+01, 1.86E+01, 1.43E+01, 1.26E+01 and 1.19E+01, respectively. With only a difference of 0.7 between the highest and the second highest Trio LOD score, there is no possible way to confidently conclude that the candidate parents with the Trio LOD score of 1.93E+01 are in fact the true parents of the 2017YCFID001 offspring. Instead of assigning parents to the candidate offspring, families were assigned to the offspring because all parent candidates with a positive Trio LOD scores for a given offspring were part of the same 2015YC family defined from the *Colony* results. When the 2015YC were defined as candidate parents and the 2016YC were defined as offspring candidates, the *Cervus* results showed some parent-offspring relationships. Only 9 2016YC ‘offspring’ candidates were assigned probable parents from the 2015YC. Seven out of the nine 2016YC ‘offspring’ candidates were solely related to probable parents from the 2015YC Family #1, meaning all candidate parents with a positive LOD score were from the 2015YC Family #1 (Figure 4). One 2016YC ‘offspring’ candidate was solely related to probable parents from the 2015YC

Family #2, meaning all candidate parents with a positive LOD score were from the 2015YC Family #2. One of 2016YC 'offspring' candidate was solely related to probable parents from the 2015YC Family #4, meaning all candidate parents with a positive LOD score were from the 2015YC Family #4. None of the 9 2016YC 'offspring' with probable parents were part of the *Colony* 2016YC full sib results.

When 2016YC and 2015YC were assigned as parents and 2017YC and sea cage samples were assigned as offspring, 41 2017YC offspring were assigned parents from 2015YC/2016YC candidate parents (Figure 5). Most of the 2016YC assigned parents were the same 9 2016YC 'offspring' candidates when the 2015YC were defined as candidate parents and the 2016YC were defined as candidate offspring. Twenty eight of the forty-one 2017YC offspring with probable parents were deemed solely related to the 2015YC Family #1 in some capacity, meaning all probable parents with positive LOD scores were from the 2015YC Family #1. Three out of the forty-one 2017YC offspring with probable parents were deemed solely related to the 2015YC Family #2 in some capacity, meaning all probable parents with positive LOD scores were from the 2015YC Family #2. Another 3 out of the 41 2017YC offspring with probable parents were deemed solely related to the 2015YC Family #3 in some capacity, meaning all probable parents with positive LOD scores were from the 2015YC Family #3. The last 7 out of the 41 2017YC offspring with probable parents were not solely related to a particular 2015YC family. In terms of the sea cage samples, *Cervus* did not assign any probable parents. Interestingly, the 2017YC that were assigned as full siblings to the sea cage samples from the *Colony* results, also did not have any

assigned probable parents in the *Cervus* results. Within the 2017YC full sib families defined in *Colony* that had 2017YC individuals with probable parents defined in *Cervus,* most offspring within each 2017YC full sib family were deemed solely related to the same 2015YC family in Cervus. For example, for Colony full sib Family #2, Cervus showed probable parents for all offspring in that family expect for offspring 2017FID053. However, all probable parents in Cervus were from the 2015YC Family #1. Another example, for Colony full sib Family #3, Cervus showed probable parents for all offspring. Offspring 2017FID007 and 2017FID032 in the 2017YC full sib family did have probable parents but were not deemed solely related to a particular 2015 family. However, all other offspring in that family that had probable parents in Cervus were from the 2015YC Family #1. The fact that the 2017YC full sib family #2 and 2017YC full sib family #3 were both assigned to 2015YC Family #1 in Cervus, suggests that 2017YC full sib family #2 and 2017YC full sib family #3 are related in some capacity, such as half siblings, first cousins, or second cousins. There is currently not enough information to assign a particular relationship to 2017YC full sib family #2 and 2017YC full sib family #3, but analysis showed that they were related.

2.4.5 BLUP analysis for Estimated Breeding Values for the natural food foraging trait

For the ASREML-R analysis, only 50 of the sea cage samples were assigned to a particular family and had PC2 data. Furthermore, phantom parents had to be used for the pedigree because the parents of the sea cage samples were not present in the brood stock samples. The 50 sea cage samples were from 17 different full sib families

(defined in *Colony*) with number of individuals having PC2 data, ranging from 1 to 10. Only four of the 17 families had more than 3 individuals with PC2 data, whereas the other 13 full-sib families had 3 or less individuals per family with PC2 data (Table 8).

The ASREML-R variance component output for random effects showed that the random animal effect had a small variance component of 0.158 with a standard deviation (SD) of 0.335 and the residual (error) effect had a large variance component of 0.952 with a SD of 0.352 (Table 9). The Wald's test revealed that all of the fixed effects were non-significant in the defined model at a 95% confidence level (Table 10).

The random effects solutions (EBVs) gave a solution for all samples in the model including the phantom parents. The offspring SLF055 (belongs to family #1) had the lowest EBV (Table 11). Likewise, the offspring SLF126 (belongs to family #5) had the highest EBV. The phantom parents for family #5 had the highest EBV among the phantom parents. Similarly, the phantom parents for family #1 had the lowest EBV among the phantom (Table 11).

2.5 Discussion

2.5.1 SNPs for parentage analysis

The final 96 SNPs were not the ideal SNPs for parentage analysis for this population of lumpfish brood stock. In this study the alternative allele frequency ranged from roughly 0.1-0.9, whereas previous studies have shown the SNPs used for parentage analysis should have a MAF greater than 0.2 (Anderson and Garza, 2006). However, the allele frequencies reported in this study were only based on the 287

samples genotyped. Therefore, since a relatively small sample size was analyzed, the allele frequencies reported do not represent the population allele frequencies. Another problematic result was that about a third of the final SNPs were not in Hardy Weinberg equilibrium. Ideally parentage analysis SNPs should be in Hardy Weinberg equilibrium. But in this study, there was a high number of SNPs that were not in Hardy Weinberg equilibrium and if all had been removed from of the analysis, the remaining SNPs would have a low statistical power to determine parentage. Therefore, even the SNPs that were not in Hardy Weinberg equilibrium were kept in the analysis. However, if this study were to be repeated, testing Hardy Weinberg equilibrium as part of the defined SNP criteria would be beneficial.

2.5.2 Colony kinship analysis

The *Colony* results showed only four different 2015YC families. It's unsurprising that there were only four 2015YC full sib families based on only 23 2015YC samples in this study. 2015 was the first year that the JBARB facility captively bred lumpfish. The first year of raising/breeding any wild fish in a captive environment presents a lot of challenges. From getting the water temperature, parameters and flow correct to providing the right diet, all these things need to be correct to keep the lumpfish alive and thriving in a captive environment. There are a lot of variables that need to be accurate and if just one variable deviates, trouble can arise (e.g., fish die). In the first year there is a lot of trial and error to get all the variables optimal for captive fish rearing/breeding, therefore mortality rates are usually high in the first year. It would be expected that the

JBARB facility experienced some struggles and had a high mortality rate in 2015 which result in a brood stock size of 23 lumpfish.

The act of capturing lumpfish from the wild and placing them in an artificial environment is stressful for the fish and not all are going to survive the process. Likewise, by bringing in wild fish to a closed environment, there is always a chance of bringing in disease to the closed environment which could end up killing the majority of the captive lumpfish. The reality is that a lot of things could go wrong while bringing wild lumpfish into a captive environment and probably did go wrong at the JBARB facility in 2015. It could have been just by chance that only 4 families of lumpfish survived the capturing process or were hardy enough to survive the suboptimal conditions at the rearing facility. Similarly, if disease was brought into the rearing facility, then it could be possible that only 4 families were resistant to that disease while the rest of the lumpfish were susceptible and died.

Within the 2016YC families and the 2017YC/ sea cage samples families, there were only full sibling relationships, no half sibling relationships. This suggests that the JBARB facility was only performing single pair mating (e.g., each fish is only mated with one other fish). If the JBARB facility was performing any sort of factorial design mating (e.g., each fish is mated with multiple fish), then it would be expected that there would be some half siblings among the samples analyzed. Unfortunately, the JBARB facility did not keep any records of the matings performed, but the parentage results support that single pair mating was used. Given the protocol for mating lumpfish at the JBARB facility, it would not be unrealistic to conclude that single pair matings were performed.

The females have to be stripped of their eggs and the males have to be killed to collect the milt (so the males can be only used once). So, if there was only one female ready to be stripped at a given time, then one male would be sacrificed for the milt to fertilize the eggs and the rest of the milt would be wasted if no other female was ready to be stripped, resulting in single pair mating.

As stated earlier, *Colony* did not produce any conclusive parent-offspring relationships among any of the 2015YC, 2016YC and 2017YC/sea cage samples. One explanation could be that the dataset was not optimal for *Colony*. *Colony* requires the sexes of the parents to be known, meaning the candidate dams and candidate sires are defined separately. However, the sexes of the parents were unknown in the dataset used, therefore all candidate parents were defined as both the candidate sire and dam which artificially inflated the sample size of the parents

2.5.3 Cervus simulations

The *Cervus*' user manual states that the more powerful the SNPs are at determining parentage the more negative the critical LOD scores are (down to -999.00) and the less powerful the SNPs are at determining parentage, the more positive the critical LOD scores are (up to +999.00) (Marshall,n.d.). The simulation results from both sets of large family size simulations where the sexes of the parents are unknown and where the sexes of the parents are known showed the simulation a) and simulation i) had the lowest critical LOD scores, meaning that the 96 SNPs were the most powerful at parentage assignments when there is just a large family with no relatives among the candidate parents. As the proportion of the full sib relatives increased, so did the critical

LOD score which means that 96 SNPS become less powerful as proportion of relatives among the candidate parents increased. This is an expected result given that the more related individuals are to one another the more similar their genotypes are to one another and the harder it becomes to differentiate the genotypes. In contrast, when inbreeding was set in the population at a rate of 2.08%, the critical LOD score roughly stayed the same compared to the critical LOD score of simulation a) and simulation i). In fact, the critical LOD score actually decreased in simulation c-3). These results mean that when inbreeding was set in the population, the 96 SNPs were just or even more powerful to assign parentage compared to when inbreeding was not present in the population. One reason that inbreeding simulations' critical LOD scores were similar to that of simulation a) and simulation i) was because the rate of inbreeding within the population was small at 2.08%. It would be expected that if the rate of inbreeding within the population was set at a higher rate the critical LOD scores would be higher.

Both sets of the small family size simulations where the sexes of the parents are unknown and known showed that as the proportion of relatives among candidate parents increased, the less powerful the 96 SNPs become at assigning parentage. However, with the small family size simulations the critical LOD scores were smaller in value compared to the same simulations done on the large family size. This could be due to the fact that the small family size simulations were easier to compute since there are less candidate parents that could be the true parents. Just like the large family simulations the critical LOD scores when inbreeding was present in the small family size simulations were similar in value to the critical LOD score of simulation b) and

simulation ii). Again, this is probably due to the low rate of inbreeding (2.08%) defined in the simulations.

When a high proportion relatives were included among candidate parents, the non-pair parent mean LOD (when true parents sampled) became positive and similar in value to the true parent pair mean LOD was a trend seen in both small/large family sizes and sexes of parents unknown and known. This trend can be explained by the fact the more related the candidate parents are to one another the more similar their genotypes become. With the non-parent pair mean LOD becoming more similar in value to the true parent pair mean LOD, the harder it becomes to determine the true parents from the parents' relatives which leads to false parent assignments. The chances that aunts and uncles of the offspring are assigned as the true parents increases as the non-parent pair LOD score increases

The results of the simulations show that the 96 SNPs are acceptable for parentage analysis when there are no relatives among the candidate parents and when there is no inbreeding or the rate of inbreeding is low (i.e., at 2.08%). However, when there are relatives among the candidate parents, the power of the 96 SNPs decreases dramatically especially when there is a proportion of full sibs among the parents.

To combat false parentage assignment due to a high number of relatives among the candidate parents more SNPs must be added to the assay. The Canadian lumpfish breeding program is in its infancy and there are already problems with both relatives among parents and inbreeding. Therefore after a few generations, the 96 SNPs will

have no power to determine the true parents because the genotypes will be too similar at those 96 SNPs. Strucken et al. (2014) concluded that a minimum of 200 SNPs are required for parentage assignment in cattle (Waygu and Hanwoo breeds), especially when there are relatives among the candidate parents. Strucken et al. (2014) showed that when there were half sib or full sibs among the sires, the core ISAG (International Society of Animal Genetics) panel (100 SNPs) had a false-positive assignment rate of 39.96%; whereas the breed-specific panel of 257 SNPs false-positive assignment rate of 0%.

The more SNPs that are added to the parentage analysis, the more expensive the panel becomes. Therefore, for the SNP panel to be cost effective, the minimum number of SNPs that can accurately assign true parentage needs to be use for the parentage analysis. In this study, the SNP genotyping (96 SNPs) cost per sample was roughly $15-$20 CAD. If 192 SNPs were used instead of 96 SNPs while still using the same Fluidigm platform, the cost per sample would roughly double ($30-$40 per sample). Obviously, the price per sample would decrease the higher quantity of samples genotyped at once, but these numbers are a rough estimate of costs. If this study was done again with the same number of SNPs but with a larger sample size of 1500, the genotyping costs would be roughly $22,500-$30,000 CAD. The question then arises what is the gain: cost ratio of genotyping. For example, Raoul et al. (2016) concluded that for parentage genotyping to be profitable (i.e., gain is larger than the cost) in sheep, the cost of SNP genotyping per sample needs to be less than €6 (~ $9 CAD). Sheep are different than lumpfish where the breeder knows the dam of offspring

and has some idea of the sire (e.g., one of the males in the group that covered the group of females), but there is still some type of gain: cost ratio for genotyping the lumpfish brood stock that could be calculated.

To limit the number of SNPs needed for the parentage analysis of lumpfish, mating records should be kept (e.g., male #1 was mated to female #2). The mating records would be the starting reference for the parentage analysis. For example, if a given offspring at 2 trios with similar high LOD scores (trio one – male#1 x female #2, trio two – male #3 x females #2), but the mating record states that female #2 was only mated to male #1, then it is more plausible that male #1 is the true father. Now, mating records would only help the parentage analysis if single-pair mating was performed or if partial factorial design (e.g., where each male is mated to multiple females but not all females) was performed. Currently, the JBARB facility is using single pair mating, so mating records would be helpful to determine true parentage. Furthermore, in this study the sexes of the candidate parents were unknown. If the sexes of the candidate parents were known, the number of probable trios would be less because when the sexes are unknown, *Cervus* tests each candidate parent to all other candidate parents. So, in reality, when the sexes are unknown the *Cervus* results might show a probable trio that are two males or two females, which are not a plausible trio but it cannot be disregarded because the sexes are unknown.

2.5.4 Cervus parentage analysis

It was not expected to have any 2016YC ‘offspring’ with probable parents from the 2015YC. As stated earlier, lumpfish take about two years to become sexually mature (Marbase Cleanerfish Ltd., 2020). Therefore, it is very unlikely that the probable parents assigned to those 9 2016YC ‘offspring’ are the true parents because the 2015YC would only be a year old when the 2016YC was hatched. However, given the inconclusive family structure, this analysis was done just to see what results were produced. One explanation could be that those fish were actually 2014YC but were mislabeled or the facility reused some of the previous year’s females and relabeled the ID number, so in fact those females were actually 2 or more years of age. The 2015YC brood stock also could have been wild lumpfish caught in 2015 not hatched in 2015, given that 2015 was the first year of captively breeding lumpfish. There are a lot of possibilities to explain what could have happened to produce these results, but it is improbable that a lumpfish hatched in 2015 was a parent to an offspring hatched in 2016. However, since the *Cervus* results did assign probable 2015YC parents to 9 of the 2016YC ‘offspring’, I concluded that those 2016YC offspring were still related to their assigned 2015YC probable parents in some capacity just not in a parent-offspring relationship.

Having only 41 2017YC/sea cage samples with assigned probable 2015YC/2016YC parents when there were 142 2017YC/sea cage samples analyzed, was unexpected. The JBARB facility sent all of the fin-clips from brood-stock that they had records of. Therefore, the probability of having all candidate parents should have

been high and the majority of the 2017YC/sea cage samples should have been assigned probable parents from the 2015YC/2016YC. The fact that none of the sea cage samples and their 2017YC full siblings (from the *Colony* results) had any assigned probable parents suggests that not all of the brood stock was included in the analysis. These results show that there are brood stock lumpfish that were used in the mating scheme that are still not accounted for. Keeping complete and detailed records of brood stock in a breeding program is paramount. If proper IDs and records are not kept, the superior lumpfish cannot be differentiated from the inferior lumpfish and appropriate strategic matings (i.e., selectively mating superior lumpfish) cannot be made because individual cannot be determined in terms of their genetic merit and family structure.

The result of 28/41 of the 2017YC offspring and 7/9 2016YC 'offspring' were solely related to 2015YC family #1 suggests that there is over representation of the 2015 family #1 genetics within the brood stock. If only a sub-sample of the brood stock was sampled then this could be just a coincidence, but in this study all the brood stock was supposed to be analyzed. It could have been possible that 2015YC Family #1 had the most successful matings (i.e., the highest fertilization rate) which would cause an over representation of their genetics in future generations. Likewise, the offspring from the 2015YC Family #1 matings could have been the hardiest and more survived to adulthood compared to the other 2015YC families' offspring. There are a lot of possibilities that could explain why there was an over representation of the 2015YC Family #1 in the brood stock.

The results of the 41 2017YC samples that were assigned 2015YC/2016YC probable parents with most having all probable parents from the same 2015YC family defined in *Colony*, is unacceptable for a sustainable breeding program. The reason that *Cervus* was unable to provide two single probable true parents (e.g., a dam and a sire) for any 2017YC offspring was because all candidate parents were related to one another in some capacity. For instance, all the assigned probable parents for those 41 2017YC offspring were 2015YC from one of the 4 - 2015YC full sib families (defined in *Colony*) or the 9 2016YC 'offspring' that were also part of the one of the 2015YC families defined in *Colony*. The more related individuals become, the more similar the genotypes become and true parents cannot be differentiated from their relatives.

To make matters worse, not only were there relatives among the candidate parents but the majority of the 41 2017YC offspring had all assigned probable parents from the same 2015YC family which suggests inbreeding (i.e., the mating between relatives) occurred. It is not surprising that relatives were mated given the lack of records kept or the lost records at the JBARB facility. If separate family tanks were not used then there would have been no way to know which lumpfish were related and which ones were not. What most likely happened was that all the lumpfish were kept in the same rearing tank and once a female was sexually mature and ready to be stripped of her eggs, the facility used whichever male was also sexually mature to fertilize the eggs. Fertilizing lumpfish eggs is a time sensitive process. Once the eggs are stripped, they need to be fertilized in a timely matter before they become unviable. Therefore, the facility probably just chose the most readily available male to collect milt from. Traits

that are associated with sexual maturity have low-moderately heritability depending on the species and study. For example, Rosa et al. (2018) reported a heritability of 0.34 for age of sexual maturity in leghorn hens (n=1398). Therefore, if a female matured early compared to the other lumpfish, most likely her siblings (e.g., brothers) also matured early, making it more plausible that the facility was mating full-sibs. The same goes for the lumpfish that mature later. The fact that the onset of sexual maturity is heritable and the time sensitive breeding process could explain why relatives were mated.

Again, 2015 was the first year that captive breeding took place, so there was definitely learning curves and struggles that were faced during the breeding process. Breeding lumpfish is not as simple as just putting a male and a female into a tank and hoping for fertilization of the eggs to occur. It is a labor-intensive process that is time sensitive, so there are a lot of things that could go wrong throughout the process. To prevent mating highly related individuals together, accurate individual ID records, family records and a detailed mating plan needs to be kept and maintained.

2.5.5 BLUP analysis for Estimated Breeding Values for the natural food foraging trait

The problem with the ASREML-R results produced was the small sample size. With the small sample size, the heritability and the EBVs of the natural food foraging trait could not be accurately calculated because there was just not enough data for statistical significance (Falconer and Mackay, 1996). EBVs are estimates of an individual's true breeding values based on phenotypes/performance records of particular traits and the pedigree data. Therefore, with more phenotype records an

animal has and/or the more individuals with phenotypic data within a family, the more accurate the estimated breeding value becomes (e.g., the closer the estimated breeding values becomes to the true breeding value).

As stated earlier, family #5 phantom parents had the highest EBV for the natural food forging trait and the family #1 phantom parents had the lowest EBV. However, family #5 and family #1 were two of the families that had the most individuals per family, 10 and 6 individuals respectively. It cannot be determined from this data set whether family #5 actually had the highest EBVs because they have the best genetics for the natural food forging trait or whether family #5 just have the highest EBVs because family #5 had the most individuals and phenotypic data. Likewise, it cannot be determined whether family #1 actually had the lowest EBVs because they do not have the genetics for the natural food forging trait or whether family #1 have the lowest EBVs because family #1 had more individuals and phenotypic data. Similarly, offspring SLF064 had the second highest EBV (0.350908205), but offspring SLF064 was part of family #17 which only had one individual. Again, it cannot be determined whether SLF064 had good genes for natural food forging trait or was it a fluke giving that there was only one observation for family #17. With such a small sample size, no conclusions can be made about which lumpfish individuals and/or families are better genetically predisposed to eating/forging for natural food prey items. Therefore, these EBVs produced cannot be used in the selection process for the Canadian lumpfish breeding program.

To calculate accurate and usable EBV's to select for lumpfish that forge for natural food prey items, a larger number of samples would need to be used. Instead of using family sizes of 1-10 individual per family, a more acceptable family size would be a minimum of 100 individuals per family. Another way to make the EBVs more accurate would be to collect repeated observations on each individual at different time periods. However, currently collecting multiple records per individual is not feasible because to collect stomach content data, each lumpfish must be euthanized so that its stomach can be dissected.

If this study would have had the appropriate sample size for BLUP, heritability of the natural food forging trait could be estimated. Forging for natural foods would be considered a behavioral trait, which usually have a low heritability between 0.1- 0.2 (e.g., phenotype is more dependent on environment rather than genotype). For example, *Chervet et al.* (2011) reported a heritability of 0.15 for different personality traits in cichlids. Similarly, *Waldron et al.* (2009) reported a heritability of 0.13 for juniper consumption/forging in goats. Given these reported heritabilities for similar traits, if this study were to be repeated it could be hypothesized that the heritability for the natural food forging trait would be around 0.1-0.2.

2.6 Conclusion

This study shows the application of SNP genotyping within the only current commercial lumpfish population in Canada. Even though, this study did not produce significant results in terms of a complete pedigree of brood stock and significant EBVs

for the natural food forging trait, it still outlines the principals and methods needed for parentage and BLUP analysis. If this study were to be performed again, more SNPs would be required to assign accurate parentage given the current brood stock relationship structure where there are a lot of relatives among the candidate parents and inbreeding is present. Furthermore, a larger sample size of families with more offspring per family would be needed to produce accurate EBVs of the natural food forging trait. Likewise, if the stomach content analysis was performed on lumpfish during a sea lice infection period, the amount of sea lice eaten could be used as the observable trait instead of PC2 scores (e.g., lumpfish that eat natural prey items instead of salmon/lumpfish pellets). Another important thing needed is, complete and accurate records of individual lumpfish or at least families of lumpfish within the brood stock population and complete records of the matings performed.

The next step in this research would be to get a record keeping system put in place for the lumpfish brood stock at the JBARB facility. Also, more research is needed to develop another SNP panel that includes ~ 200 SNPs instead of 96 SNPs. Even if the Canadian commercial lumpfish population is not selectively bred for sea lice eating efficiency, SNP genotyping is still a way to manage inbreeding within the population. Once a larger panel of SNPs is produced, more lumpfish deployed in the salmon sea cages can be genotyped and stomach content analysis can be performed. With a large enough sample size, heritability of the sea lice eating efficiency trait could be estimated which would be useful to predict the genetic gain of the lumpfish population over a certain amount of time. Similarly, a genomic-wide association study (GWAS) could be

performed on the commercial lumpfish population to see if there are any SNPs associated with genetic variation within the sea lice eating efficiency trait. That way if there were any associated SNPs, the genotypic data could be combined with phenotypic data to produce more accurate EBVs.

3 Chapter 3: Practical Implications for Industry

The following discusses what the JBARB facility is currently doing for their breeding program and proposed future improvements to achieve the maximum gain for the sea lice eating efficiency trait. This chapter will discuss the importance of starting with a good base population of brood stock, the proposed selection goal(s) and the phenotypes needed to achieve those selection goals, the different mating schemes and lastly the reproductive technologies needed to breed lumpfish in captivity.

3.1 Base populations available

The base population available is the starting point to any breeding program (Falconer and Mackay,1996). The base population is the foundation of the breeding program, therefore if the base population is inferior the breeding program will also be inferior. The base population is the main source of genetic variability for the breeding program. The more genetic variance a population has the greater genetic gain that can be achieved throughout the selection process. Hence, if the base population has little or no genetic variance then little genetic gain can be achieved because there is little genetic difference in the population to select for.

Given that lumpfish are an undomesticated species, all of the base population must come from the wild. Collecting lumpfish from the wild has certain challenges for obtaining superior brood stock because wild lumpfish do not come with impressive pedigrees like other commercial livestock species. Since nothing is known about the traits of the different lumpfish wild populations, it would be wise to collect lumpfish from

numerous wild lumpfish populations in Canada, permits permitting. By collecting from numerous wild lumpfish populations there would be greater genetic diversity within the base population and therefore more selection opportunities. If lumpfish were only collected from one or two wild populations, then there is a higher probability of those lumpfish being genetically similar (e.g., related to one another) and therefore the genetic diversity of the base population is low which would not be a good starting point for a breeding program. Whittaker *et al.* (2018) reported fairly low F_{ST} values among three locations in the West Atlantic (Frenchman Bay, Cobscook Bay and Witless Bay) ranging from 0.013 - 0.03 between these locations, meaning that these three populations in the West Atlantic have very little genetic variance between the subpopulations. Given that the wild lumpfish population in North America already has quite low genetic diversity among the different subpopulations, it is crucial to maximize the available genetic diversity for the breeding program by collecting at as many locations as possible. It would be important to record the collection site of each collected fish, because it could indicate whether a particular subpopulation have lumpfish that naturally eat more sea lice on average than other subpopulations.

In 2016, there was 9 female and 68 male lumpfish brood stock within the JBARB breeding program. However, not all 68 male lumpfish were used for breeding since there was only 9 females and females can only be bred twice per season. Therefore, the maximum number of breeding males in 2016 was 18 males which means that the effective population size (Ne) was ~24 lumpfish, using the equation Ne= 4 *Nf*Nm /Nf+Nm, where Nf is the number of breeding females and Nm is the number of breeding

males (Frankham, 1995). The effective population size is the equivalent population size in an ideal random mating population with an equal number of females and males (Frankham, 1995). This means that the 2016 brood stock was most likely equivalent to a random mating population of 12 males and 12 females.

To put the Ne into context, generally the minimum viable effective population size to avoid inbreeding depression is Ne greater or equal to 50 (Frankham *et al*., 2014, Frankham, 1995). This assumes that the starting breeding population is unrelated. If the JBARB lumpfish breeding program continues to use a breeding stock with an Ne=24, then it will not be long before the offspring in future generations start showing signs of inbreeding depression. Inbreeding depression is a reduction in biological fitness which reduces productivity and profitability. At the absolute minimum, the JBARB breeding program needs an Ne=50 to have any success in the short term. For example, if the JBARB had 50 unrelated females and 25 unrelated males (each male covers 2 females), the Ne= 67 which would be sufficient to avoid inbreeding depression.

However, successful breeding programs that achieve genetic gains in the long term, do not just avoid inbreeding depression. Instead, they want to maintain genetic variability within the breeding population to maximum genetic gains while minimizing adverse effects of selection (e.g., inbreeding). The minimum effective population size for sufficient genetic variability is generally equal to or greater than 500 (Frankham *et al*., 2014, Frankham, 1995). Therefore, if the JBARB wants to maintain genetic variability within the breeding program they would have a Ne=500. For example, if the JBARB had 360 females with 200 males, the Ne= 514 which would be sufficient to

maintain genetic variability. Note that 360 females and 200 males would not have to be part of the core breeding stock at the facility. Some of the lumpfish could come from other facilities or even other countries, to help maintain genetic diversity. Effective population size is related to the rate of increase in inbreeding per generation (delta F) with the equation of delta F = 1/(2 *Ne). Going back to the JBARB 2016 YC breeding stock of 9 females and 18 males (Ne=24), the rate of increase in inbreeding per generation if the Ne was maintained would be 2.1% per generation (1/(2*24)).

3.2 Selection goal(s)

Selection goals are the reason why breeding programs start, if no selection goals are set then how does one know what they are breeding for. In terms of a lumpfish breeding program, the main selection goal is to increase the sea lice eating efficiency in the commercial population. The more sea lice the commercial lumpfish eat off the salmon, the less sea lice infections there are which leads to more profit for producers. There have been preliminary studies showing that there is a genetic component to the amount of sea lice a lumpfish eats (Imsland *et al.,* 2016a). Therefore, the breeding program would select lumpfish that eat more sea lice and less pellets to be the breeding stock for the future generation and cull the lumpfish that eat more pellets and less sea lice. After multiple generations of selecting superior sea lice eating lumpfish and culling the lumpfish that eat pellets, the population mean for sea lice eating efficiency would be greater than the original population if there is genetic variation for the sea lice eating efficiency trait within the population.

However, it is not common for commercial breeding programs to only select for one trait, normally they selected for multiple traits. Another trait that the JBARB breeding program should consider selecting for is disease resistance. Some of the leading diseases common among commercial lumpfish populations are vibriosis, furunculosis and pasteurellosis (Erkinharju et al., 2020). Having any pathogens present would decrease the sea lice eating efficiency of the lumpfish and the pathogens could infect the salmon. Selectively breeding disease resistant lumpfish could also be more cost effective than vaccines that protect against common diseases.

Another trait that could be selected for is body size (either weight or length). As stated earlier, numerous studies have shown that the smaller lumpfish are more likely to forage for sea lice instead of eating the pellets because they cannot physically fit the pellets in their mouth. Therefore, the breeding program could select smaller lumpfish to be the breeding stock for the next generation and cull the larger lumpfish that are more likely to eat the pellets. After multiple generations of selecting smaller lumpfish, the population mean for body size would be smaller than the original population. However, it is also important to understand the correlation between different traits. Some traits are positively correlated meaning that as one trait is selected for, the other trait also is unintentionally selected for. Similarly, some traits are negatively correlated meaning that as one trait is selected for, the other trait is adversely selected against. For example, body size is correlated with age at sexual maturity, as smaller body sizes are selected for, earlier sexual maturity is unintentionally selected for.

Lastly, lumpfish could be selected for temperature tolerance especially in the more southern regions. Lumpfish are a naturally cold-water cleaner fish, with preferred broad water temperatures ranging from 0-11 degrees Celsius and an optimal temperature range of 4-7 degrees Celsius (Eriksen *et al.,* 2014). The problem with using lumpfish in the salmon sea cages in Newfoundland is that some parts of the coast experience water temperatures higher than 11 degrees during the summer months. These higher water temperatures cause the caged lumpfish to dive deeper in the water column and become inactive and stop eating the sea lice off the salmon. However, the challenge with selectively breeding for temperature tolerance is that there are physiological limits with regards to enzyme denaturing. All enzymes have an optimal temperature range where the enzyme functions properly and the rate of reactions are maximized. As the temperature deviates from the optimal temperature the enzymes' rate of reactions are lowered. If the temperature deviates to the point where the temperature is no longer in the enzymes' functional temperature range, the enzymes start to denature which means that the rate of reactions are non-existent. If the enzymes can no longer perform as catalysts, the whole function stops occurring which has negative physiological effects to the organism. The enzyme temperature optimum is dependent on the natural environment of the species. It might be beneficial to collect lumpfish from southern subpopulations of wild lumpfish because those subpopulations would have a higher enzyme temperature optimum compared to the northern subpopulations. Therefore, the lumpfish from the southern populations would be more

likely to thrive with the increasing water temperatures due to global warming even when placed in the sea cages in the northern regions.

Another important aspect to consider when selectively breeding lumpfish is the effects of domestication. The majority of the brood stock is from wild populations, so by collecting wild lumpfish and putting them in an artificial environment for captive breeding there will be adverse selection pressures that are not directly selected for. The process of collecting fish from the wild produces a lot of stress for the captured fish that can cause negative effects on the biological system. Common stressors would include netting, handling of the fish, decreased water quality (e.g., increase levels of ammonia) and the transportation to the rearing facility (Milla *et al.,* 2020). In most cases, some of the captured fish would not be able to handle the stresses of being collected and would die off within the first few days of being captured. Just by collecting the fish there would be selection pressures, fish that can handle the collection stresses and acclimation into captivity are adversely selected for. At the collection point of the process, the sea lice eating efficiency of the lumpfish would be irrelevant because if the superior sea lice eating lumpfish cannot handle the stress of collection and die off, their good genetics cannot be used in the breeding program. Even after the initial acclimation period, the lumpfish would be adversely selected based on their behavior within the captive environment. Skittish fish that are afraid of human interactions would stay at the bottom of the tank and would not come up and feed which would lead to starvation. In contrast, lumpfish with bold or curious personality traits would handle human interactions and

therefore feed readily. In this case, bold lumpfish would be unintentionally selected for as they would thrive in a captive environment.

It is also important to note some of the most common effects of domestication that are inevitable. Early domestication for most species involves selecting for tameability (positive human interactions) which is thought to be correlated with these development regulatory genes (Trut *et al.,* 2009). In particular, neoteny is common effect of domestication. Neoteny is when animals reach sexual maturity while still retaining juvenile characteristics. Domesticated animals also tend to reach sexual maturity at an earlier age to their wild counterparts, which is another characteristic of neoteny. In terms of lumpfish the effects of domestication might be positive for the breeding program, as wild lumpfish reach maturity at 2-5 years of age depending on the subpopulation. If the domestication process causes future generations of lumpfish to mature in 1-3 years through neoteny, it would decrease the generation interval. Decreasing the generation interval is a way to capitalize on the genetic gains per year through the breeders' equation which is defined as $\Delta G=(A\sigma_g i)/l$, where ΔG is genetic gain per year, A is the accuracy of the selection traits, σ_g is the standard deviation of the additive genetic variance, i is the selection intensity and l is the generation interval. Therefore, as the generation interval decreases the genetic gain per year increases within the physiological limits of the species.

3.2.1 Phenotypes needed for selection goal(s)

Once the selection goals are established, there needs to be a way to determine which animals are superior for those selected traits. Phenotypes are the observable

characteristics of an individuals, that are the sum of the genetic effects and the environmental effects (Phenotype = Genetic effects + Environmental effects). As stated earlier, the main selection goal for the lumpfish breeding program is sea lice eating efficiency. To determine which lumpfish are better sea lice eaters than others, the amount of sea lice eaten for each lumpfish needs to be determined. One way to determine the amount of sea lice each lumpfish eats is by morphological stomach content analysis which is the process of dissecting the stomach and a trained taxonomist identifies the contents. The problem with morphological stomach content analysis is that the fish has to be sacrificed in order to perform the stomach content analysis. To determine which brood stock are superior sea lice eater's progeny testing will have to be performed. The progeny from the brood stock would be deployed in the salmon sea cages to eat the sea lice off the salmon, then some of the progeny for each brood stock could be sacrificed for stomach content analysis. The results from the progeny would determine the value for the brood stock under the assumption that if the progeny are superior sea lice eaters, then the parents (brood stock) would also be superior sea lice eaters because the brood stock were the ones that passed down the superior genes to the progeny. To get accurate estimates on how well the brood stock perform for sea lice eating (the genetic merit the lumpfish has for the sea lice eating trait), numerous offspring will need to be tested, the bare minimum number of offspring tested for each brood stock should be around a 100 progeny. Keep in mind, the more progeny tested the more accurate the EBVs for the brood stock become.

As for disease resistance selection, the phenotypes needed are whether or not the lumpfish acquires the prevalent disease within the environment, in this case vibriosis or furunculosis. In most cases it is not just whether or not the lumpfish is infected with the disease but also the severity of the infection. To record the phenotypes of disease resistance, disease trials could be performed with some of the brood stock progeny in a closed environment. The progeny would be exposed to either vibriosis or furunculosis depending on the trial, and each progeny would be observed and the response would be recorded (e.g., the lumpfish is resistant or not and the severity of the disease). Then the data recorded from the progeny would be related back to the brood stock under the assumption that if the progeny has disease resistance than the brood stock passed down the disease resistance genes to them. The brood stock with the most disease resistant offspring in the disease trials would be the selected candidate brood stock.

Likewise, for temperature tolerance selection, the phenotypes needed are whether the lumpfish can thrive at higher water temperatures or not. To record temperature tolerance, temperature tolerance trials could be performed using some of the brood stock progeny in a closed and controlled environment. Different progeny from the brood stock would be placed in different water temperature tanks (e.g., increasing water temperatures) and the physiological response would be recorded for each fish. Again, the records would be related back to the brood stock under the assumption that the brood stock passed down the genes for temperature tolerance to the offspring. The brood stock with the most temperature tolerant offspring in the temperature trials would be the selected candidate brood stock.

For the body size selection, the phenotypes needed would be either body length or body weight or a combination of both. Body length and body weight can be taken on the brood stock, but to determine if the body size is due to genetics and not the environment, progeny testing would also have to be done. Measurements of body weight and body length would be recorded at different defined life stages and again would be related back to the brood stock. The brood stock with the smallest offspring in terms of body weight and length would be the selected as the candidate brood stock.

3.3 Mating schemes

The principal behind a mating scheme is how the brood stock are mated to one another to maximize the response to selection (genetic gain) while maintaining genetic diversity within the population. There a multiple mating schemes that are commonly implemented in fish breeding including mass, single pair mating, nested, partial factorial and full factorial ((Dupont-Nivet et al., 2006, Fiumera et al., 2004, Busack and Knudsen, 2007). Each of these mating schemes have pros and cons that have to be considered before implementing them into the lumpfish breeding program.

Mass mating involves pooling the sperm of all the breeding males then fertilizing all the eggs with the pooled sperm. Out of all the mating schemes, mass mating is the most cost effective because it is the least labor intensive. However, one problem with mass mating is sperm competition (Dupont-Nivet et al., 2006). Just because the sperm is pooled at equal proportions among all males, does not guarantee that the fertilization rate would be in equal proportions among the males. Sperm competition would result in an over-representation of only a few males' genetics within the offspring. In addition,

there is no method to track or record matings when implementing mass mating because all the males' gametes are mixed. With known matings, it makes it easier to determined parentage because the offspring's genotype can be predicted with the genotypes of the individuals for each recorded mating.

Single pair mating involves mating one male to one female; therefore, each male and female only has one mate. The advantage with single pair mating is the simplicity of recording and tracking matings because each brood stock only has one mate. However, the disadvantage of single pair mating is that each brood stock's genetics are only captured with one mating. Therefore, if one of the matings failed for a variety of reasons (e.g., sterile male, immature eggs, human error), then both individuals in that mating would not contribute any genetics within the next generation (Fiumera et al., 2004).

Nested mating design involves mating sires to multiple dams while the dams are only used once. Nested mating design is better at distributing the mating fail risk on the sire side because the males are mated to multiple females, but on the dam side the risk of mating failure is still present because the dams are only used once. The nested design works well for single offspring bearing species, where the females are only limited to one mating per cycle. However, fish are a multiple bearing species meaning the dams are not limited to one mating, therefore the females could also be mated to more than one male. Another disadvantage of the nested mating design is that the females cannot be selected for at the same intensity of the males because they are only used once which means that some of the potential genetic gain is lost.

Factorial mating designs are becoming increasingly popular for fish breeding because it reduces the risk of mating failure for both the sire and dam. There are two types of factorial mating designs: partial factorial and full factorial. In a partial mating scheme, a subset of males is mated to a subset of females and vice versa. For example, if a breeding program had 10 males and 10 females, then 5 males would be mated to 5 different females and the other 5 males would be mated to another 5 different females. This example would be classified as a 5X5 partial factorial design which would create 50 different matings. In a full factorial mating scheme X number of males are mated to all X number of females. The disadvantage to a full factorial design is that it is time consuming and complicated because each male has to be bred to every female which creates a lot of matings that all have to be properly executed (Dupont-Nivet et al., 2006). From a practical standpoint, the partial factorial mating scheme would be less complicated and more time effective compared to full factorial because it would require less matings.

One advantage of the full factorial mating scheme is that it increases the effective population size while reducing inbreeding over multiple generations (Busack and Knudsen, 2007). Studies have also shown that the full factorial design produces the greatest amount of genetic gain compared to the single pair mating and nested mating when the heritability, population size and selection intensity stay constant among the schemes (Dupont-Nivet et al., 2006). Partial factorial mating schemes are comparable to the full factorial mating schemes in terms of increasing effective population size,

reducing inbreeding and maximizing genetic gain only when the partial factorial mating design used large mating sets (e.g., 10X10 partial factorial) (Dupont-Nivet et al., 2006).

Given that a breeding program can get comparable genetic gains when implementing a partial mating scheme with large mating sets to a full factorial design while being more time and cost efficient, a partial mating should be considered for the lumpfish breeding program. For example, if the breeding program had 100 males and 100 females after applying the selection intensity threshold, then a 10X10 partial mating (10 sets of 10X10) or a 20X20 partial mating (5 sets of 20X20) or even a 25X25 partial mating (4 sets of 25X25) could be applied.

3.4 Reproductive technology

At the JBARB facility at Memorial University (Newfoundland), where currently all lumpfish are commercially produced, the captive females are assessed during spawning season and when the females are ready to spawn, they are put in holding tanks. Once the eggs are ready to be released, the females are stripped of their eggs (Monk, 2019). Female lumpfish can spawn twice during a single spawning season, roughly 2 weeks apart (Fry, 2018). On the other hand, the sexually mature males are sacrificed for their milt (semen). The gonad is harvested from the males and the milt is extracted and purified, removing any tissue (Monk, 2019).

Once the eggs are fertilized, they are artificially raised in incubators tanks with high water flow and a stable water temperature of 9 degrees Celsius. After about 300-degree days the eggs begin to hatch. Just prior to hatching, the eggs are transferred to

the larval tanks and any unhatched eggs are removed after 3 days post hatching (Monk, 2019). The larvae are fed brine shrimp three times a day for the first 35 days post-hatching. At 45 days post hatching, the juvenile lumpfish are fully weaned on dry food (Monk, 2019). Around 50 days post hatching, the lumpfish are graded by weight where the fish go through a grader that is 3 mm in diameter. At this time, the juvenile lumpfish are transfer to 3000L (~80 U.S gallon) tanks. The lumpfish are continuously graded for weight on a biweekly basis until they are deployed to the salmon sea cages.

One benefit of the labor-intensive breeding practice is that producers have control of the system at every stage of the process. This is important because it allows the producers to keep and maintain records at every stage. For instance, since the producers have to physically strip the eggs from the females and milt from the males and then add the milt to the eggs, this allows the producers to record the different matings of the fish. This would allow the producers to monitor and manage inbreeding and their selection goals.

A major challenge with the current breeding practice is that the process will it inevitably limit the genetic diversity because it is so labor intensive. By stripping the female of the eggs and the male for the milt, it would be more time efficient to let's say use only 10 superior males to cover all the females instead of sacrificing and collecting milt from 100 superior males. If the breeding program only uses 10 superior males to save time and money instead of using 100 superior males available, it bottlenecks the genetic diversity available for the future generations by 10X. If the breeding program was established then the selection intensities could be stricter to allow for greater

genetic gains per generation. However, since this breeding program is just in its infancy, so it is too early to limit the genetic diversity or else there will be nothing to select for in future generations.

The last major challenge in the current lumpfish breeding practice is that the male lumpfish have to be sacrificed to collect the milt. The problem with this is that male lumpfish can only be used once no matter how good the genetics are for the desired trait. Furthermore, without the ability to use the male lumpfish more than once it would at the upmost importance to preserve the milt from the superior males for future use. However, to date cryopreservation (freezing of semen) is not widely used in lumpfish breeding. Cryopreservation would be a vital tool for breeding lumpfish because it would not only preserve superior genetics but would allow for the transportation of milt to exchange genetic material with other commercial breeding facilities to aid in genetic diversity within the population. Pountney et al. (2020) reported that lumpfish sperm can remain mobile up to 3 hours after collection and when diluted in extender solutions can stay viable for up to 14 days after collection when stored at 4 degrees Celsius. In addition, there have been studies that have shown cryopreservation can be successfully used for lumpfish semen (Norðberg et al., 2015). Going forward, it would be beneficial for the JBARB lumpfish breeding program to look into proven cryopreservation protocols for lumpfish milt so that the breeding program can preserve superior genetics on the male side for future generations.

Summary of my recommendations for the Canadian Lumpfish Breeding Program:

1. Collect wild lumpfish from as many locations and subpopulations that is financially feasible (e.g., collecting 20 lumpfish at each site with a minimum of 20 different collection sites).
2. Keep the effective population size above 50 and aim for closer to 500.
3. Multiple selection goals like sea lice eating efficiency, disease resistance, temperature tolerance and body size
4. There will be adverse selection due to domestication (rearing lumpfish in captivity) that have to be accounted for.
5. A partial factorial mating scheme could be implemented to achieve the greatest genetic gain.
6. Preservation of milt should be considered, so that males can be used in multiple matings
7. Estimated Breeding Values should be performed with BLUP.

Chapter 2: Tables and Figures

Table 1: Table of number of lumpfish fin-clip samples from each of the different year classes and location. All of the lumpfish from the Joe Brown Aquatic Research Building are possible parents. All of the lumpfish fin- clips caught in Olive Cove Newfoundland have stomach content data associated with them and are the progeny of the possible parents.

Source Population	Year Class	Number of fin-clip samples
Joe Brown Aquatic Research Building	2015	23
Joe Brown Aquatic Research Building	2016	98
Joe Brown Aquatic Research Building	2017	82
Olive Cove sea cage Newfoundland	Unknown (Most likely 2017)	84
Total number of samples		**287**

Table 2: The 4 - 2015YC families defined from the Colony kinship analysis. All offspring within a family had to have a probability of 0.85 of being full sibs with each other to be classified as full sibs. 2015YC Family #2 had the largest full sib family of 8 individuals and 2015YC Family #4 had the smallest full sib family of 2 individuals

2015YC Family #1	2015YC Family #2	2015YC Family #3	2015YC Family #4
20152C0FID1B	20152CFID10B	20152CFID15B	20152CFID21B
20152CFID11B	20152CFID14B	20152CFID18B	20152CFID22B
20152CFID9B	20152CFID12B	20152CFID16B	
20152CFID6B	20152CFID4B	20152CFID19B	
20152CFID8B	20152CFID13B		
20152CFID7B	20152CFID3B		
	20152CFID2B		
	20152CFID5B		

Table 3: The 6 - 2016YC full sib families from the Colony kinship analysis. All offspring within a family had to have a probability of 0.85 of being full sibs with each other to be classified as full sibs. Family sizes ranged from 3 individuals to 25 individuals.

family1	family 2	family 3	family 4	family 5	family 6
20162CFID126A	20162CFID166A	20162CFID64	20162CFID117A	20162CFID128A	20162CFID158A
20162CFID144A	20162CFID174A	20162CFID65	20162CFID118A	20162CFID152A	20162CFID27
20162CFID176A	20162CFID53	20162CFID68	20162CFID119A	20162CFID168A	20162CFID30
		20162CFID69	20162CFID120A	20162CFID18	20162CFID31
		2016CFID67	20162CFID125A	20162CFID23	20162CFID33
			20162CFID133A	20162CFID52	20162CFID42
			20162CFID135A	20162CFID9	20162CFID50
			20162CFID143A	2016CFID43	20162CFID51
			20162CFID145A	2016CFID48	20162CFID75
			20162CFID159A	2016CFID66	20162CFID80
			20162CFID167A		20162CFID81
			20162CFID169A		2016CFID76
			20162CFID20		20162CFID26
			20162CFID22		20162CFID4
			20162CFID28		20162CFID5
			20162CFID32		20162CFID82
			20162CFID37		20162CFID83
			20162CFID44		2016CFID59
			20162CFID49		2016CFID63
					20162CFID17
					20162CFID56
					20162CFID62
					20162CFID7
					2016CFID73
					20162CFID60

Table 4: The 25 - 2017YC and sea cage samples full sib families from the Colony kinship analysis. All offspring within a family had to have a probability of 0.85 of being full sibs with each other to be classified as full sibs. Family sizes ranged from 2 individuals to 12 individuals.

family 1	family 2	family 3	family 4	family 5	family 6	family 7	family 8
20172CDIDO67	20172CFID001	20172CFID004	20172CFID035	20172CFID062	20172CFID063	20172CFID070	20172CFID078
20172CFID079	20172CFID003	20172CFID006	20172CFID038	20172CFID073	SDS0F127	20172CFID080	SDS0F051
20172CFID081	20172CFID031	20172CFID032	20172CFID039	SDS0F003	20172CFID071	SDS0F059	SDS0F156
SDS0F055	20172CFID034	20172CFID049	20172CFID041	SDS0F082	SDS0F111	20172CFID075	SDS0F091
SDS0F056	20172CFID042	20172CFID027	20172CFID043	SDS0F105	20172CFID064		SDS0F119
SDS0F108	20172CFID053	20172CFID009	20172CFID044	SDS0F107	SDS0F006		
SDS0F130	20172CFID002	20172CFID036	20172CFID046	SDS0F110	SDS0F077		
SDS0F137	20172CFID054	20172CFID007	20172CFID051	SDS0F112	SDS0F090		
20172CFID066			20172CFID050	SDS0F113	SDS0F092		
20172CFID076				SDS0F126			
SDS0F104				SDS0F078			
family 9	family 10	family 11	family 12	family 13	family 14	family 15	family 16
SDS0F063	SDS0F066	20172CFID005	20172CFID011	20172CFID016	20172CFID058	SDS0F065	SDS0F129
SDS0F081	SDS0F080	20172CFID008	20172CFID029	20172CFID045	20172CFID059	SDS0F073	SDS0F089
SDS0F095	SDS0F094	20172CFID037	20172CFID033	20172CFID056	20172CFID060	SDS0F120	
SDS0F115	SDS0F121	20172CFID047	20172CFID040				
	SDS0F068						
	SDS0F069						
	SDS0F070						
family 17	family 18	family 19	family 20	family 21	family 22	family 23	family 24
20172CFID074	20172CFID077	SDS0F062	SDS0F079	SDS0F086	SDS0F088	20172CFID030	20172CFID068
SDS0F064	SDS0F057	SDS0F103	SDS0F093	SDS0F114	SDS0F128	20172CFID055	SDS0F087

family 25							
20172CFID069							
SDS0F002							

Table 5: Cervus simulation results when sexes of parents were defined as unknown. Simulations with a large family size (10,000 offspring and 100 candidate parents) are simulations a), c-1), c-2), c-3), d-1), d-2), d-3). Simulations with a small family size (500 offspring and 10 candidate parents) are simulations b), e-1), e-2), e-3),f-1),f-2) and f-3). The most important trend for both the large and small family size is that when the proportion of relatives among the candidate parents increases so does the critical LOD score. Furthermore, as the proportion of relatives among the candidate parents increases the Parent Pair: Non-parent pair (true parent pair sampled) Mean LOD becomes positive and similar in value to the Parent Pair: True Parent Mean LOD.

Simulation	Pair Parent Assignment Rate- Strict (0.95 confidence level)	Pair Parent Unassigned Rate	Critical LOD – Strict (0.95 confidence level)	Parent Pair: True Parent Mean LOD (SD)	Parent Pair: Non-parent pair (true parent pair sampled) Mean LOD (SD)	Parent Pair : Non-parent pair (one true parent sampled) Mean LOD (SD)	Parent Pair: Non-parent pair (neither true parent sampled) Mean LOD (SD)
Large Family Size							
a)Large Family Size	95% (9462 offspring assigned)	0% (0 offspring unassigned)	-7.50	35.81 (5.72)	0(0)	-7.88 (7.44)	-35.23 (7.49)
c-1) Large family size with 0.2 relatives	95% (9464 offspring assigned)	0% (0 offspring unassigned)	5.00	35.96 (5.67)	27.19 (5.60)	4.63 (8.03)	-6.47 (6.40)
c-2) Large family size with 0.5 relatives	95% (9536 offspring assigned)	0% (0 offspring unassigned)	9.00	35.90 (5.78)	31.19 (4.88)	9.53 (6.88)	-2.34 (6.59)
c-3) Large family size	95%(9464 offspring	0% (0 offspring	13.50	35.90 (5.67)	25.86 (7.16)	13.18 (6.08)	4.62 (6.04)

with 1.0 relatives	assigned)	unassigned)					
d-1) Large family size with full sib mating	95% (9490 offspring assigned)	0% (1 offspring unassigned)	-7.00	36.11 (5.89)	0 (0)	-7.36 (7.52)	-31.56 (6.42)
d-2) Large family size with half sib mating	95% (9480 offspring assigned)	0% (1 offspring unassigned)	-7.50	35.97 (5.66)	0 (0)	-8.22 (7.57)	-32.93 (8.10)
d-3) Large family size with 1st cousin mating	95%(9479 offspring assigned)	0% (1 offspring unassigned)	-8.50	35.98 (5.66)	0 (0)	-8.42 (7.54)	-29.38 (9.28)
Small Family Size							
b) Small Family Size	92% (461 offspring assigned)	0% (0 offspring assigned)	-14.00	35.33 (5.65)	0 (0)	-19.94 (10.32)	-50.15 (3.47)
e-1) Small family size with 0.2 relatives	94% (471 offspring assigned)	0% (0 offspring assigned)	-5.50	35.61 (5.44)	0 (0)	-7.15 (12.56)	-33.21 (0)
e-2) Small family size with 0.5 relatives	95% (473 offspring assigned)	0% (0 offspring assigned)	-2.00	35.92 (6.01)	0 (0)	-3.85 (10.28)	-16.02 (11.97)
e-3) Small family size with 1.0 relatives	95% (474 offspring assigned)	0% (0 offspring assigned)	2.50	35.99 (5.65)	0 (0)	2.91 (9.24)	-20.92 (0)
f-1) Small	94% (469	0% (0	-14.00	36.55	0	-17.07	0

family size with full sib mating	offspring assigned)	offspring assigned)		(5.93)	(0)	(9.88)	(0)
f-2) Small family size with half sib mating	92% (461 offspring assigned)	0% (0 offspring assigned)	-14.00	36.23 (5.83)	0 (0)	-19.33 (11.45)	0 (0)
f-3) Small family size with 1st cousin mating	97% (485 offspring assigned)	0% (0 offspring assigned)	-19.00	36.02 (5.47)	0 (0)	-15.88 (9.10)	0 (0)

Table 6: Cervus simulation results when sexes of parents were defined as known. Simulations with a large family size (10,000 offspring with 50 candidate mothers and 50 candidate fathers) are simulations i), iii-1), iii-2), iii-3), iv-1), iv-2), iv-3). Simulations with a small family size (500 offspring with 5 candidate mothers and 5 candidate fathers) are simulations ii), v-1), v-2), v-3), vi-1), vi-2) and vi-3). The most important trend for both the large and small family size is that when the proportion of relatives among the candidate parents increases so does the critical LOD score. Furthermore, as the proportion of relatives among the candidate parents increases the Parent Pair: Non-parent pair (true parent pair sampled) Mean LOD becomes positive and similar in value to the Parent Pair: True Parent Mean LOD.

Simulation	Pair Parent Assignment Rate- Strict (0.95 confidence level)	Pair Parent Unassigned Rate	Critical LOD – Strict (0.95 confidence level)	Parent Pair: True Parent Mean LOD (SD)	Parent Pair: Non-parent pair (true parent pair sampled) Mean LOD (SD)	Parent Pair: Non-parent pair (true mother unsampled) Mean LOD (SD)	Parent Pair: Non-parent pair (true father unsampled) Mean LOD (SD)	Parent Pair: Non-parent pair (neither true parent sampled) Mean LOD (SD)
Large Family Size								
i)Large Family Size	95% (9463 offspring assigned)	0% (0 offspring unassigned)	-7.50	35.87 (5.55)	0 (0)	-7.78 (7.45)	-7.52 (7.47)	-26.10 (5.36)
iii-1) Large family size with 0.2 relatives	95% (9492 offspring assigned)	0% (2 offspring unassigned)	15.00	35.89 (5.71)	29.10 (5.33)	14.82 (7.16)	14.75 (7.30)	-3.45 (6.37)
iii-2) Large family size with 0.5 relatives	94% (9374 offspring assigned)	0% (1 offspring unassigned)	19.88	35.96 (5.67)	30.13 (6.15)	18.85 (6.40)	18.90 (6.02)	7.76 (5.89)
iii-3) Large family size with 1.0	92% (9162 offspring assigned)	0% (1 offspring unassigned)	24.03	35.99 (5.58)	29.31 (5.32)	21.90 (6.10)	21.83 (6.03)	11.58 (4.19)

relatives								
iv-1) Large family size with full sib mating	95% (9468 offspring assigned)	0% (1 offspring unassigned)	-7.50	36.15 (5.90)	0 (0)	-7.83 (7.32)	-7.53 (7.50)	-24.89 (7.92)
iv-2) Large family size with half sib mating	95% (9495 offspring assigned)	0% (1 offspring unassigned)	-7.50	36.07 (5.71)	0 (0)	-7.42 (7.20)	-7.26 (7.74)	-26.85 (7.05)
iv-3) Large family size with 1st cousin mating	95% (9484 offspring assigned)	0% (0 offspring unassigned)	-7.50	35.89 (5.73)	0 (0)	-7.55 (7.52)	-8.15 (7.10)	-28.70 (8.41)
Small Family Size								
ii) Small Family Size	93% (467 offspring assigned)	0% (0 offspring unassigned)	-24.00	35.61 (5.38)	0 (0)	-25.62 (9.79)	-28.52 (10.88)	-53.18 (2.11)
v-1) Small family size with 0.2 relatives	94% (469 offspring assigned)	0% (0 offspring unassigned)	-9.00	36.00 (5.37)	0 (0)	-14.02 (17.18)	-12.60 (18.67)	-65.51 (0)
v-2) Small family size with 0.5 relatives	96% (478 offspring assigned)	0% (0 offspring unassigned)	5.00	36.24 (5.62)	31.19 (0)	1.07 (14.41)	7.18 (9.71)	-6.81 (0)
v-3) Small family size with 1.0 relatives	97% (486 offspring assigned)	0% (0 offspring unassigned)	9.00	36.16 (5.53)	0 (0)	12.34 (7.51)	13.07 (10.41)	0 (0)
vi-1) Small family size	92% (458 offspring	0% (0 offspring	-19.00	35.89 (6.24)	0 (0)	-24.65 (12.34)	-23.44 (10.69)	-52.45 (11.66)

with full sib mating	assigned)	unassigned)						
vi-2) Small family size with half sib mating	94% (471 offspring assigned)	0% (0 offspring unassigned)	-24.00	35.99 (6.03)	0 (0)	-24.40 (11.54)	-23.30 (9.78)	-53.12 (10.30)
vi-3) Small family size with 1st cousin mating	93% (467 offspring assigned)	0% (0 offspring unassigned)	-19.00	35.59 (5.49)	0 (0)	-23.74 (10.65)	-22.84 (13.77)	-38.79 (0)

Table 7: Cervus simulation results when there was 200 females and 10 males. The most important trend is that when the proportion of relatives among the candidate parents increases so does the critical LOD score. Furthermore, as the proportion of relatives among the candidate parents increases the Parent Pair: Non-parent pair (true parent pair sampled) Mean LOD becomes positive and similar in value to the Parent Pair: True Parent Mean LOD. When half sibs were among the candidate parents the critical LOD scores were not as high in value compared to when there were full sibs among the candidate parents.

Simulation	Critical LOD	Assignment Rate	Mean LOD True parent pair	Mean LOD Non-parent pair (true parent sampled)	Mean LOD Non-parent pair (true mother unsampled)	Mean LOD Non-parent pair (true father unsampled)	Mean LOD Non-parent pair (neither true pair sampled)
a) 200 females, 10 males producing 10,000 sampled offspring	-8.50	95%	35.86 (5.68)	0.00 (6.69)	-1.65 (9.01)	-17.31 (9.01)	-30.42 (8.83)
b-i) 200 females, 10 males producing 10,000 sampled offspring. 20% of females are full sibs	12.00	94%	35.90 (5.73)	28.70 (5.05)	20.60 (6.22)	-16.77 (8.66)	-19.88 (5.32)
b-ii) 200 females, 10 males producing 10,000 sampled offspring. 20% of females are full sibs. 20% of males are full sibs.	17.50	94%	35.97 (5.60)	29.42 (5.77)	21.05 (6.09)	2.50 (12.64)	-8.42 (11.88)
c-i) 200 females, 10 males producing 10,000 sampled offspring. 50% of females are full sibs.	19.94	94%	36.08 (5.64)	30.27 (6.15)	23.89 (5.28)	-14.91 (7.78)	-14.05 (8.67)
c-ii) 200 females, 10 males producing 10,000 sampled offspring. 50% of females are full sibs.50% of males are full sib	21.50	93%	35.98 (5.62)	30.52 (5.76)	23.98 (5.57)	10.80 (7.57)	3.87 (6.75)
d-i) 200 females, 10 males producing 10,000 sampled offspring. 20% of females are half sibs	-6.00	96%	35.96 (5.67)	0.00 (0.00)	7.82 (6.49)	-17.82 (8.83)	-28.01 (6.58)

d-ii) 200 females, 10 males producing 10,000 sampled offspring. 20% of females are half sibs. 20% of males are half sib .	1.00	95%	35.90 (5.70)	35.33 (0.00)	6.69 (6.82)	-9.92 (9.82)	-22.79 (8.24)
e-i) 200 females, 10 males producing 10,000 sampled offspring. 50% of females are half sibs.	1.50	94%	36.12 (5.67)	30.53 (0.00)	10.72 (6.33)	-16.99 (8.71)	-23.37 (7.39)
e-ii)200 females, 10 males producing 10,000 sampled offspring. 50% of females are half sibs. 50% of males are half sib.	5.00	95%	35.90 (5.68)	28.74 (12.92)	10.58 (6.18)	-4.75 (8.83)	-14.68 (8.42)

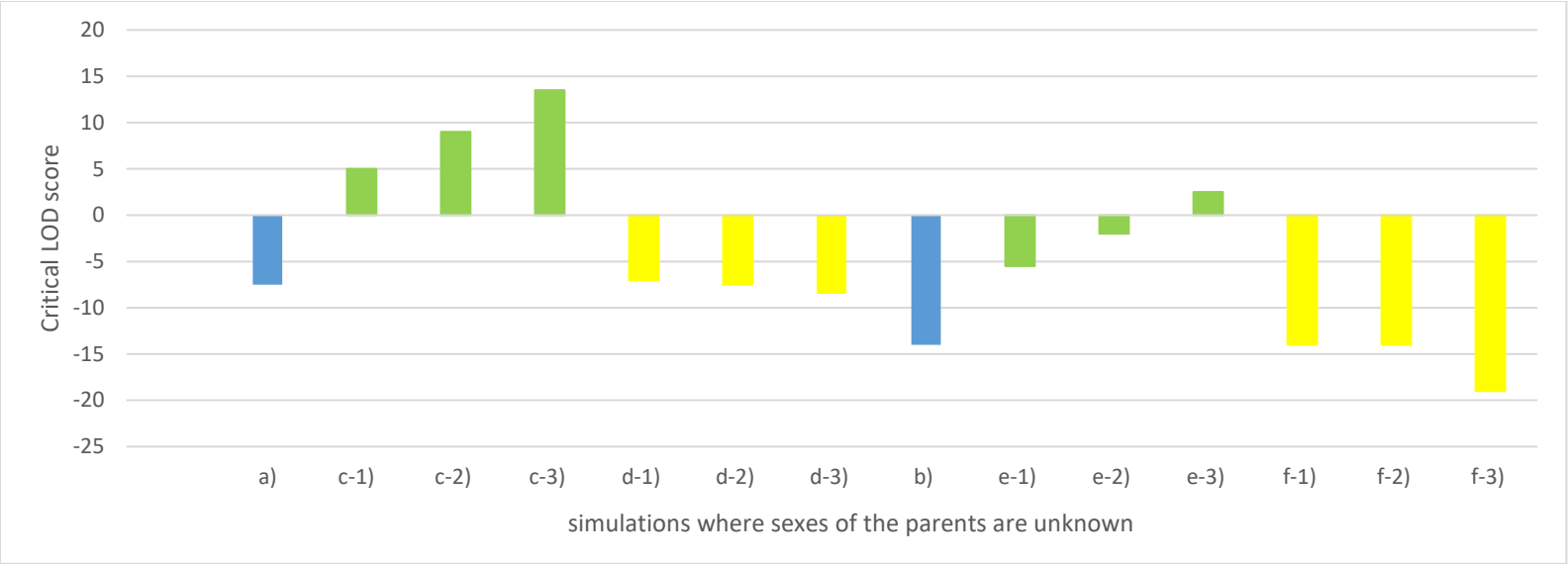


Figure 1: The critical LOD scores plotted against simulations where sexes of the parents are unknown. Blue bars are simulations with no inbreeding and no relatives among candidate parents, green bars are simulations with relatives among the candidate parents and the yellow bars are simulations with inbreeding present. Within the large family size simulations (a), c-1), c-2), c-3), d-1), d-2), d-3)), simulation d-3) large family size with 1st cousin mating had the lowest critical LOD score of -8.50 and simulation c-3) large family size with 1.0 relatives had the highest critical LOD score of 13.50. Within the small family size simulations (b), e-1), e-2), e-3), f-1), f-2), f-3)), simulation f-3) small family size with 1st cousin mating had the lowest critical LOD score of -19.00 and simulation e-3) small family size with 1.0 relatives had the highest critical LOD score of 2.50.

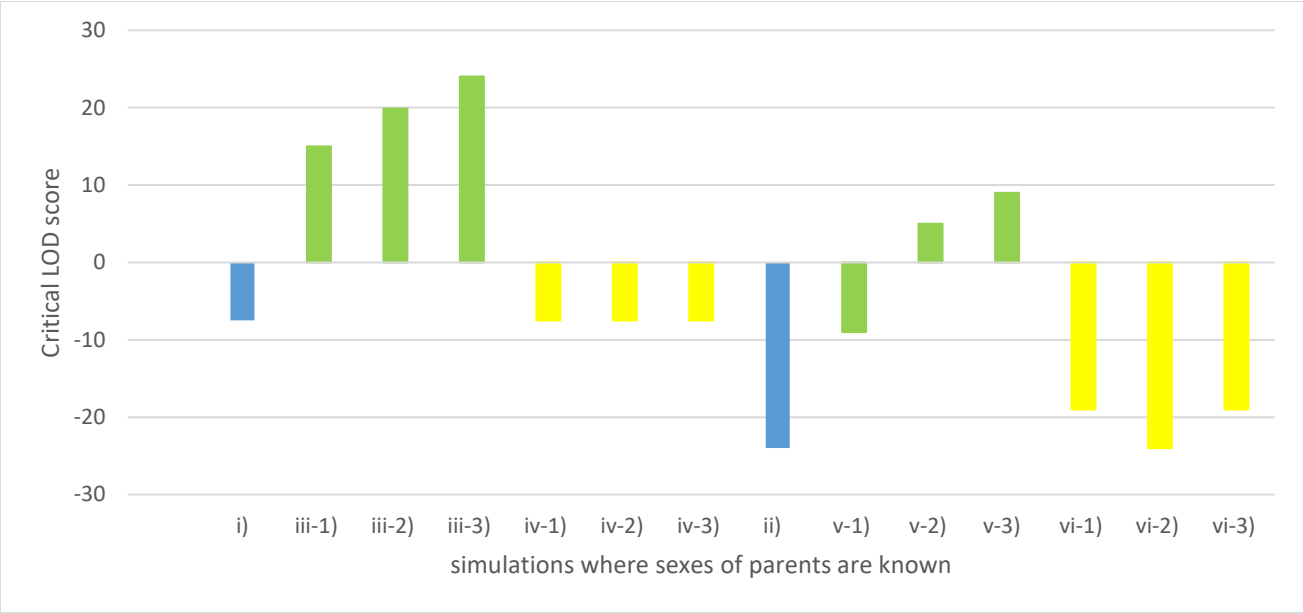


Figure 2: The critical LOD scores plotted against simulations where sexes of the parents are known. Blue bars are simulations with no inbreeding and no relatives among candidate parents, green bars are simulations with relatives among the candidate parents and the yellow bars are simulations with inbreeding present. Within the large family size simulations (i), iii-1), iii-2), iii-3), iv-1), iv-2), iv-3)), simulation i) large family size had the lowest critical LOD score of -7.50 and simulation iii-3) large family size with 1.0 relatives had the highest critical LOD score of 24.03. Within the small family size simulations (ii), v-1), v-2), v-3), vi-1), vi-2), vi-3)), simulation ii) small family size had the lowest critical LOD score of -24.00 and simulation vi-3) small family size with 1.0 relatives had the highest critical LOD score of 9.00.

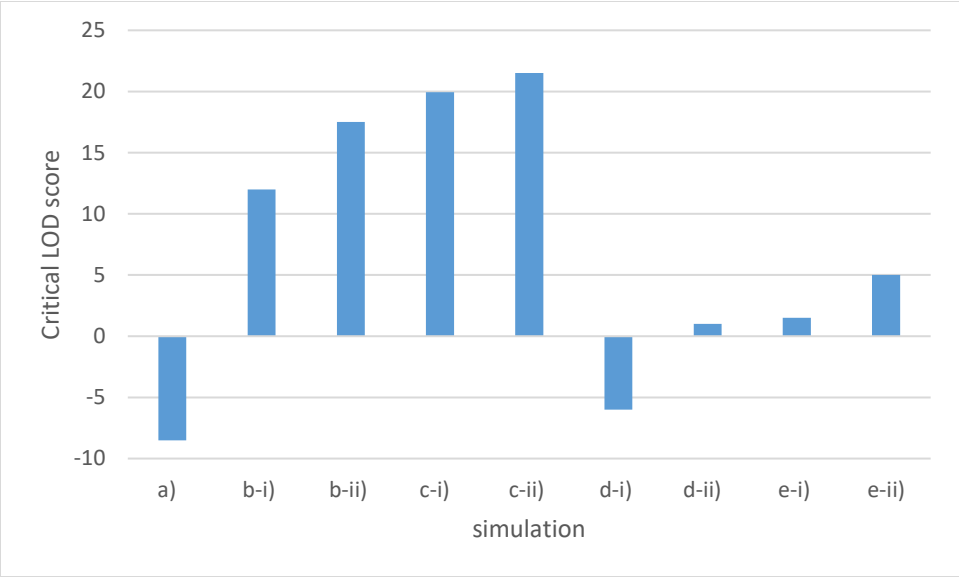


Figure 3: The critical LOD scores plotted against the practical application simulation sets where there is 200 females and 10 males. Simulation a) 200 females and 10 males have the lowest critical LOD score of -8.50 and simulation c-ii) 200 females and 10 males with 50% of females are full sibs and 50% of males are full sib has the highest critical LOD score of 21.50.

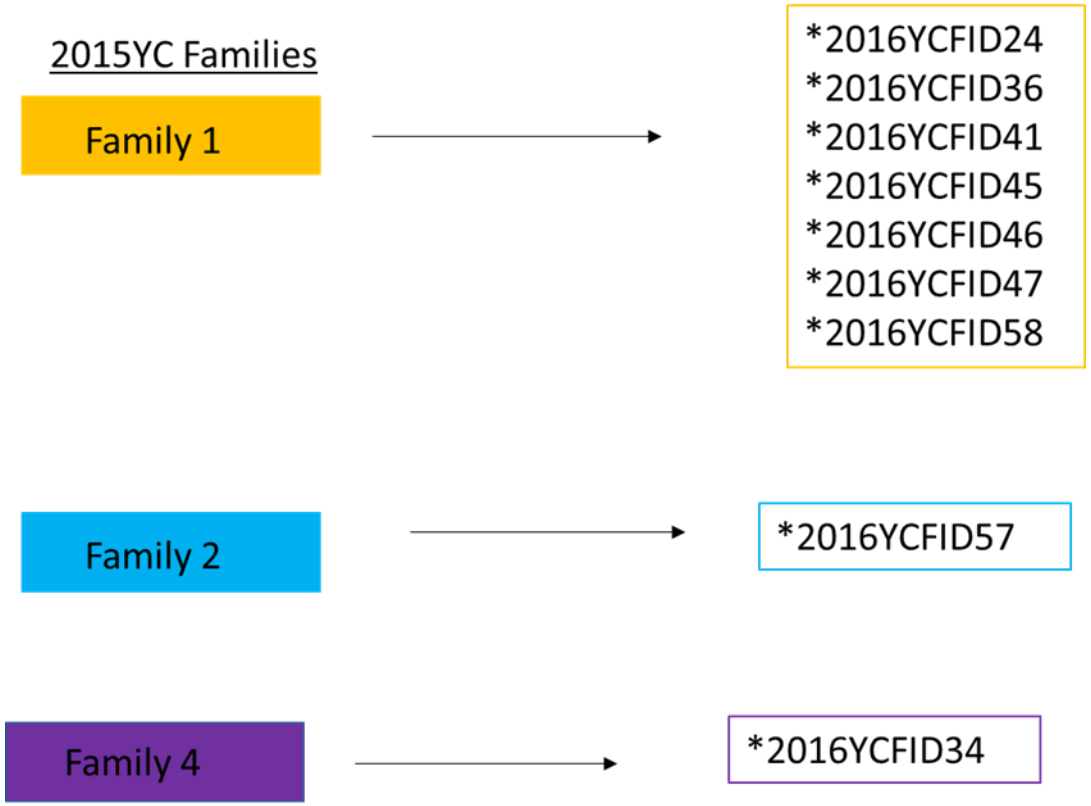


Figure 4: Cervus results when 2015YC was defined as the candidate parents and 2016YC was defined as the offspring. The results showed that 9 - 2016YC 'offspring' were solely related to 2015YC family #1, 1 – 2016YC 'offspring' was solely related to 2015YC family #2, and 1 – 2016YC 'offspring' was solely related to 2015YC family #4.

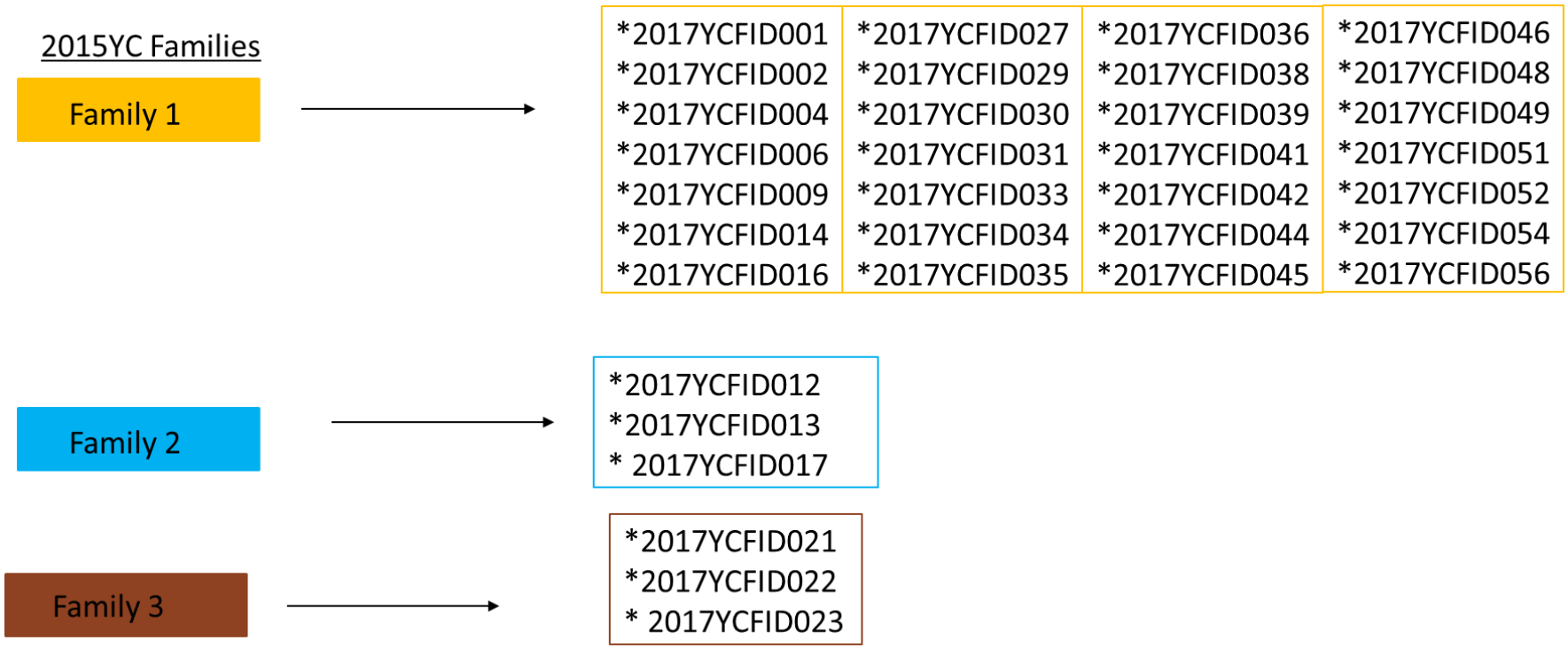


Figure 5: Cervus results when 2015YC and 2016YC were defined as the candidate parents and 2017YC/sea cage samples were defined as the offspring. The results showed that 28 - 2017YC offspring were solely related to 2015YC family #1, 3 – 2017YC offspring were solely related to 2015YC family #2, and 3 – 2017YC offspring were solely related to 2015YC family #3.

Table 8: The 17 different full sib families (defined in Colony) with number of individuals having PC2 data, ranging from 1 to 10. Only four of the 17 families had more than 3 individuals with PC2 data, family #5 had 10 individuals, family #10 had 7 individuals, family #1 and family #6 both had 6 individuals. Whereas, the other 13 full-sib families had 3 or less individuals per family with PC2 data.

Family 5	Family 10	Family 1	Family 6	Family 8	Family 9	Family 16	Family 19	Family 20	Family 21	Family 22	Family 24	Family 25	Family 7	Family 15	Family 17	Family 18
SLF003	SLF066	SLF055	SLF006	SLF051	SLF063	SLF089	SLF062	SLF093	SLF086	SLF088	SLF087	SLF002	SLF059	SLF073	SLF064	SLF057
SLF126	SLF068	SLF056	SLF090	SLF119	SLF081	SLF129	SLF103	SLF079	SLF114	SLF128						
SLF082	SLF069	SLF104	SLF077	SLF156	SLF095											
SLF078	SLF070	SLF130	SLF092													
SLF112	SLF080	SLF137	SLF127													
SLF105	SLF094	SLF154	SLF111													
SLF110	SLF121															
SLF113																
SLF107																
SLF157																

Table 9: The ASREML-R output of the variance components of the random effects when PC2 was the response variable. The residual error had a larger variance component then the animal effect

	Component	Std. error	z. ratio	Bound	% ch
Animal effects	0.1581543	0.3348035	0.4723795	P	0.9
Residual	0.9518992	0.3515920	2.7073970	P	0.0

Table 10: The ASREML-R output of the wald test for the fixed effects. None of the fixed effects were significant when PC2 was the response variable

	D.F.	Sum of Sq	Wald Statistic	Pr(Chisq)
Intercept	1	0.1197	0.1258	0.7229
Sex	2	2.0967	2.2027	0.3324
Cage	5	8.4592	8.8866	0.1137
mcTLmm	1	0.0386	0.0405	0.8404
Residual		0.9519		

.

Table 11: The EBVs calculated using ASREML-R and the defined model. Offspring SLF126 (belonging to family 5) had the highest EBV and offspring SLF055 (belonging to family #1) had the lowest EBV. Among the phantom parents, phantom parents #1 had the lowest EBV and phantom parent #5 had the highest EBV.

Animal	EBV for the natural food forging trait	Std. error	Z ratio
PS7	-0.037242725	0.3899477	-0.09550698
PD7	-0.037242725	0.3899477	-0.09550698
PS1	-0.134941357	0.3712866	-0.36344260
PD1	-0.134941357	0.3712866	-0.36344260
PS5	0.250850009	0.3615761	0.69376832
PD5	0.250850009	0.3615761	0.69376832
PS6	-0.074492177	0.3766735	-0.19776324
PD6	-0.074492177	0.3766735	-0.19776324
PS8	-0.037785987	0.3803339	-0.09934951
PD8	-0.037785987	0.3803339	-0.09934951
PS9	-0.044547161	0.3800917	-0.11720108
PD9	-0.044547161	0.3800917	-0.11720108
PS10	0.133291337	0.3746665	0.35575996

PD10	0.133291337	0.3746665	0.35575996
PS15	-0.018036527	0.3900921	-0.04623659
PD15	-0.018036527	0.3900921	-0.04623659
PS16	-0.054731647	0.3854769	-0.14198426
PD16	-0.054731647	0.3854769	-0.14198426
PS17	0.175454102	0.3899649	0.44992278
PD17	0.175454102	0.3899649	0.44992278
PS18	-0.109196769	0.3896641	-0.28023308
PD18	-0.109196769	0.3896641	-0.28023308
PS19	-0.075387206	0.3882397	-0.19417697
PD19	-0.075387206	0.3882397	-0.19417697
PS20	0.025156905	0.3853194	0.06528845
PD20	0.025156905	0.3853194	0.06528845
PS21	-0.052046220	0.3857598	-0.13491872
PD21	-0.052046220	0.3857598	-0.13491872
PS22	-0.006832352	0.3854075	-0.01772760

PD22	-0.006832352	0.3854075	-0.01772760
PS20	0.025156905	0.3853194	0.06528845
PD20	0.025156905	0.3853194	0.06528845
PS21	-0.052046220	0.3857598	-0.13491872
PD21	-0.052046220	0.3857598	-0.13491872
PS22	-0.006832352	0.3854075	-0.01772760
PD22	-0.006832352	0.3854075	-0.01772760
PS24	0.013310727	0.3913721	0.03401041
PD24	0.013310727	0.3913721	0.03401041
PS25	0.047177049	0.3895876	0.12109485
PD25	0.047177049	0.3895876	0.12109485
SLF002	0.094354099	0.3699626	0.25503683
SLF003	0.243327858	0.3456938	0.70388259
SLF006	0.081177848	0.3695899	0.21964303
SLF051	-0.141205574	0.3618370	-0.39024632
SLF055	-0.266849709	0.3555867	-0.75044897
SLF056	-0.118141737	0.3558326	-0.33201496

SLF057	-0.218393538	0.3702848	-0.58979877
SLF059	-0.074485451	0.3714770	-0.20051160
SLF062	-0.068524903	0.3756157	-0.18243354
SLF063	-0.083840100	0.3623645	-0.23136955
SLF064	0.350908205	0.3715495	0.94444549
SLF066	0.110719054	0.3629735	0.30503341
SLF068	0.166316584	0.3636206	0.45739039
SLF069	0.223103517	0.3626186	0.61525666
SLF070	0.140832955	0.3625680	0.38843181
SLF073	-0.036073055	0.3720830	-0.09694894
SLF077	-0.040935511	0.3590139	-0.11402208
SLF078	0.217077638	0.3460882	0.62723222
SLF079	0.027271816	0.3681738	0.07407322
SLF080	0.171034914	0.3579621	0.47780170
SLF081	-0.044911773	0.3624275	-0.12391933
SLF082	0.274652105	0.3489990	0.78697100
SLF086	-0.097248218	0.3721551	-0.26131103

SLF087	0.026621453	0.3774215	0.07053507
SLF088	-0.015778539	0.3720089	-0.04241441
SLF089	-0.059096186	0.3720468	-0.15884074
SLF090	-0.100549098	0.3695804	-0.27206284
SLF092	-0.158903283	0.3671350	-0.43281976
SLF093	0.048198900	0.3677574	0.13106167
SLF094	0.112486305	0.3594452	0.31294425
SLF095	-0.049436771	0.3631872	-0.13611925
SLF103	-0.157636714	0.3744576	-0.42097349
SLF104	-0.171731732	0.3559053	-0.48252085
SLF105	0.253768664	0.3474474	0.73038001
SLF107	0.228434148	0.3495807	0.65345184
SLF110	0.238593491	0.3469577	0.68767306
SLF111	-0.139171040	0.3615551	-0.38492341
SLF112	0.197266611	0.3461847	0.56983060
SLF113	0.260587168	0.3470517	0.75085982
SLF114	-0.058890442	0.3668948	-0.16051043

SLF119	-0.059166871	0.3638672	-0.16260566
SLF121	0.141837365	0.3633064	0.39040698
SLF126	0.507919275	0.3468072	1.46455795
SLF127	-0.163064155	0.3575168	-0.45610212
SLF128	-0.004718517	0.3657386	-0.01290134
SLF129	-0.105098753	0.3663488	-0.28688166
SLF130	-0.083453703	0.3561085	-0.23434910
SLF137	-0.138564902	0.3567685	-0.38838885
SLF154	-0.165847716	0.3565102	-0.46519769
SLF156	0.049228495	0.3630705	0.13558936
SLF157	0.337723137	0.3468167	0.97377995

REFERENCES

Anderson, E.C., and J.C. Garza. 2006. The power of Single-Nucleotide Polymorphisms for large-scale parentage inference. *Genetics*, 172(4): 2567-2582.

Beacham, T.B., C. Wallace, C. MacConnachie , K. Jonsen, B. McIntosh, J.R. Candy, R.H. Devlin, R.E. Withler.2017. Population and individual identification of coho salmon in British Columbia through parentage-based tagging and genetic stock identification: an alternative to coded-wire tags. *Canadian Journal of Fisheries and Aquatic Sciences*, 74(9): 1391-1410

Berry, D.P., J.W.M. Bastiaansen, R.F. Veerkamp, S. Wijga, E. Wall, B. Berglund and M.P.L. Calus. 2012. Genome-wide associations for fertility traits in Holstein–Friesian dairy cows using data from experimental research herds in four European countries. *Animal,* 6(8): 1206-1215.

Brumfield, R.T., P. Beerli, D.A. Nickerson, S.V. Edwards. 2003. The utility of single nucleotide polymorphisms in inferences of population history. *TRENDS in Ecology and Evolution,* 18(5): 249-256.

Busack, C., and C.M. Knudsen. 2007. Using factorial mating designs to increase the effective number of breeders in fish hatcheries. *Aquaculture*, 273: 24–32.

Cooke Aquaculture Inc. 2019. Unpublished PowerPoint: Cooke Aquaculture Inc. Cleaner Fish Program. Presented at the NAIA/Global Cleanerfish Workshop 2019, September 25th 2019, in St. John's Newfoundland Canada.

Dupont-Nivet, M., M. Vandeputte, P. Haffray and B. Chevassus. 2006. Effect of different mating designs on inbreeding, genetic variance and response to selection when applying individual selection in fish breeding programs. *Aquaculture,* 252:161-170.

Dussault, F.M. and E.G. Boulding. 2017. Effect of minor allele frequency on the number of single nucleotide polymorphisms needed for accurate parentage assignment: A methodology illustrated using Atlantic salmon. *Aquaculture Research,* 1-5. https://doi.org/10.1111/are.13566

Eliasen, K., E. Danielsen, A. Johannesen, L.L. Joensen, E.J. Patursson. 2018. The cleaning efficacy of lumpfish (Cyclopterus lumpus L.) in Faroese salmon *Salmo salar L.* farming pens in relation to lumpfish size and seasonality. *Aquaculture*, 488: 61-65.

Eriksen, E., C.M.F. Durif, and D. Prozorkevich. 2014. Lumpfish (*Cyclopterus lumpus*) in the Barents Sea: development of biomass and abundance indices, and spatial distribution. *ICES Journal of Marine Science*, 71(9): 2398–2402

Erkinharju, T., R.A. Dalmo, M. Hansen, T. Seternes. 2020. Cleaner fish in aquaculture: review on diseases and vaccination. *Reviews in Aquaculture*, 13:189–237.

Falconer, S. and T.F.C. Mackay. 1996. Introduction to Quantitative Genetics, 4th Edition. Essex, England: Longman.

Fernando, R. L., and M. Grossman.1989. Marker assisted selection using best linear unbiased prediction. *Genetics Selection Evolution*, 21(4):1-11.

Fiumera, A.C., B.A. Porter, G. Looney, M.A. Asmussen, J.C. Avise. 2004. Maximizing offspring production while maintaining genetic diversity in supplemental breeding programs of highly fecund managed species. *Conservation Biology*, 18: 94–101.

Fluidigm. Genotyping with the 96.96 IFC Using SNP Type Assays or the SNP Trace Panel Quick Reference. Fluidigm Corporation. March 2015.

Frankham, R. 1995.Effective population size/adult population size ratios in wildlife: a review. *Genetics Research, Cambridge*, 66: 95-107.

Frankham, R., C.J.A. Bradshaw, B.W. Brook. 2014. Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses*. Biological Conservation*. 170: 56–63. doi:10.1016/j.biocon.2013.12.036. ISSN 0006-3207.

Gjedrem, T., N. Robinson, M. Rye. 2012. The importance of selective breeding in aquaculture to meet future demands for animal protein: A review. *Aquaculture*, 350-353: 117-129.

Havenstein, G.B., P. R. Ferket, M. A. Qureshi.2003. Carcass Composition and Yield of 1957 Versus 2001 Broilers When Fed Representative 1957 and 2001 Broiler Diets. *Poultry Science ,*82:1509–1518

Hayward, C.J., M. Andrews, and B.F. Nowak. 2011. Introduction: *Lepeophtheirus salmonis*—A Remarkable Success Story. In Salmon Lice: An Integrated Approach to Understanding Parasite Abundance and Distribution, First Edition. Simon Jones and Richard Beamish, editors. John Wiley & Sons, Ltd.1-28.

Heaton, M.P., G.P Harhay, G.L. Bennett, R.T. Stone, W.M. Grosse, E. Casas, J.W. Keele, T.P.L. Smith, C.G. Chitko-McKown, W.W. Laegreid. 2002. Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle. *Mammalian Genome*, 13: 272-281.

Heaton, M.P., K.A. Leymaster, T.S. Kalbfleisch, J. W. Kijas, S.M. Clarke, J. McEwan,

J.F. Maddox,V. Basnayake, *et al.* 2014. SNPs for Parentage Testing and Traceability in Globally Diverse Breeds of Sheep. *PLOS ONE*, 9(4): 4851.

Henderson, C. R. 1975.Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 423-447

Igboeli, O.O, M.D. Fast, J. Heumann, J.F. Burka.2012. Roles of P-gylcoprotein in emamectin benzoate resistance in sea lice. *Aquaculture*, 344:40-47

Imsland, A. K., P. Reynolds, G. Eliassen, T. A. Hangstad, A. V. Nytro, A. Foss, E. Vikingstad, T. A. Elvegard. 2014. Assessment of growth and sea lice infection levels in Atlantic salmon stocked in small-scale cages with lumpfish. *Aquaculture,* 433: 137–142.

Imsland, A.K., P. Reynolds, G. Eliassen, A. Mortensen, O.J. Hansen, V. Puvanendran, *et al.* (2016a). Is cleaning behaviour in lumpfish (*Cyclopterus lumpus*) parentally controlled? *Aquaculture,* 459: 156-165.

Imsland, A.K., P. Reynolds, A.V. Nytro, G. Eliassen, T.A. Hangstad, *et al.* (2016b). Effects of lumpfish size on foraging behaviour and co-existence with sea lice infected Atlantic salmon in sea cages. *Aquaculture*, 465:19-27.

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

Jones, O. and J. Wang.2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*,10: 551–555.

Kalinowski, S.T., M.L. Taper and T.C. Marshall.2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, 16: 1099-1106. http://dx.doi.org/10.1111/j.1365-294x.2007.03089.x

Liu, Z.J. and J.F. Cordes. 2004. DNA marker technologies and their applications in aquaculture genetics. *Aquaculture,* 238:1-37.

Lommen, S.T.E., K. G. Koops, B. A. Cornelder, P.W. de Jong, P.M. Brakefield. 2019. Genetics and selective breeding of variation in wing truncation in a flightless aphid control agent. *Entomologia*. DOI: 10.1111/eea.12810

Lorance, P., R. Cook, J. Herrera, L. de Sola, A. Florin, & C. Papaconstantinou. 2015. *Cyclopterus lumpus*. The IUCN Red List of Threatened Species 2015: e.T18237406A45078284. Downloaded on 24 July 2019.

Marbase Cleanerfish Ltd. Environmental Preview Report for Marystown Marbase Cleanerfish Hatchery Project. Marbase Cleanerfish Ltd. October 2020.

Marshall, T. *Cervus* Reference Guide. Field Genetics Ltd. Accessed October 11, 2021 http://www.fieldgenetics.com/pages/aboutCervus_Overview.jsp

Marshall, T.C., J. Slate, L.E.B. Kruuk and J.M. Pemberton. 1999. Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, 7 : 639-655.

McBriarty, G. 2019. Unpublished PowerPoint: Cooke Aquaculture cleanerfish program development: Land-based growout and transportation. Presented at the NAIA/Global Cleanerfish Workshop 2019, September 25th 2019 in St. John's Newfoundland, Canada.

Milla, S., A. Pasquet, L. El Mohajer, and P. Fontaine.2020. How domestication alters fish phenotypes. *Reviews in Aquaculture,* 13: 388–405.

Monk, J. 2019. Unpublished PowerPoint: Memorial University's Cleaner Fish Production and Research Program. Presented at the NAIA/Global Cleanerfish Workshop 2019, September 25th 2019, in St. John's Newfoundland Canada.

Morrissey, M.B. and A.J. Wilson. 2005. The potential costs of accounting for genotypic errors in molecular parentage analyses. *Molecular Ecology*, 14: 4111-4121.

Nielsen, R. 2000. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics*, 154: 931-942.

Norðberg, G., A. Johannesen, R. Arge. 2015. Cryopreservation of lumpfish *Cyclopterus lumpus* (Linnaeus, 1758) milt. *PeerJ*, 3: e1003; DOI 10.7717/peerj.1003

Pampoulie, C., S. Skirnisdottir, G. Olafsdottir, S.J. Helyar, V. Thorsteinsson, S. Jonsson, A. Frechet, *et al*. 2014. Genetic structure of the lumpfish *Cyclopterus lumpus* across the North Atlantic. *ICES Journal of Marine Science*, 71(9): 2390-2397.

Pountney, S.M., H. Migraud, A. Davie.2020. Short term cold storage and sperm concentration assessment of lumpfish (*Cyclopterus lumpus. L*) Milt. Aquaculture. *Aquaculture,* 529: 735646

Powell, A., J.W. Treasurer, C.L. Pooley, A.J. Keay, R. Lloyd, A.K. Imsland and C.G. de Leaniz. 2018. Use of lumpfish for sea-lice control in salmon farming: challenges and opportunities. *Reviews in Aquaculture*, 10: 683-702.

Premachandra, H.K.A., H. Nguyen and W. Knibb.2018. Effectiveness of SNPs for parentage and sibship assessment in polygamous yellowtail kingfish *Seriola lalandi*. *Aquaculture*, 499: 24-31.

Raoul, J., I. Palhière, J. M. Astruc, J. M. Elsen. 2016. Genetic and economic effects of the increase in female paternal filiations by parentage assignment in sheep and goat breeding programs. *J. Anim. Sci*,94:3663–3683

Rosa J.O., G.C. Venturini, T.C.S Chud, B.C. Pires, M.E. Buzanskas, N.B. Stafuzza, G.R. Furquim., V.A.R. Cruz, G.S. Schmidt, E.A.P Figueiredo, V.F.M.H. Lima, M.C. Ledur, D.P. Munari. 2018. Bayesian inference of genetic parameters for reproductive and performance traits in White Leghorn hens. *Czech J. Anim. Sci.,*63, 230–236.

Strauch, O., J. Oestergaard, S. Hollmer, R.-U. Ehlers. 2004. Genetic improvement of the desiccation tolerance of the entomopathogenic nematode *Heterorhabditis bacteriophora* through selective breeding. *Biological Control*, 31: 218–226

Strucken, E.M., B. Gudex, M. H. Ferdosi, H. K. Lee, K. D. Song, J. P. Gibson, M. Kelly, E. K. Piper, L. R. Porto-Neto, S. H. Lee, C. Gondro. 2014. Performance of different SNP panels for parentage testing in two East Asian cattle breeds. *Stichting International Foundation for Animal Genetics*, 45: 572–575

Trut, L., I. Oskina and A. Kharlamova.2009. Animal evolution during domestication: the domesticated fox as a model*. BioEssays,* 31:349–360.

Vandeputte, M. and P. Haffray.2014. Parentage assignment with genomic markers: a major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. *Frontiers in Genetics*, 51(432): 1-8.

Waldron, D.F., C.A. Taylor, J.W. Walker, E.S. Campbell, C.J. Lupton, T.D. Willingham and S.Y. Landau. 2009. Heritability of juniper consumption in goats. Journal of Animal Science, 87: 491-495.

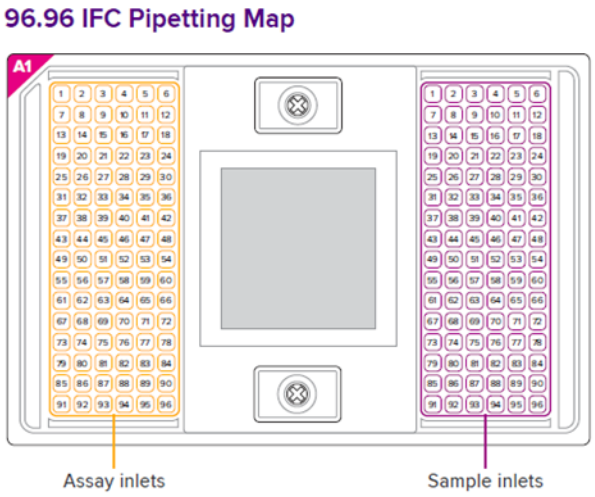
Wang, J. Institute of Zoology webpage for COLONY. Accessed October 12, 2021 https://www.zsl.org/science/software/colony

Whittaker BA, Consuegra S, Garcia de Leaniz C. 2018. Genetic and phenotypic differentiation of lumpfish (*Cyclopterus lumpus*) across the North Atlantic: implications for conservation and aquaculture. *PeerJ,* 6: e5974 http://doi.org/10.7717/peerj.5974

Yue, G.H. and J.H. Xia. 2014. Practical Considerations of molecular parentage analysis

in fish. *Journal of the World Aquaculture Society*, 45(2): 89-103.

APPENDICES



Appendix 1: 96x96 integrated fluid circuit (IFC) used for SNP genotyping. The 96x96 integrated fluid circuit is loaded with the 96 primers on the left and 96 DNA samples on the right. When the circuit is loaded into the Biomark machine, 96 PCR reactions are performed consecutively on all 96 samples (Fluidigm 96.96 SNP Type Assay Protocol). When the circuit is analyzed in the Biomark machine, SNP genotypes are determined for each sample at all 96 SNPs based on the PCR products produced (Fluidigm, 2015).

Appendix 2: The chosen 96 SNPs for the 96SNP panel for genotyping with their 100bp flanking regions on each side of the SNP

SNP CONTIG/POS	PRIMERS with SNPs
jcf7180000027996_1416962	CTTACTGCTATAAGGCAGCATTTCTATTGATTGATATTGGGTTAAATAGATGAGAAGAAATGACTTTATTGCAGCACCAGGTCTGTCGTTTGGACATTTC[T/A]CTTGAGGGTGAAGAACTGCACCAGAGTCACTATATAATTCCACAGAGATTGCAGCAAAATTAAAAAGGGGGATGTGAACAATTTCCCATTATAAAAGTTG
jcf7180000027999_1568475	TTAGATTCTCATTTAGTTTTTACTTGCGTGACTTAAAACAAAAACTAATGATATAGAGCAATGTGATTTGGCACAAAAACATGTCAAGTGACACATGTAC[A/G]TGTAACCTACATGTAACAGGTTAATGCGGATTAAAGGATGGTGGAAAAATTGAGAGAAATAGCTGTACAAGAATAAGAGAGTAAGATAAATCTGAGGCTA
jcf7180000028000_36613	CACAAAGAACATGACTTCCATGACTCTCATTAAGGTGACCTCTCACCACCTCTCAGTAGAACGGAGTTGAGGCCTCTGCACACCAAGTTCATGTTTTTCG[T/G]ACACGATTTTTAAATCCGCTACGCACAGAAACCTTGCACACTGAGTCTGATGCGTTTGAATAACCTATTCATTGTGTGAATGTGCGATATGAGACAAAAC
jcf7180000028002_1631588	GAGTAGGCATGGATATTGTGGGTCCCCTGCCCAAAAGTGCCCGGGGCCACCAGTACATCTTGGTCATAGTGGACTACGCCACCCGTTATCCGGAAGCCGT[T/C]CCTTTAAGAAAGGCCAATGCTAAACAAATAGCAAAAGAACTGTTCCTCTTCAGCAGCCGCGTGGGGATTCCAAAGGAGATTCTTACCGACCAGGGGACCC
jcf7180000028006_153962	TGTGTTTAATGTAAATCATGTTTGCATCTCCACATCATCTAAACAAAGTATTGATCTGATCAACAATGGACTTAATACTGGATGAAACATGGTAAGATAT[G/T]GTCTTGGAAGTGTGTAACAAAATAATCGCTATGAATAAAAACAGCATTTTTGCTTTGTACAACTAAGAAAATGTTTACAGACAAGCAAAACTAAATGTAC
jcf7180000028023_247833	TATTTATGTGTGCATTGGAATGCATTTGTGAATCCTTCTGTGTGCATTTGTGAGTCTAATTTATTTGTAAATCCTTCTGTGTGCATTTGTAAATCGAACA[A/T]ATTTGTGAATTGTTCTGTGCACATTTGTGAATCTTTGTGTTTACATTTGTAAATGCACACTGTGCATTTGTGATTATTGAGACTGATCTGACCCCATATA
jcf7180000028031_421776	AACCGCTACGATTTCTAAGCCTATCAACATAATACTGTGAAACGGGATCTTACAACTGTGGAAAGTTGTCACGGTGACAAGTGATCTACGGGAAAACTGT[T/C]GAAATAAGCCATTCATTCACGTCATTAAGCAACTTATGCACCAGGAATCTGCATCCATTGCTGGATGGCGTCACGCTATATTGTGGCAAAAGTTAAGGTT
jcf7180000028033_270344	GCGAACAGGAGATGAGAGTATCCAGACCATCAGTTTTAAGACTTTAAGAGCTTGGCTCTCTGCTTTGAAGTTAAGGTCAGCACAAGTGTGAGTGAAAGTG[C/T]GTGTGTATTTGTGGGAGAAACACACATGCCTACTGACACACTCACACATGAACACTGTCATATAAACAGATTCAGCAATAGTTCAGATTGATGGCTGATA
jcf7180000028	GACTGTGTTGGGCTGTCGTAACTACGGCAACTGACTCCGGTCTCGCGATAGTTCCGCAGCGGGTTATGCTTTCCTCTCCCTCATAAATAATCTGCGTTTA[A/C]ACTACA

041_24 4011	GTTTATGAGTTTACCGCAGACCGTTTTTATTTATGAATCTGTTGGTTTGTTGTTGT GAACTTTTCCCTCCATGATTTTGTTGATTTATCTTACA
jcf7180 000028 043_11 27979	TTGAGAACAGCATGTTTGAAGAAGAGCCGGATGTTGTGGATTTGGCCAAAGACT CTGCTGCGTTCCCAGTATGCGATCATGTGATCCTATGCCATGCATG[C/T]TATACA TGTGTTAATTTTAAGGACCAGCTTGGCTCAATAAATTGCATTTCATCGAAAGCATA TTTCCAGTTATTATTTACATTACATTGTAAACACTGCA
jcf7180 000028 047_11 0248	GCATTAATAACCGTTTACAACAGCAATCAATATGACAAGATCAAGTAGAGTATAAA TAATACAATGAAATCAATATAATTTATTTAGTTTTCCCACCAGA[G/A]GGGGAAAAC AGCCTCTGTTGACAAGCTGATTGACACGTCGCTGTGTCAATCATGCCTCCAGGT GCCGCTGTTTGCAGAAAGCAAACGCAACAAAACAGAT
jcf7180 000028 057_70 3958	AATAAACTGAATTGAATTGATGTAAGCTGACAGACTGAAAGACATTGCCAGGTTTT GATTTTGTTCATAGTAAAGACCCTTCATGGCATTCAAGCAACTC[T/C]GTGATTGC TCAAATTATACAAAGTGCATCATCTGCTGGCACTGAAACTTGTCAGCATTTTTAGA CTTCAAGAGAAATAAAACCATGCAGTTATAGGACAC
jcf7180 000028 061_60 4071	CTCATAAAGACCCACTGAATTAAACGACATTTACATTAACATGATTACATAGTTAC ATACATTATTTACTGTTGGACTGCAATCCCTTCAGCATAAAACT[T/A]CTTTAAAGG TATAGGATGCAGTACGATAATACACATTGTCAAATACAACATCACACTTTGATATT AACTGTAAACAATTGAGAAGCTAGTGTTTGTGACA
jcf7180 000028 074_50 2167	TGGGGCTGAATCAAAATAAAAGTAAATGCAAGCATAAATACTCTGGTGCAATGGA TGGATCTAAATGTGTTCAATGTGTGCTCCTTGGTAAAATCACAAT [A/G]GAGTGCCACTGAGGTTTCCTAAGCCACACTTGCAATTATTTTGGATCGTCTT GTTTACATCCATAATGAACCAACGGCAACATTACATCCCTCAAGTAAG
jcf7180 000028 084_13 29220	TGAATTCACCAAGAGTAACACGTGAATAAAAGACTAATGTATGTTAGTATTCATAA GTGTGTTCTTTTGTTTCGGATTTAACAGTAGTCGACATGTCCTC[G/A]TACAACAG TTCGTATGATATCTTATGAAAAGTTATTCAACATTTGAACAATTTCCATACTAACAG CCTTTTCAAAATAAACAACTTGGTTAGGTTTAGGG
jcf7180 000028 091_13 4765	CGAAGATATAAAGATATTCAGGAGAGATTTCTGTTGTCTTTGTGCTTGCAAAGTCT CTCAAGAGAAGGCCAACAGCATTTGTACATGTGTTTGTCAAACC[A/G]TAGAGCTG ATTTCCTCGATCTATTTTTAACAATGCATAAAATGGCCATGTGAAAACAATGGGAC AATGTGAAAGGGGTCACTCATAGTAATAACACTACA
jcf7180 000028 102_10 50587	ACATACAGCGTCTAAGCAGGTTATGTTTAGCGAGTATACTTTCTACGATCTTCTAA CAGTATATTAACTCTCCTCACCCCCATTTCTCTTTCCGTTCCAA [T/C]AGGGTGTGATGGAGAGTTTCCTGGGAACAGCCCTGGCTGGCTCTGTCTTCT GTCTCTTCAGCGGTCAGCCCCTCATTATACTCAGCTCCACTGGACCCAT
jcf7180 000028 122_59 7590	ATGACTGCTCGGAGTCCAAATAGCAGGGAGATTTGGGTGATTGGAGATGGAAGG AAGACGGAACAAAATGACGAAGGAGAGCTTGTCAGATCAAGGGGGA[T/C]GCAAT GATCGGATAATGAAGAATTAGAAGTCAACATGACTATGATTTCAGGGAAATTCAA GTGCATGACAAATGCATTAAATGTTCACACTTCCAGGACA

jcf7180000028126_463279	GACCTCACCCGCTGACACACCAAGACAACACTGGTCGGAAGGGAAGCTGCAGTTCCCTTATGGTCAGATTGATTCTCTGGAATACTGGGAAAGAAGCACT[C/T]CAGCCCCTTTATGTGAAGCCAGCTCTGGTGTAGGATTAGAGCCAGTCACTGTGGTTACCGTCAGTCTTAACTGTACAGAAGTTGTGCAGATGTATTTTAC
jcf7180000028136_188534	AACTCTCGGGCAGCAGCTCCATGCTGGCAGCGGGCAAGACGGCCAACAGCTGCTGGAAGGGCATGAAGGGTTTCTCAAGGTCAAAGGTCAACTTTAACCC[C/T]GATATGTTCTTGATGTCGGAGAGGAAGGGGGCGTAGTGGTAGGGGTAGTACCTGAACAGACAGGCACACATTATTAAATATGATTTACATATTGAACAAA
jcf7180000028148_82326	GTGCTTTTCCATTTTGATGAGCGTTATCAGATGTTTCTTTTTATTTGATCCCCTGTTCAAAAATCATTCACAGAGAAGGAGTGAGGGCCATGGCTAAGCA[T/A]TGTTGTTGTGCTTGGTTCGAGCAACCCCCAGCACCGGACTGAAGCCACGAGTGAAACGTCTCGTACGGCCTGTCGAGATGGACTGCACAGTCTGGGAATG
jcf7180000028151_41545	CACTGAAAGGATGCATCAGAGCCCTTGTGCACCATTTGACAGCCCTCTTGAGTGTGCTGTTTAGTGTCCCACAATGCAATGCAGAAAGGAAAGGGAGGAG[T/C]TTATCACAGCTGGAAAAGATTAAAAACAAATTAAGAAACCAAACACTATAAATATTGTATTCTCTGTCTCTCTATATAAAAATGTGGACTCTTCTATGGG
jcf7180000028160_524267	ACAGCAGTAACTCATGATATTTTATGTCATGCATTAATTATATAACTTTCTAAATTACTCAGTAACACTAAAGTCTCACTTTATTCTGTTGCCAGAAGGT[T/G]GGGACTCTACTTCTTATGTAATCACAGCTGTAAATGAGAACTAAAAAGCATATCTGAATGTCAGAAATGTAATTTGTCATTTTATTCTACAAATATAAAA
jcf7180000028168_718255	ACGTGGGCAGACATTACACAATGCACCTTTCAACCACTCAATGCAGTGATACTACTTTGTTATCCTACTACGTATGATATTACCATAGTATTATATTAAA[G/T]ATGCATTAATATGTAAGAAGCATTTTACTCTTGCGGTTGGTCGAGGTGCATCATATCTTAGAATCTAATTTGATATTTTATAAGCAAATCTTGTTTTGTA
jcf7180000028174_68139	AGAGGTAGTTAGCCAGGCCTAGTTATGTTTTGTTAATTTACATTAGTTTGGCAAGAGCTGAGCCTTCAGTTATCAAGCTCCTCTTTTATGGAACCAGCTT[C/T]CACTTTCAGTCCAGGAGGCAGACACAGTCACCTTATTTAAGAACAGACTTAAGACTTTCCTCTTTAATAGAGCTTATAATTAGGGCTGAATCAGGTTTGC
jcf7180000028178_313522	CCTGAAACCCCCTCCCGTGTGAGCCAATTAGGGATGTAAAGATAATGTATATTCATTAATTCAAGTATTCACTTGAAAAACAATACTGGAGTTCATAAAA[G/A]AAACACATTTACCTCATTCACCTCCAAAAAGTTAGACATTTCCTCCAAACCTAATGCAAACTGTTAATGCATTTCTTGTTTTAGTTGCATTCATGTTTTG
jcf7180000028184_362747	AGAATGAGAAGTACACAGTTTACCTCTATCTGCCTGTTAAGGTACACTTATATCTGTCACGTTGCTCCCAGATGTGGGTTTGGATCCCTGATTGTTTTCA[G/C]CAGCATCGCTAAGCTGAAGGGATTGAGTGAATGATTCCATTCAGTCCTTGGATCATGGCACAGGGAATCCAAAATGGCACACAGTCACCTCATTTCTGCT
jcf7180000028214_39	TCCAAGCTTGTCAAAAACCCACCGTTTGGTCAGTGGCTCGCTGGGTCTGTTTCCTTTAGCATCTTTCAACTAACTCGATGTAAACCCATCCGTCTCAGGA[A/G]AAGAGTTTGGTTTCACCCAGGAAACCGCTGTTTGTGTCCCGTGTGAAAGCAAGTCAACATC

5096	GACTTCCTTTACCTTCCTTTTGTTTCATTACTGACTAAC
jcf7180000028276_120881	AAGTGTTAAAATACGAAAGGTTTGGAAGTGTGTCAGAACCGGAGCCCATTGGTCAGTGGATGGTAACTGGCAAACAGCCTTGTGTTGGTCTGGTAGTGAG[G/A]AAGTGTAGCATAAAGACGGATAGCTCAGACTGAGGTATCCGAGTTGAAGTGAAGGACACGGGCTCCATTTGTTTTGTGCTATTTTTAAACATAATGTAAC
jcf7180000028297_243746	TTTCGCAAACATTACCTTAAAGGACGGCGACGTTAACTCTAGTGGCAGTATTCCATTTCGCAAACAAAGGCTTTACAACTACAGAATAGGCAATTACCGT[C/T]ACCATCAATTATCAAAAGCACAAAACAGATGCAACAATTGTGACGCTACATTAGCATGTCAGTATCTAAAAGTCATTACCTGGAAACGACAACGCCGTTG
jcf7180000028304_207685	TGTTACAAAAAGGTTGAACATGTTTAATGCCGCGGAGCGAAAGATACGCAGCGTCCGCACGTTGTGTTAATGCAGATATGTGTAAACCGAGCTTTGCTGG[A/G]TGACAGCTGCGGTAGTTGGGAGCGATGTCAGAGGGTCATGCTGAGTGTCAGTTAAATGGCTGTCGAGTTCCACGCGGGAAATTAAATAGCCGGTGCATGA
jcf7180000028337_97839	ATATACATTAATCATTTTAAATGCTCGTCTCTGAGTCTGACCGACAGAAAACACAGAAAATAGCTGCCTGGAACGGCCATTACAGCTAACGGCACAGAGA[C/T]CTACATCCCTACATCCGTCACTGCAGGAAAAGTTAGAACATTAAATATGAATAAAATGTGCAAACATGGTTCAGGCAAGGATTGCTATCAGTTTCTCCCC
jcf7180000028353_103572	ACCTGTTTGATGCTTCACGGTGTAAAGATGGCCAACTGGTGAAGACCGAAGACAGACCTATGCGTCCTCCGGTCCGGCTCGATGGTCCGACGTAGAGAAA[T/C]GCACCGTCTCCCTCCTAAGAGGGACAACTTGTCGCAGCCTTTGGAGGGCGAGAGACCCGCAACTCCGACCAGTGGTCACAAGGCTACACGCGACTTCAGA
jcf7180000028366_68458	GCTGGAACAGTTCATTCATTTAGTGTTTTCTGTTAAAATCCAAATGGGTTTTGGCTATTGACCTCTGGCTTTTTATTAAAACCGTCTGTGTATAAATACA[T/G]CTATGAAACCCATTCAACTCCAATTCAGTTAAAATGTTTGTAGAAGCGATGAATCAGATTTCAGGTCGCTTTAGTTCCAACTGACAGCCCACACTACAAA
jcf7180000028379_58837	AAGTCCGACTGTTAGAAGTCTATGAGAAAATGACCACATCTCACCTTTATTTTATCAGTAAACAATGTGAACATGCACCTATTGTTAGTCACTTTGGATA[A/C]AAGTGTCAGCTAAATGACATGTAATGTAATTTAACATGAGTTTATTTATTGTCTCAATTGCTAGTTTCAAGTATTTTTCAATCAACGGTTGGTATGATAT
jcf7180000028412_96282	CGTAATTTCTTTAAGGAGCGCAATAGTTGCAGCTCTATGACAGCATCTAAGCTGCAACCCCACCATCTCCCTCCGCCTTTATATCTGCACCTCTATCAGA[G/A]AACATGTGGACAGAGGCTCAATGAGAGAGGGCACGGAGAACAGGCCGGCTGAGGAGGGAGGTCATTTCTGCTGTCTCACCTCAGCGCTTTGTTCATTCAA
jcf7180000028475_857541	GTGGGTCAGCGGTTAGCAGGTCCGTCTTTCAATCAGAGGATTGGCGGCCCGCTCTAGTCGATGTGTCCTTGAGTAAGACACTTAAACCCAAATTGCTCCC[C/T]GTAGCTGTGTCTACGATGTATGGTTGTAAGTTGCTTTGGATAAAAGCGTCATGTAAATGTGTCTTGCCCAAGGACACATCGACATGGTCTTGCGCTCCGC
jcf7180	GAGCAGCAGATTTAAAGACTTTTTAAGACGTTTTCCTGAACAGGAAGTGAGGGGT

000028480_237326	TTTCAGTTGCCACCATACACTGGAGAGAAGGTGTTCGTGAGACTT[T/G]TTTAGTTTGACGGTCTAAGATTTTGGAACCTTGTGATAACAGTTTGAGGAAATGGTACACATTTCAGAAAATATGTCTATGTGGCTGTGAAGAGCTATAA
jcf7180000028498_214895	ACAGAGGCCGTTCCTTGAACTAATCTGCCCGTGTGCCGGCCTTGATTCAGCGAGAAGCCACAACTGATCCAAGCCGGGTATTTCGTAACGAGGTTATCCC[A/G]ACGATTCCTTTCTGAAGACAGGTTTATTGTCTTCTCTTATACAAATCCTCAGGGTGTGTACTTAATGCCACACCCGCGTGTGGATTCAATCTCATTTGAT
jcf7180000028507_61462	AACAGTGTTCCTAGATGGCAGGGAGATTTGTAAAAAGATACATCAACTGACCACTATTGCAGAGATTCTTGATGCCTGCAAAAGTTAATTGACTTTGATA[G/T]ACTTCTAAAATGTAACACAGATTGTAATGAATTGATTGACACAGACATCAATGGTAATGTCCACAACTTTGACAAATATTACATTCTGTTCCAGTCAATC
jcf7180000028543_102539	CATAATGTCTCTTCACTAGAGGCTGCAATAATCCTCAGACTTTCACTTTTCAGACTAGTGCCAAACCTTTCAAGTTGTGATACTATTTGCAAAGACATTC[A/G]ATGAACAGTTGACAGCTAGTACTGCCCACCCAACGGAGTTTTGCTTCAGTGCATCATATTTTGATTTATATTGAAAAAGATAAAATATTCATTTGAATGC
jcf7180000028598_288560	TTATTAAATGTTTCTGCGTCTGGGAGGCGTGTTCTTGCGTTTAAATTCTTGCGAATAAATTCTTTGTAGCATAACTCAACATCTTAAAACAGTCCCAGGA[G/T]TCTGTATACCGATAGGCTGCAGGTTTCTCTCTGGCGGCCACACTTTGACGCTAAAAGCTAAACCATCCCCAACTTTTCATTTCAGCTCGCTGCTGATCTG
jcf7180000028624_123250	AGGTTGAACACCAGTCGAGCAGCTGCATTCTGGATGAGCTGTAGAGGTCGAATGGCATTAGCAGGTAGACCTGCCAGGAGGGAGTTGCAATAGTCTAGCC[A/G]TGAGATGACCAGAGCCTGGACCAGAACCTGTGCCGCCTTCTGAGTGAGAAGAGGTCTTATTCTCCTGATGTTGTACAACATGTACCTACAGGAGCGTGTT
jcf7180000028782_159441	GTGTATTAATGAAACTGACTGACAGTCAGATTAATTTCTGTTGGCTGACACACCTGCTTTATGGAGCCAAGACTGAAAAGAACAGCCTCAGGGGACAGCA[G/A]GCTGCACAAGGATTTAACCACAGTTGGGAAACATTACAGACAGGATCGCAGTTTAAATACGGGCCCCGTGAAGGAAACTTCATTATTGGCAACAGATCAA
jcf7180000028800_37826	TGACCACGGCACTTCTCTTCCAAGAAGTGAAGAAATCTCTCCTCCAAGACCCGTTTGAGAGCAGAACTTTGGCGGAGAAACTACAGGTAAAGCAGCTGGG[T/C]CCAGATCAACCTCACAAGAAGACAACAGACTGGCTGAAAGGGAAGCGGCACAGGAAGCTTCTGTAGAGACCGGTGCACCCGGAAGGTTTGGCTAGCTGGA
jcf7180000028972_1494704	CGGGGCTCCTCCGAGTCGACGGGCAGTACGGCGGCGGGGACCGGGGCATGTCGATACCGGAGCACGGATTCTGCCAGCCCATCTCCATTCCGCTGTGCAC[T/G]GACATCGCGTACAACGAGACCATCATGCCGAACCTGCTCGGCCACACCAACCAGGAGGACGCGGGGCTGGAGGTGCACCAGTTCTACCCGCTGGTCAAGG
jcf7180000029032_198407	TTCCAAGTTTAAGAAAATATCAACATGATAAATACTTTTACAAAAATACCAAAGATGCTGTTGATAAATGTCTTTTCAGTTCGACCCTTTGACTCTGATC[T/G]CCATCTGAGCAGCATGACACAACATCACAACAACAAAAAGTGAACACCAACATAAGACCATCCACTCACCAATTAGGAGACTCCACTCAAGGTAATATGG

jcf7180000029116_131451	AAAAATTAGTAGAGTGATTTTACACAGTAGATTGTTTTTCTTTTACGTGTTTTCGTCTTGGACTTTTTGCCAGCAGAAACTCTTTAATTCTAATGGATTC[G/A]TTTTGCTCTTCCAGTAATTTAGGTGATGACAGGAGTCCAACTGCGGTGTTGTGTTCATCAGAGCAGGAGAGCTCTTACTCTGCCTTGTCCACAGCCCAGC
jcf7180000029204_266368	GCCATGTCCCATTTCCTGACTGTGATTTACACGCAACGGTAAGCACACTGGAAGAGGAGTTAATGTCACGGACGAGCAAATTACATGATGCAGTTTTTGT[A/T]TGCGGATCAAAAACCATGAGTGTTTTTGATTCGGGCGATACCGCTCTCGGTTTATTTCTCATCCACAACCGACACACAAATCGGCGCTTGATCATTTACC
jcf7180000029235_29829	GCTAGTTTGCTTCCCCGCGACCAGCTTGGCTCCGATTAGAAATCCAGCCAGCCGGAAGTCTGCAGACCTCCCGTCCGCTGGCAGCACGAGCTTCACACCG[G/C]TATGTGTAACGTGACGTCTCTGTGCCGTTTAGATGAAGATGTGCACAAGTGCGGACGCTGTCAGTCCGAGTTCTCCACGCTGGAGGCTTTCATCCAGCAC
jcf7180000029346_280766	ATGTAAGACAGCTGGTTGAGAAATCAATTACCATTGCATTACACCGGCTATATGTGGCCTCGTGACGCCACTGCTGCCGGTGGAAAGATAGCGCTGCAGC[C/A]CTCCTGTGGTTTAATGCTTTGCCTCTGCACTCTCTATCTATGGTGAAGATCCTGTTCTTTGTACAAGATAATACAGATAGTTTATAATGTGTATGTTATT
jcf7180000029371_54578	GCACAAAAGTAAGTTCAGTGCAGCAAATCTCTGTGGATTGCTCACATCTGATCGACACCTTCATGGTCAGTATTGGCTCTCAGTCTGACTGATTGAATCG[G/T]TGTCTTCCTACAGAGAATGGGTGTGTTGAGACGGAAAAGACAACTCTTATAGTAAAAATAGTGCCCATGTCATAGTGTTGGACTGATGTTTTTACAATAA
jcf7180000029531_42893	AAAGGCTCTGTCCTTTAGTTCTGTCCTTGTGATAACTGTATTCATTAGTTTGCAGTATAAATTGTAGGCCAAACCCATATTACCCCAAAACAATACGAAC [A/G] AACCTAGGTCTACTTCCCCAAACGAAGAGAACTGCAGGAAATCTTAAACTTCCAGTGCATCAAATCATTGACTAGCCTAGCCTGAACGAGGGGTGGCTCC
jcf7180000029592_7410	ACAATGGCCGTAATCCAAGATGGCGACGGCAAACTCGTTGAACTCGAGGCTTCGAAAAACCGCCGTCCCCGAAACCAACGGGTGTCATCGTCACGACGAC[G/T]ACGACCACCGTCATTGGTTCTATTCTATGTCGCAGAGCAACGTCTCTGCAGCTTGAGTTCAACGCACACTTTATTTCAGGAATCTAAACCAAACTTACTC
jcf7180000029733_15458	TTAATGTAAATCAATGTCAATATCTCGTATGCTGGTCCTCATGTTTTGCGTTATGTTGACTTAAATCTAGCTTGTTTATCCCGGTATTTAAAGTAATCCT[C/T]GTGTGTTATACGTCGTATTTTAATGCCTCGAAGTGTTTACTTCCTACCTGGCCTTCGCTGAGCCTCACCTCGTCGTCTTTGTCGCCACTCCGATCGTCGG
jcf7180000029973_46671	TCTTTCAGCACCTTCTGCTTCTCATCGTATTCTTCTTCAAGCTGGACTTCCATTTGCTTTAGCTGGACACACGGAAATAAAACCTCAGCAATGAGTACAT[G/T]CATAGACCGGTCTGATCCCAGTGGGACGGAAAGTAGAGCTCAACATTTATTATTATTTAAGAATATTAACAATTATGTTACATCCCTCTAGGTCCTGACA
jcf7180000030271_11	GTGGCTGCAGAGATTTTGGTTTGCTCAAAGCAACAAGTTGAATTCTGTAGCTAAAATAGTTCGGTGAGACGTGTGGGCTGAATTCACCGCCTGTCACCGG[G/A]GATGTCAGTTCTGCTTAATTAAATAACCCGACAGCCATTAAAAAGGTAACTAATCGGTTCAT

583	GTTTCAAGAGAACAAGAAACACAATACTGATGTGAAGC
jcf7180000031487_52697	AGGCTCAGCTGATTATGTGATGATGTGGTGAAAATCTAAAAGCTGTCTATATGTTCTGTTGAATTATAAGCTCTATGTATATGAACGCAGCAGAGCTTTT[T/C]CCATTAAACACTTTATTTCTAGAGTGCCTTGCAAACATTACTTTTAGTTTCCCGCGTTCACTTTATTTTGTTTACCTGTCGGTCATTGTTTACCTTCTCT
jcf7180000031959_6311	GTTCTAGAAAGACCTGAAGCCTTAAATAGAAAGACCTGAAGCCTTTAAGGGGTTCTAGAAAGACCTGAAGCCGTAAAGAGGTTCTAGAAAGACCTAAAGC[A/G]GTTAGGTTCTAGAAAGACCTGAAGCCGTAAAGAGGTTCTAGAAAGACCAGAAGCCTGAAAGAGGTTCTAGAAAGACCTAAAGCCAGGTTAGGTTCTAGAA
jcf7180000032674_21131	CTGGACCACCTTGGGACTCAACTTGTCCGCATACGTCCTGCGAAGCCAAAGAAAAGCGAGTTGACGTAATCTTTCGGGGTTTTTAACGAAGCTTTACGCT[T/A]TAATTTGTTTTTAAAAGGCCCTGAAGCTCTTACCCGGTGCTGATGGCCAGGATCTGCAGCAGGCGGGTGGAGGCCACCTTGAGGGCGGTGCTGAGCTGAG
jcf7180000033464_55389	ATATCCATCTATACGCAACATCTATTAGCCAGTGAGATTAAATACTTTAGAAAAGTGTTGCTTGGTATCGATTATTCACGGATGAACTGAGGTGGAAATG[C/T]GGCAAAAATGATTTCATTCTGATCTTTAATTCGACAAAACATTGAGTAAAAGCATCTCATGTCAGCATTTTGCAATACATACACACAAGGACACACATAT
jcf7180000033679_248374	ATGTGGCCAGGCTGACAGATGACATCCAGGCTGCATCTCTGTTCCTGGGTCATCCATTGCTCTTGGAGTGATAAGTGGCTCTGTCGTGGTGTGGCTCCTG[G/C]TGCTGCATACACAGTCAATCACAGTGCGGAGCCACAGTACAACAGCAGCCTGTGGAGGACAGCAGAGCAATAACTCGAGATGACATCTGTATCACCTGAA
jcf7180000033682_2207464	AAGACACTTAAAGAAACATTTAAGAACCTCCAGTCCTCACTGCGAGAAAAACAATAGTTGAAACACTGGTATAAACTAACAACTTCCCTTTTGAGGTACA[C/T]GAACAGCGAGATGATGAATATGATTAAATAAATGAAGACAGACCTTGAAGACTGGAGAAAGTGCTGTGTTTGGTTTAGAACCACAATTACTATTATAACC
jcf7180000033687_801937	GGAGAGAGCACATCACCCACCAAGTACAGCTCTACCGACAAAAGATTAAAGAAGATGAAGAAGGACGAGTAAAGATGGAGTTCAGGCTGTTGAAGATGCA[G/T]TTGGCAGAAAAGGGCGCTGCAGCAACAAAGAAAAGCATTAAGCCCCATAAACAAATGTCAGCCATGGAGGACATCTCTGTACAAATTGAGGAAGCAGTAG
jcf7180000033693_294152	CTGTATGTTGATTATGTTGCTATCCCTTCTGCAACCACCATTCATTCTCTCAATTTGTATTAAGCACTCACCGATGGGAAGATTTCTTAAAACAACATAA[C/A]ATCTAATTGTATGGAGAATAATTGAAACACAAAAGAGATGGGATATTATTGTGGGAAAAGATGGGTAGAGTGAGGGAATTAGAGAAGGGTGAACTAAAGG
jcf7180000033702_10109472	TCTAGGAATCCATCGGCAAGCAGTGTGACGACAGACCCACCTATGGGGGCAAAGAGGCGAAACAATGTGGATGACCAGTCATTCAAATATTTAGTTAAAT[G/T]ATGTTATATCATTATCTATTTGCAGGATGTTTACAATCTAAAGAATAAGGCTGGCAATGACTTCATTCTACTGACAAGCAGGATCAACAAAGCATTTTAT
jcf7180	ATATGTGCGTACTGTGCATGCGACAGACCAGAGGGGAATCACTGGTCCTTGATT

000033704_2530758	TTATGATACATTATTTTCTCTGAAAAGGGTTCGAGGCAGAAGAAAA[G/C]ACAGTGACACATTTTCCGTCCTATGGTATCAAGATTCTTTTCTTTTTACATTAGTTGTACAGTGAGTGTATCATTTCTGAGGGGGATATTAATAAAAACC
jcf7180000033715_694445	CAACAACTCCCCATTGGAACATCTCAATTACGTATTATTAAATATACGGAACAGCGATTTCTGAATTTTAACAACCCTCCACTTTGTTCTGTGGACTACA[C/T]GCTGAGCTTTTCAAAAATACTTACATAAGTCTCTAACCCAAGAGATGTTTTGCTGAGAGATGACCGAGGGTTGTGGGACTTAACTGTTTCATTCTGCAGA
jcf7180000033723_241992	CTGATATTGTTCTGGATTTGAATGTGTTACTGGACTTTTGACCCACAACAGTGTCTTTCAATCGCATGTGTATTGTAACAAATGTTGGGCATGCCACAAA[A/T]AATAAATATGATTTGATTTCCCAGCATCTATGTTGACAACAGGTGAAAATAAAACATGATATTGAACTGTATGTAACAAACACTAACTTGTTAGGATAAC
jcf7180000033725_342062	CGACAGCGTAGAGAGGAGAGTGTCAGTGGCAAAGTTAGGCAGCATGAGTGAGAAGGAGTCAGTTGAAGGGAGGGCTGACAGAACAGACGAGGCCAGAGAG[A/G]AGGGTGAGAGGGTGCAGATGTTGTGACGGACAGGTGCAGAGTCCGTTGTGGGAGGGTTGTCAGTTCCAAAGAGTGGGAGAGAGTAAGAGATGAAGAAGTA
jcf7180000033734_186988	TACAGTTGACAAGACCTTAGGATGGGCAGATGTGTAGCGGGGATGAAACCAGGAATTTGTTCTCTATTCAAATGTCGAATAGGGAACACTATCCAGGACA[G/C]CACAGACTTTGTGAAGCGCACTCCACCCCTCCACAAATCCCACAGCGCCCCTCTTGTTCCTCTCTCCGTCAGCTGCACCACCCACATTAATCAGCTGAGC
jcf7180000033738_441920	CGGTTAAAGATGTCGTCTGCGTACAGATTGTAAAGCCCTCTGAGGCAAATGTGTGATGTTGGGCTATACAAATACATTTAATTAAATTTAATTGAAAAAT[T/G]TAATATTAGGTTAATAAAGAGGAAAATGGTTATTGGAGGTTTACCTTTATTTAGCTACTGATTCACATATGTTTGCTATTGACTGATTTAAGTAGTTTAT
jcf7180000033745_294551	TTAATACGACTTGGATATTGACGTGAATGAAATTCACCAATCGGAAGAACCTAATAGTTCTCATCCACCGATGAAGACTGTGATATTAAGTGAACCGTGT[G/T]TTATTTTGTCATGCATTATACCTTTGAATAGCTTTATTGGTTCTAATTTGTATCCACAAAGAGAAGAGTTTTCCAAACGACCCCAACAATGCACTTGCTC
jcf7180000033750_128999	CGGATACTTTCATCTTATAATACAGTCTAGAATTAATTTGAATGAAGCATATTTCTAATCTCTCGCTTTGAGTTCATCGTCACATCTTTTAAAAGTGCAT[C/T]GTCCCCATTGCCCACAGTTGGGGGTCACCCACAGCTTACTGGGGGTCGCAAGCTTAACAGAGGACCTTTGACAAGCAACTCGTTCAATAAATACAATTTC
jcf7180000033774_309631	CTTCCACCGTGGCTCAGCGGGGCCACATATTGGAGGCCAGTAATCTCAGCCACGGATTTAGTGCCAGAGTCAAAAGGTCAACAACCTCTCCATCCACACA[G/C]AACTCTGACCATGTGATTGATAATCTAGACCGCAGCAGGAAACAACAAGTCCGGGCCATATTATTTTCTGTCCGCCAGTACCGACGGCAGCCACCGATTG
jcf7180000033792_292440	GTATTGTGGGGCGCCTGGACCCCAGAATCTGAAGGAGCTCTGGGACTTGTCGGACCAAGTCCAGCGCACTCGGTACAGACCGATCCAACCGTAATATCCG[T/C]CTAATTGTGCAGCGACCGCTTTGTTTTCTTCATCATTCTCGATCATAGCCAAGTCGGTGTGGTGTTCTCGGCAGTAGTCGCGGGCATCGGACCACTTTTT

jcf7180000033794_278028	AAAGACTGAGAGCAGGGGAAACGGCACCTCTGAAGCTCACTAATTAAAAGGTTCTACATGACATGCAGGATTCAGAGCATTAACAGCAGCATACAACACA[G/T]TCCTACGGTGGCTAAATTACTGAGCTGACAGCATTGAGTTACATTATGATGCGTTCAAGCTCTATTCAGTGAAAATGGGTATTGTGCGAAGGTAAATCGT
jcf7180000033806_453803	GCTCGTTGATTAACCGAGCTGATTACACGTTAATTAGCTATTTAATAATAAACAAAGCGTGCTCGGGATATTCACGCATGTAGTAACCAACAGGAAAACA[A/C]CACATTTTGGGCGCACGCCGTCAGCGCAGAGCAGCTGCGGGTTCGCCGAGTTTCATTCAACGCGTTAGCTCGGAGTTAGCTTCCTGTAGCTAACGCTAGC
jcf7180000033844_1094302	ATTCCTGCTGGTGATCAGACTGGATGCACGCTTTACTGAAGAGGAGAAGAAAGCTGTCAAGTGGTTTGAAGACAACTTCGGTGAGGAATTCTCCAATTAC[T/A]CTCTGGTGCTCTTCACTCGGGGCGACGCGCTCAAGGGGAAGTCCATTTGGGAGTATTTAGGCGAAGATCCTGATCTCAAGGAGTTAATCAGTGTCTATAA
jcf7180000034005_239257	ACGATGTGAGACGGACTTCACGGAGTCGTTGGAGAGCCGACGGTGTTTGTAGGAAGCTGCCAATTCCATGGACTTCAGATCAGCTTAACAGACATCCTAC[T/G]GCAGCCTCCGAGGGGAACAACAGTCGGCTGTGATGAGCTGACACTTCCAGTCCCTCAGTGGAAGGCCAAAGAAAACACGGCAGAATGGGAGAGGAGGTGA
jcf7180000034094_76521	TCTGAGTCTTCTGTACTTGTGCTAATATTCCACCAGCACAGATACACATTTAATAGTTACTACCACAATAGAGTAAAACACGTTTTATAATTGCTATCAC[G/A]ATAAAGTAAAACACAAACTGAACACATGCTTTCAGTCGCTGTGATTTCCTCCCGTTGCCACAAAGTTCCATCATCTTGCTTCAATGTGATAATGTGTCAT
jcf7180000034097_39292	AACACCGATACGCTGTTCACTAAACCTATTTCCTCAAAAAGGATTGTTTTATTTCCTCGCTGGAGACCCGAGCGGTGTTGTGAGTAAACAGGAAGTAGAG[T/C]GAAACAATAAAAGCCCCAGAAGGAGGTTCATTAGAGCATTTAGTTGTGGTGTTTACAGTGAAAACAAGCCGTGACGTTCAGGCTCTGAGGCGCAGAGGGA
jcf7180000034306_22398	TAAGACACTGGTCTGGACACATAAAGGGACATCTGAGAGACAAAGACCAGCACATTCTTTAAATACCTGCTAGTGAGTCATAGTATTCATCCTTGTTGTT[T/A]GTGCATTGTATGTATTTTGGATTAGTACTTCCTGTTTGACGTCTATCTTGTGAATCTTTAGAAATATTTATCTTGTGAAATAATGATCAGACTTTATATT
jcf7180000034419_227988	ATGATTTATTTGGCCTTCCCTTCTCTCTGCGGTGCTTCTGCATGGAGACTCTCCCACTGCGACTGTTTAATTAGTCACCAATGTCAGTGTTTCAGTGCCG[T/A]GCCGCTCGACTCCCTCGGTGACAAACTGTTGCCAAAGAAACACATCCGCTCAGATCCTCTGATGCCGAGGGGCGGCTTGGAGAGTTCAGGGGGTGATTGA
jcf7180000034432_3442488	AGTTTAAAAGTATAGTATAGCAAAGTTAACTGTCTCAGTCTTCATGTTCACCTATAGTTACTGGATGATCCAGATAGATTGATTTATACTAGTCTAATTT[C/T]GATTCGATTATTTTGACAATGTAGAGAACTGTTAATGTGTCCACATTAACAATATATAAAACACATCACTTTATGTGGAAAATGTAGCCATTTGTTAGTC
jcf7180000034438_10	CGAACCCTCCGACCCTGTTGATCCCTCATGCGAGCCTTCGGCCTGAACCAGCTGCTTAGAAACTTGGAGCACTACGACAGATCACTGCAACACAAGATCA[A/G]CTCTTTCCTTTACAGAACATCAATATGCAAGAGAATACTTCATATTCATGTTTCATATTGTTA

92426	ACACCTGATTTCAGTACAATTCTGATTGCAATGTGCA
jcf7180000034444_81871	TAACGACAACATCAAATACAGACATTCTGTCACAATATCAGATTCCACCTTCACTAGCAATATTATCATAATACCTACAGAAGAAAAATAATAGTGCTAT[A/C]TGTCAAAATTCATCTTTATTTTTGGTTAGTGTGCGTGATTACACTAAAAATGTATCGAGGTGGAGAGGGAGTCTGTTACCATGACCCTCAAGTCCTCTTT
jcf7180000034465_87845	TTCTAATGCCATTTAATCTCTGACTGTTATTTTCCAACTTCCTGTCTCTCAGTTACAGCGTTGAAAGGGCCTTCAGTCAAGCCTTTTGTCATATTCTACT[T/G]AATTATAATCTTCGGAGCAGCTCCCAGCCTCTAAATCTCATAATGGTGCATAGCTCAGCCTGTAGAGAGTCCTTTCCAAGATTGTTGCGATTATGCTTTC
jcf7180000034467_857969	ACTGCATGATCTCATACCTGAGGACAGAGGATTTTAAGTTTGAAGAGACATAATTAGGGCATTTTCACTTTCCAATTATATAATCATAATTTCTGCAATG[T/C]TCCAACTAGTTGATGAGTACTGACCAGTAGTCTGCATGGGAGAGTTTGAGTCTGGGCTGGGCAAGGGTGATCTGTCTCTCCCTAATGACAAAGGTCAACA
jcf7180000034473_234958	ACAACTCTGATCTGGATTTGTTTTATTGCTGGTTCTGGTCTCTTCAGGGGAGGGTTTGAACACGATGAGACGATAAAGCAATGCCAGCTTCACCCTCCGT[C/T]TTTCACTCAAAATGTTGTCGTGCCCCTTTTGTGTTGCTTCCCTGAAATGCGTATCTCCTCCCCTCAGTGTTAAAACGGTGGGCCTGTTCCCACGCCGGTG
jcf7180000034477_132223	CGGACTGGGCACCCGGCGTTTTAGAGGTAAACGTCGGCGAAACCGCCCGTCGGTGACTGGAGACCACGTTCTCTAGGTAACACAGGGTCAGACCGAGGGG[T/C]GTGGCTAAAATCAGGGCCAAGACCAGGGACTGGCTCGCCGTCAGCATCACCGCCAACAGGAAGACGACTAAGAGGCAGGGCAGGAACAGAGGGGAGGGAT
jcf7180000034503_66077	AGAATTTCCTCTCCGGGGTTAAATAAAGTCTATAGAGTCAGTGAGTAAGTGTCCACCTTGTCACTGTGCTTTATATAAAAGTAATTTATGTTGCCGTTAG[T/C]GCCGACAGGTGCGTGGGAGAGACGAGTTCAAGACGATGGTGTAAATGCCATCAAATATTGATCGAAGCAGAATATTCCATTGGATAAAGACAAAGTTGTG
jcf7180000034509_145901	TGATGGTGCCGTCGATGTCAGAGATTATGACCTCGTCGTCCCAGTTCCACAGGTAGACGGTGCCCTCGCAGCGGCACGTCCCCTGGTACTGAGTGGTGAT[A/G]CTGAAGGTGACGTCATTCGGCCCCTCCCTCAGCTTCAGGCTGGCCTGCAAGAGACAGAAGACGAGTGAGTTTAAGTGAGTGCCTCTTATTAAACAGCGAT
jcf7180000034510_1446864	CTCCTGCGCGCTATTCTCGCGGAGCATGTCGAGGGTCCGGACCGCTCGACGCACGGGAGTCATCAGCGTGCAGTCATCCAGCTGACTCCCAACACTTTGG[A/G]CATACCTTCCTGAGAAAGAAGACAAAGAAGACCGGAGCGTTGTTCAACACCACCTCGACCGAGTGACTGCTCCCGTGCGGGTCGGATCACATCCGTCGCT
jcf7180000034525_228049	TGTCATGTAGTGTTAAAGTGCCTTGAGTGGTCGGAAGACTATAAAAGCGCTATACAAGTAGGTATATTTATCATTTACCATAGCATAACATCATTAGTGA[T/A]TATCTGTCGGGGTGTCCACACCATATCTCATACACACGACGTACAGGTACACGTGTTGCTGTTGTCAGATCTCTCTTCAGTATTGAGCTGCAGTGCATTGT
jcf7180	CTGTATGTGAGTTTATTGACTTTGTCTTCACCACTGCAAGTGATGCACTTTAAAGC

000034529_825249	ACCTGTTTGTTGTTGATTAGTACTATCACCACAACTACCACAAG[C/T]ACGTCTCCATCTTTTTGTTTCAGGCCCTACATTTACGCCATGCACTCTGAACAACCAGACACATACGGACACAAAGTGGGGCCCATGAGCACAGAAGTGA