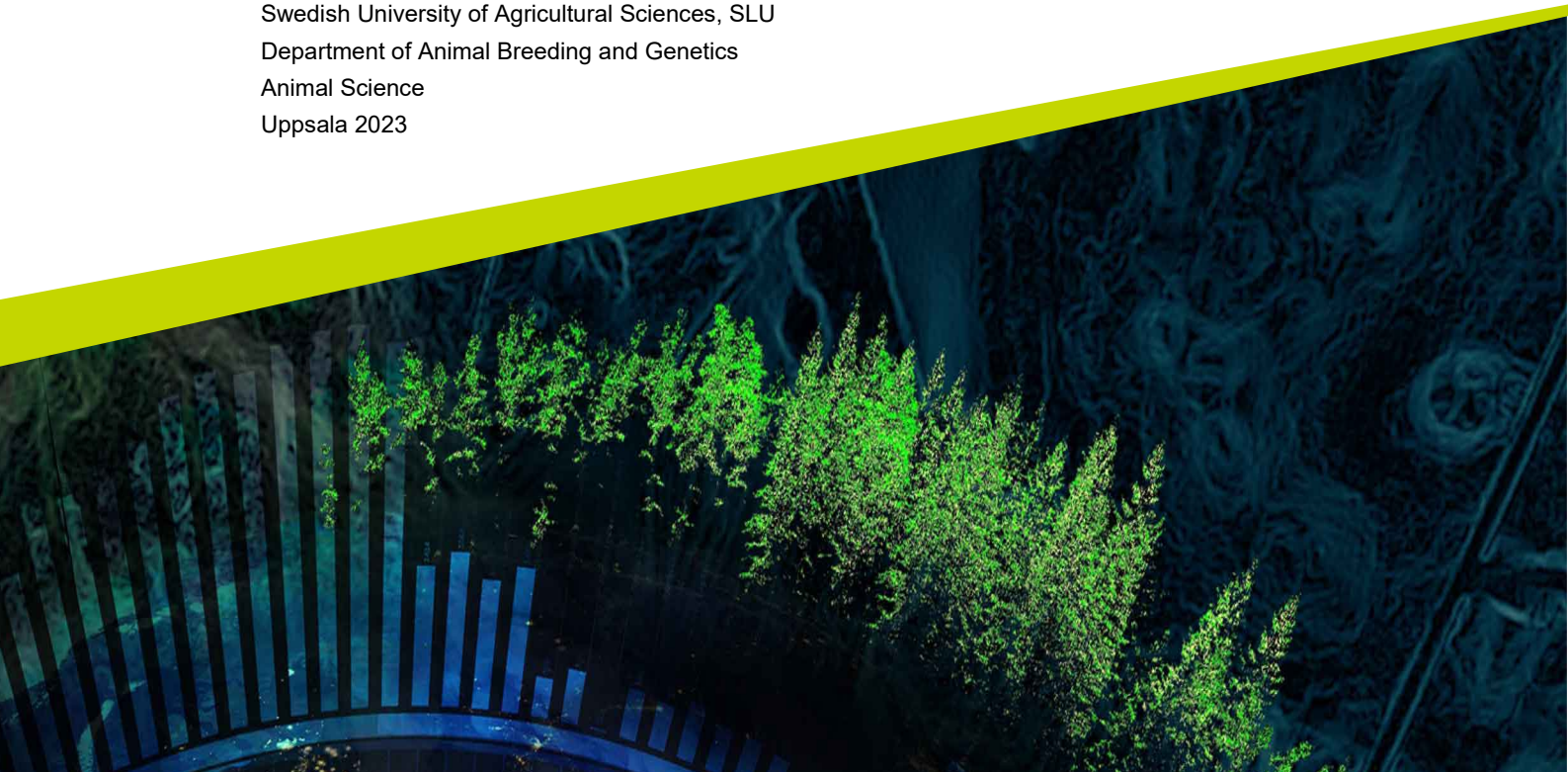
Parentage Assignment using genetic markers in the common carp (Cyprinus carpio)

Sven Kristian Feddersen

Independent project in Animal Science • 30 credits
Swedish University of Agricultural Sciences, SLU
Department of Animal Breeding and Genetics
Animal Science
Uppsala 2023



Parentage Assignment using genetic markers in the common carp (Cyprinus carpio)

Sven Kristian Feddersen

Supervisor:	**Christos Palaiokostas, Swedish University of Agricultural Sciences, Department of Animal Breeding and Genetics, Applied Genetics**
Examiner:	Susanne Eriksson, Swedish University of Agricultural Sciences, Department of Animal Breeding and Genetics, Applied Genetics

Credits:	30 credits
Level:	Master's level, A2E
Course title:	Independent project in Animal Science
Course code:	EX0870
Programme/education:	Animal Science
Course coordinating dept:	Department of Animal Breeding and Genetics
Place of publication:	Uppsala
Year of publication:	2023

| **Keywords:** | single nucleotide polymorphism, Parentage Assignment, aquaculture, genetic markers |

Swedish University of Agricultural Sciences
Department of Animal Breeding and Genetics

Abstract

This study investigated the effect of varying single nucleotide polymorphisms (SNPs) marker density, error threshold selection, and different SNP selection strategies on parentage assignment accuracy in carp populations. Using a marker set of 15,615 SNPs, we found a positive correlation between the quantity of SNP markers and the accuracy of parentage assignments, consistent with existing literature. We discovered that error threshold selection significantly influenced assignment accuracy and recommended its careful consideration based on population genetic characteristics. Furthermore, our interval-based SNP selection analysis showed that SNP density crucially impacts parentage assignment accuracy. Notably, random SNP selection across the genome yielded more accurate assignments than chromosome-specific selection, emphasizing the importance of diverse and representative SNP sets. These findings provide valuable guidance for future genetic research in parentage assignment.

Keywords: single nucleotide polymorphism, Parentage Assignment, aquaculture, genetic markers

Table of contents

List of figures

Abbreviations

AKVAFORSK	Norwegian Institute of Food, Fisheries and Aquaculture Research
APIS	Auto-Adaptive Parentage inference Software
CGIAR	Consultative Group on International Agricultural Research
FAO	Food and Agriculture Organization
GIFT	Genetic Improvement of Farmed Tilapia
ID	Identifier
MAS	Marker Assisted Selection
PCR	Polymerase Chain Reaction
RAS	Recirculatory Aquaculture System
SDGs	Sustainable Development Goals
SLU	Swedish University of Agricultural Sciences
SNPs	Single Nucleotide Polymorphisms
SSRs	Simple Sequence Repeats
UN	United Nations
USD	United States Dollar

1. Literature Review

1.1 Global population growth and hunger

According to predictions, by 2050, the human population is expected to reach 9.7 billion. This higher demand calls for a higher supply of food. In light of the populational increase and the stagnation of fish production by traditional fishing industries, aquaculture will play a significant role in bringing food and economic growth in a sustainable manner, across the globe. The decline in fish stocks as well as overfishing highlights the importance of an alternative and more sustainable food production system. Compared to livestock, aquaculture can produce a higher yield of fish per unit area, which means, a more significant potential to produce more food in less space, a crucial aspect for meeting the world's growing demand for food (FAO, 2022). This rise in demand for food comes with challenges such as competition for resources and overfishing, and the need to reduce the food system's environmental impact, along with the effects of climate change. To address these issues and ensure sustainable food production and food security, a comprehensive global strategy is necessary. Selective breeding in aquaculture with a special focus on fish farming is in its early stages and therefore offers an opportunity to develop this production system in the most sustainable way possible (Godfray et al.,2010).

1.2 Aquaculture Production

Aquaculture is the quickest-growing segment in the global food production sector with an annual growth rate reaching an average of 5.8% from 2000–2016 (FAO, 2018). In 2020 alone, production reached a peak of approximately 97 million tons, out of which more than half involved fish consumed by humans. This resulted in a total worth of USD 259 billion (FAO, 2022).

On the other hand, capture fisheries production has remained stable since the late 1980s just over 90 million tons. Of this total, 87% (79.3 million tons) came from marine fisheries and 13% (11.6 million tons) came from inland fisheries (FAO, 2018). To address this increased demand for food, global aquaculture

production is projected to grow 37% over 2016 and reach 109 million tons by 2030 (FAO, 2018)

The growth of aquaculture also comes with challenges that will need to be addressed in order to minimize environmental degradation, disease outbreaks, and competition for resources with humans. According to the FAO (2022), social challenges, such as the need for responsible management, fair access to resources, and good working conditions for workers should also be included in the challenges for the sector's development.

1.3 Aquaculture and UN's Sustainable Developing Goals

Aquaculture can play an important role in achieving the United Nations' Sustainable Development Goals (SDGs), especially in developing countries with important participation in the aquaculture sector (Nasr-Allah et al. 2021; Sampantamit et al., 2020).

Firstly, aquaculture can contribute to the reduction of poverty by giving small farmers options for income, generating jobs, and increasing food security (SDG 1) (Sampantamit et al., 2020). In addition, in terms of combating hunger (SDG 2), aquaculture is important to increase food security and nutrition, by providing high-quality protein sources, essential fatty-acids (Troell et al., 2023), minerals and vitamins, including in poorer communities (Sampantamit et al., 2020).

The availability of a variety of nutrient-rich fish and seafood items from aquaculture can also help to promote healthy diets and lower the risk of diseases like heart disease (SDG 3). Aquaculture can also empower more women by expanding their access to resources, giving them additional earning options, improving their decision-making power and promoting equal wages (SDG 5)(Cavalli et al., 2021). The adoption of Recirculating Aquaculture Systems (RAS) minimizes water use and can have positive effects on water quality (SDG 6). By treating and recycling water in the production system, RAS lessens the demand for water from natural bodies, as well as providing clean water for the farmed species. This approach positively affects the health and growth of aquaculture species and contributes to the sustainable use of water (Cavalli et al., 2021).

Moreover, aquaculture can boost the economy and produce jobs, especially in rural areas (SDG 8) (Cavalli et al., 2021). By reducing waste and encouraging resource management, it can encourage sustainable production and consumption behaviors (SDG 12). Moreover, by relieving the pressure on wild fish populations and restoring damaged aquatic ecosystems, aquaculture can help in the preservation of marine biodiversity (SDG 14) (Cavalli et al., 2021). Lastly, aquaculture supports sustainable land-use practices by the adoption of responsible production systems aimed to mitigate impacts on the environment and also by reducing the demand for land-based agriculture since it has the potential to replace to a certain extent less sustainable land-based agricultural systems, therefore contributing to the preservation of wildlife and natural habitats (SDG 15) (Troell et al., 2023).

1.4 Environmental Impact

The survival of the organisms raised in aquaculture is largely dependent on the quality of the water. Water quality is impacted by variables such as dissolved oxygen, pH, and nitrogen levels. Aquaculture practices like feed consumption and the release of waste by cultured organisms also have an impact. Various cultural systems produce varying amounts of trash, which has its own effects on the environment. A type of organic pollution called eutrophication can happen when substances like nutrients, leftover feed, faeces, and dead fish are dumped into bodies of water. Poor water quality can cause stress and disease in cultured species in intensive culture systems with high stocking numbers. There is a significant reliance on chemicals and medication in the form of antibiotics to address these issues (Martinez-Porchas and Martinez-Cordova, 2012).

The expectations for growth, raises concern about aquaculture's impact on the environment (Waite et al., 2014; Naylor et al., 2000), especially the impact of the inputs and resources, such as water and land usage, feed, and energy (Waite et al., 2014), given its dependent relationship with the environment. In light of this, in order to develop a sustainable, long-term production system, it is crucial that good practices are adopted from early stages. Aquaculture has the potential to become the primary food production sector, not only in economic terms, but also from an environmental and social perspective (Naylor et al., 2000).

Because aquaculture depends on the environment, there is an interaction between both in which aquaculture influences the environment while the environment also influences aquaculture. There have been numerous attempts to identify the various environmental risks associated with aquaculture (Martinez-Porchas and Martinez-Cordova, 2012). The effluents from aquaculture operations flow back into the very water bodies they depend on, leading to changes in the microbial communities and introducing potentially toxic chemicals. These effects are especially noticeable in lakes and slow-moving canals that provide water for these farms (Ozbay et al., 2014). When assessing water quality, lime is often used to control pH levels, to promote a more adequate environment for the animals. On the other hand, the use of fertilizers, in the form of both inorganic (nitrogen and phosphorus) and organic compounds (animal manure) plays a significant role in increasing productivity by promoting the growth of phytoplankton, a base component in the aquatic food chain. However, overfertilization increases the levels of nutrient, leading to eutrophication of the water, which affects oxygen levels and harms aquatic ecosystems (Martinez-Porchas and Martinez-Cordova, 2012).

1.4.1 Resource Use/Inputs

Small fish are being collected and turned into fishmeal and fish oil to make aquaculture feeds despite the fact that the natural population is decreasing. The environment suffers when wild fish are used excessively to provide aquaculture feed because sustainability is impossible. Around 15% of the 891 million tonnes of fish produced between 2007 and 2012 were used for feed and other non-food purposes, according to FAO (2014), however, this percentage was said to be declining. This drop can be related to the increase in plant-based components as well as other animal byproducts used in aquaculture feed formulations.

The largest consumer of water is the agricultural industry. Apart from the expanding demands for domestic and industrial usage, one of the main factors contributing to the rise in global water demand is the share used for agriculture, including crops, cattle, and aquaculture. The majority of this freshwater requirement is used for crop irrigation. Agriculture is responsible for over 70% of all freshwater usage globally, but in most developing nations, this percentage may reach 90%. (FAO, 2011). Freshwater use for agriculture in developed nations is relatively modest (up to 5% of total consumption), while more water is tapped for industry and energy generation (15%).

Since water is the most essential component in aquaculture, water extraction is unavoidable. Less than 1% of the world's renewable and accessible freshwater is used for aquaculture (Boyd and McNevin, 2014).

1.4.2 Water Use

Both organic and inorganic components included in aquaculture wastewater have the tendency to add to the load in the environment where the effluent is released (Andreotti et al., 2017). Due to eutrophication, both large-scale aquaculture and dense small-scale family farms have a tendency to limit the growth of aquaculture in the region where they are located as well as, indirectly, globally (Andreotti et al., 2017). The economics of shrimp cultivation in Thailand's Krung Krabaen Bay and Welu wetlands are unsustainable as a result of nutrient enrichment. In the heavily populated state of Kerala in India, eutrophication in the receiving rivers was caused by the direct effluent discharge from Macrobrachium farms. There are species that can be used for aquaculture that are fed and those that are not, with the latter having the aim of reducing resource usage in the form of feed while simultaneously ensuring environmental integrity (Li et al., 2020). Nevertheless, feeding practices frequently result in significant nutrient loading in receiving water bodies due to uneaten feed, feces, and other biological waste generated during metabolism (Han et al., 2020). Poor feeding practices and the use of low-quality feed can cause nutrients to be discharged into receiving waters, where they may either be decomposed or build up and cause pollution, depending

on the conditions in the receiving waters and the concentration of the nutrients released (Romero-Soto et al., 2018).

1.4.3 Genetics

Beyond extraction, humans are also attempting to restore fish populations that have been depleted through techniques including stock improvement, fish introductions, and marine ranching. Stocking has been used in inland and coastal fisheries to increase fish populations, but the results have been mixed economically. While some methods have been effective, others have not. There have also been reports of biological effects, such as high rates of wild salmon death as a result of cannibalism by stocked fish (Naylor et al., 2005).

The chance of establishing exotic fish for aquaculture as well as the potential for any unfavourable impacts after establishing them determine the dangers of doing so. Only a handful of the several introduced species in Hawaii managed to establish themselves, and important predictors included the quantity, maturity period, and water depth (McKindsey et al., 2007).

Genetic modification and escapes of farmed species can result in genetic diversity loss or deterioration, extinction, and hybridization. The use of risk assessment is advised to determine the consequences of genetically modified organisms because they may provide unidentified dangers to natural populations (Naylor et al., 2005). Fish escapes from aquaculture facilities can result in fitness problems, disparities between wild and farmed populations' partner preferences, and reproductive failure. Male chinook salmon raised in captivity had better sperm fitness than their wild counterparts (Lehnert, Heath and Pitcher, 2012). The relationship between the effects of escape, the population of escapees, and their capacity for reproduction leads to believe that the migration of populations with lower levels of adaptation can promote the spread of maladaptation in wild populations. Inbred lines of prawn can develop in aquaculture as a result of ineffective broodstock management and breeding (Singh et al., 2010).

1.5 Domestication and Selective Breeding

Domestication has been defined by Price (1984) as the process in which animals become adapted to man and the captive environment that they find themselves in, by changes in their genetic material over generations and developmental processes caused by the environment across each generation. The same evolutionary processes that allow free-living populations to adjust to changes in their environment will be used to produce phenotypic adaptations to the captive environment. The main distinction is that man can use artificial selection to speed

up phenotypic changes in captivity that would otherwise not occur or endure in nature.

Selection is the primary mechanism of domestication. Animals that are most adapted to a specific habitat will, through the process of natural selection, generate more offspring that survive than those that are less adapted. Artificial selection will also occur in a farming environment since the farmer would often choose animals with the best behavior and the fastest growth, or whatever the trait of interest might be. The animals' productivity will rise as they live longer, are less anxious, and have more disease resistance. In simple terms, the fish that have been domesticated are more suited to farming circumstances (Vandeputte and Prunet 2002).

Selective breeding can produce even more significant improvements and efficiency than the process of domestication alone due to its ability to adapt to the farming environment. The majority of aquaculture species have first been targeted for growth rate improvement. As the growth rate increases, production time and maintenance needs will also automatically decrease. This indicates that a higher production can be obtained while still using the technological resources available to each farm (Gjedrem, Robinson and Rye, 2012).

1.6 Selective breeding in aquaculture

1.6.1 Salmonids

The first large-scale family-based breeding programs for salmonids were developed in Norway, during the 1970s and are now recognized as the industry standard for the genetic improvement of aquaculture species (Gjedrem, 1985). The development of effective and sustainable salmon farming in Norway has relied heavily on the introduction of selective breeding programs on Atlantic salmon. Beginning in the 1970s, AKVAFORSK began a national selective breeding program by collecting fertilized eggs from more than 40 populations of Norwegian river fish. Throughout the 1970s and 1980s, a number of private selective breeding operations were also started. Using a combined family and within-family selection technique, the Norwegian national breeding program was created to gradually integrate all economically significant traits in the breeding target (such as growth, age at sexual maturation, disease resistance, and quality traits). As a result, the Norwegian salmon industry has more than saved more than US$ 230 million annually due to the higher feed efficiency of the selectively bred salmon (Thodesen, et al., 2006; Janssen, 2019).

1.6.2 Nile Tilapia

The Genetic Improvement of Farmed Tilapia (GIFT) program was launched in the Philippines in 1988. The GIFT project, initiated in 1988 and still ongoing, serves as a cornerstone for global and Asian initiatives to enhance tropical fish stocks. The GIFT project has been successful in creating tilapia with higher growth rates and survival rates with a shorter harvest period, thus drastically changing tilapia farming worldwide (CGIAR). As tilapia is a fairly cheap fish, its introduction and distribution have increased rural incomes, created jobs, and improved human nutrition, particularly for the poor. Tilapia farming offers an attractive source of income for hatchery owners and fish farmers, and GIFT has significantly contributed to the creation of jobs, including for small farmers. Since many of the new fish farmers are women, this increases the local availability of high-quality, reasonably priced protein, generates money, and improves household nutrition while also empowering local women. Furthermore, the GIFT technology has potential environmental and genetic risks, and improvements in management practices and infrastructure could increase the yield and profitability of the local strains even if genetically improved strains are not introduced. These improvements also will ensure the realization of the full potential of introduced strains (Ansah, Frimpong, & Hallerman, 2014; CGIAR).

1.6.3 Common carp

Originating from central Eurasia, the common carp, with a cultural history of about 4000 years in China, is the most domesticated fish worldwide (Hulata, 1995). It is divided into subspecies such as the European-Transcaucasian, the Amur-Chinese, and the South-East Asian carp, each reflecting differently.

Genetic improvement focuses mainly on traits like growth rate and disease resistance (Hulata, 1995). A five-generation mass selection program on the European common carp for enhanced growth rate yielded little success, attributed to decreased additive genetic variation from prior domestication and uncontrolled non-genetic variation. Still, family selection showed promise, particularly in low heritability situations (Hulata, 1995).

Efforts to decrease the intermuscular bone count have not been successful, whereas breeding for disease resistance, specifically against infectious dropsy, saw significant progress (Hulata, 1995). The dropsy-resistant Krasnodar common carp now thrives in commercial fish farms. The ongoing project aims to improve dropsy resistance further, identify major genes for resistance, and understand the mechanisms of susceptibility to the disease (Hulata, 1995).

1.7 Biotechnology

The use of genetic markers has become increasingly important in fish production, as they can help assign parents, improve breeding programs, and increase the efficiency of aquaculture (Nguyen et al., 2022). In particular, single nucleotide polymorphisms (SNPs), microsatellites, and marker-assisted selection (MAS) have been widely used to achieve these goals.

Microsatellites, also known as simple sequence repeats (SSRs), are short sequences of 2-6 base pairs that are highly polymorphic and widely distributed across the genome. These markers have been extensively used in fish production for parentage assignment, as they can provide a high level of resolution due to their high variability (Radanović et al., 2022). Depending on their variability in the population, a high level of assignment can be achieved using fewer than 10 to in some cases 50 microsatellite markers (Glaubitz, Rhodes, & Dewoody, 2003). Because of their low variability, approximately six times more SNPs than microsatellites are required to ensure the efficiency of a particular parentage assignment (Griot et al., 2020). Microsatellite markers can be used to distinguish between individual fish, making them valuable tools for assigning parents and establishing pedigrees in breeding programs.

SNPs are single base-pair differences in the DNA sequence that occur between individuals. They are abundant throughout the genome and can be used as markers to identify and trace specific alleles in a population (Zhang et al., 2022). SNPs have become the marker of choice in many fish breeding programs due to their high abundance, stability, and ease of genotyping. High-throughput SNP genotyping platforms have allowed for cost-effective and efficient identification of a large number of markers, which can be used to assign parentage and track genetic diversity (Garcia et al., 2023).

The use of genetic markers, specifically microsatellites and more recently SNP's has become an essential tool in modern aquaculture breeding. These markers have allowed for improved breeding programs, better management of genetic resources, and increased efficiency in aquaculture, ultimately contributing to a more sustainable and productive industry (Nguyen et al., 2022).

1.8 MAS and Genomic Prediction

Marker-assisted selection (MAS) is a method that utilizes genomic markers, such as SNPs or microsatellites, to select individuals with desired traits for breeding purposes. It is crucial to note that MAS works most effectively when the traits of interest are controlled by genes with large effects (Vallejo et al., 2017). However, many traits, in the likes of growth rate, disease resistance, and stress tolerance, which are important for aquaculture success, appear to be polygenic, involving the

contribution of a large number of genes each having a small effect (Palaiokostas et al., 2016). Despite this, MAS can still significantly improve the efficiency and accuracy of breeding programs by reducing the time and cost associated with traditional selection. In fish production, MAS has been used, but the polygenic nature of many traits presents challenges that need to be addressed for its optimal application (Vallejo et al., 2017).

In fish production, marker-assisted selection, allows researchers to pinpoint particular genetic markers connected to desirable traits and select individuals carrying these markers for the next breeding. The use of MAS in fish breeding programmes speeds up genetic advancement as well as it offers solutions for problems with phenotypic selection, such as the environmental influence on traits. Breeders can directly choose breeding candidates based on their genetic value and avoid these environmental influences by utilizing MAS, allowing for a more accurate and efficient selection process (Palaiokostas et al., 2016). Genomic prediction is an enhanced methodology that uses genome-wide markers to estimate an individual's genetic merit, improving the use of MAS even further (Meuwissen, Hayes and Goddard, 2001). The genomic prediction uses all available markers across the genome, capturing the effects of all genes, including those with tiny impacts, which enhances prediction accuracy in contrast to MAS, which concentrates on a small number of markers linked to the traits of interest. Genomic prediction enables early parent selection by making possible the calculation of the breeding values of individuals without phenotypic records. As a result, the marriage of MAS with genomic prediction offers a promising path forward for aquaculture breeding programs that are effective and long-term (Palaiokostas et al., 2018).

1.9 Parentage Assignment

The assignment of parentage and, consequently, pedigree information are crucial components of animal breeding, particularly aquaculture, a field with particular difficulties (Gjedrem, Robinson, & Rye, 2012). Conducting efficient breeding operations, assessing genetic parameters, and promoting the genetic development of a species over generations all depend on having accurate pedigree data (Vandeputte & Haffray, 2014). The fundamentals of parentage determination are founded on the idea of Mendelian inheritance, in which the progeny receives one of two alleles from each parent at each locus (Grashei, Ødegård, & Meuwissen, 2018). Most modern breeding programs are built on these principles, and their success depends on their capacity to precisely track pedigree data (Lillehammer, Meuwissen, & Sonesson, 2013). Keeping track of pedigree is often easier when breeding terrestrial animals. However, because of their aquatic habitat and way of reproduction, aquaculture species, especially those classified as mass spawning species, present special challenges (Gjedrem et al., 2012). Since many of these

species lack practical ways to monitor their pedigree, genetic markers must be used to collect and maintain correct pedigree data (Vandeputte & Haffray, 2014). Although effective, this strategy adds additional complications and factors to the breeding process. The need for precise sample processing, data analysis, and genotyping error management are just a few of the challenges that need to be addressed (Grashei et al., 2018). In practice, parentage assignment is anchored on the identification of genetic markers like microsatellites or single nucleotide polymorphisms (SNPs) (Houston et al., 2014). Once these markers are identified, statistical methodologies are used to identify the most probable parent-offspring relationships. This process is a vital component of breeding programs in aquaculture. Some parts of the process, such as taking tissue samples and extracting DNA, can be time-consuming and expensive. Using PCR and electrophoresis to work with DNA markers adds to the cost. However, thanks to newer, more affordable methods, it is becoming easier to determine parentage in fish farming (Houston et al., 2014). In addition, with the latest technology, the use of SNP panels has become more available and even analyzing the entirety of the genome, providing significantly more information to work with (Houston et al., 2014).

Parentage assignment primarily relies on two computational methodologies: exclusion-based and likelihood-based methods (Grashei et al., 2018). The exclusion method operates on Mendelian segregation of alleles. It is direct and effective but can be sensitive to genotyping errors. This approach rules out potential parents if an offspring's alleles at a specific locus don't match the candidate parents. When error rates are moderate, a minor number of allele mismatches between offspring and parents are permissible (Grashei et al., 2018).

On the other hand, likelihood-based methods deal with parentage assignment from a probability point of view, attributing the parentage to the couple that seems most probable (Grashei et al., 2018). Unlike exclusion methods, which primarily rely on direct allele matches, likelihood methods estimate allele frequencies and take genotyping error rates into account. They often identify more parent-offspring matches, especially when the marker sets are less powerful. However, these methods can yield inconsistent results. Information about sibling relationships becomes invaluable here, enhancing the efficacy of these methods. These methods are typically more computationally intensive and may require the use of specialized software tools to carry out the necessary calculations (Grashei et al., 2018). The likelihood method was created to handle cases where the exclusion method does not work (Chakraborty, Meagher, & Smouse, 1988). This method calculates the chance that a set of parents are the true parents of an offspring by looking at all the available genetic markers (Marshall, Slate, Kruuk, & Pemberton, 1998). Offspring are simulated based on the genotypes of the parents, according to Mendelian inheritance rules. The difference in likelihood between the most likely and second most likely parent pair is calculated, and the difference between the

simulated individuals assigned to their true parents and simulated individuals assigned to an incorrect parent pair is then used to define a threshold value, usually at the 95% confidence interval. This threshold helps ensure accurate assignments while maintaining a good assignment rate." (Griot et al., 2020). The development of parentage assignment methods has impacted the estimation of genetic parameters in aquaculture. It allows the estimation of heritability and genetic correlations, which are vital in evaluating expected genetic gains and designing breeding programs (Vandeputte & Haffray, 2014). The ability to use accurate pedigree information through genotyping has arguably been the most significant contribution to aquaculture genetics in recent years (Gjedrem et al., 2012). Mixed family rearing designs for efficiently estimating genetic parameters have enabled the study of heritability for a variety of traits in several fish species (Tsai et al., 2016). Moreover, these developments have also facilitated the use of more advanced selection strategies such as genomic selection, which relies on marker information across the entire genome to make selection decisions (Lillehammer et al., 2013).

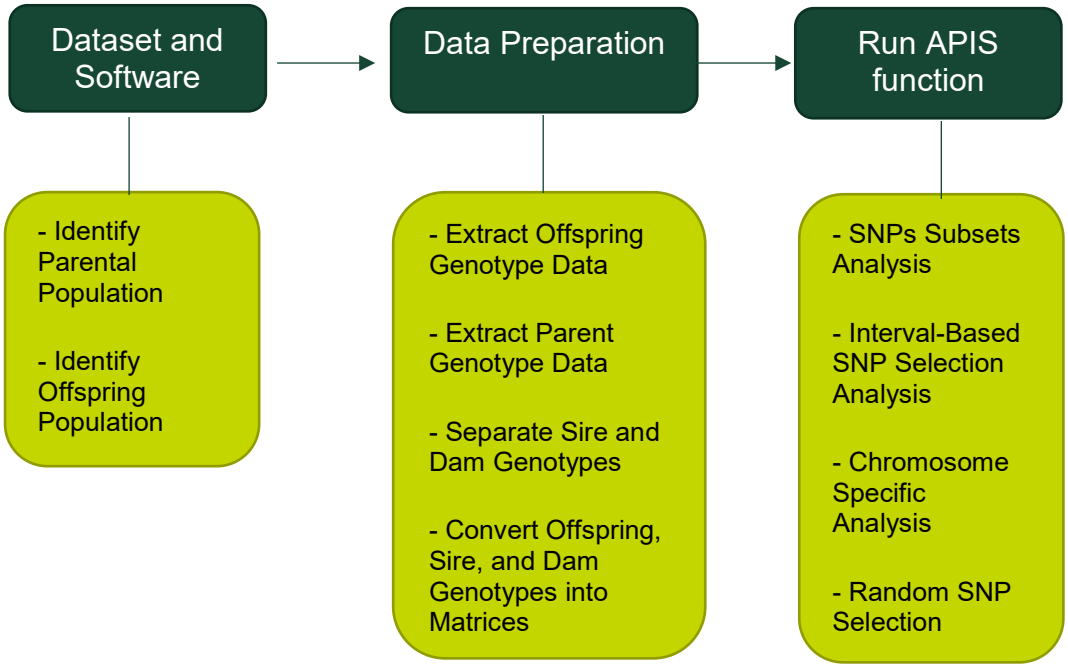
In the aquaculture setting, especially for early-stage organisms, the tagging of individual fish to monitor their performances can be impractical or impossible, thus restricting the depth of genetic studies that can be carried out at these stages (Palaiokostas et al., 2016). While technologies are advancing, the nature of early-stage organisms may continue to impose inherent limitations on individual data collection. Therefore, alternative strategies or methodologies need to be considered to overcome this challenge in aquaculture genetics (Palaiokostas et al., 2016).

The availability and continual development of dedicated software, such as COLONY, Cervus, and APIS, provide powerful tools for parentage assignment (Griot et al., 2020). These software packages allow for the effective analysis of genetic data, further facilitating the application of parentage assignment in aquaculture. In fact, the combination of modern genotyping technologies and sophisticated software tools has been fundamental in the increasing effectiveness of parentage assignment in aquaculture (Griot et al., 2020).

In conclusion, parentage assignment plays a crucial role in aquaculture. It significantly contributes to understanding the genetic basis of traits, designing effective breeding programs, and studying various traits that are important for the aquaculture industry (Vandeputte & Haffray, 2014). Despite existing challenges, advancements in genotyping technologies, such as next-generation sequencing, and developments in statistical methodologies promise a brighter future for parentage assignment. Refined statistical techniques have simplified the analysis of complex genetic data, and new high-speed sequencing methods allow us to study more markers simultaneously, enhancing the accuracy of parentage determination (Grashei et al., 2018).

2. Materials & Methods

In this study, we aimed to conduct parentage assignment and assess its accuracy using the APIS (Auto-Adaptive Parentage Inference Software) package in RStudio (version 2023.03.0) on a common carp (Cyprinus carpio) population. The individuals used in this study originated from a breeding program initiated at the University of South Bohemia in České Budějovice, Czech Republic, in May 2014. This program employed artificial insemination techniques and involved four factorial crosses consisting of five dams and ten sires each, totalling 20 dams and 40 sires.



2.1 Dataset and Software

2.2 The dataset used in this study ("carp_geno.txt") consisted of a total of 1345 common carp individuals, including 20 dams and 40 sires, which constituted the parental population. The remaining 1285 individuals represented the offspring resulting from the mating of the previously mentioned sires and dams. An exclusion based methodology had been previously applied to the population, and so the parentage of the offspring individuals was already known and informed in the sire and dam columns of the dataset. The results obtained using APIS (likelihood method) was later compared to the exclusion based method. It is important to note that the sire and dam columns contained "NA" values for the parental individuals, as the dataset consisted of only two generations. The remaining columns corresponded to 15,615 Single Nucleotide Polymorphisms (SNPs), and the genotypes of the individuals were represented as 0, 1, 2, or "NA," indicating the absence of an allele, the presence of a single copy, the presence of two copies and missing information, respectively.This study consisted in analyzing genetic data, meaning that prior steps such as tissue collection, DNA extraction, DNA quantification and quality check, DNA amplification, SNP genotyping and its quality control had been previously done and were not performed in this study.Data Preparation

The initial step involved reading the "carp_geno.txt" dataset into R using the "read.table" function. Subsequently, the genotype data for the offspring population (comprising lines 61 to 1345) was extracted from the dataset. Row

names were assigned to the offspring genotype data based on their corresponding IDs, and the ID column was then removed.

The parent genotypes were extracted from the first 60 rows of the dataset. Similarly, row names were assigned to the parent genotype data based on individual IDs, and the ID column was subsequently removed. The sire and dam genotypes were separated by filtering the rows based on the IDs ending with "_M" and "_F," respectively. This process resulted in separate datasets containing sire and dam genotypes. To facilitate further analysis, the offspring, sire, and dam genotypes were converted into matrices.

The scripts used for the following analysis can be found in the Appendix section.

2.3 APIS Analysis

The APIS analysis was performed using the APIS function provided by the APIS package in RStudio following the methodology described in Griot et al., 2020. The APIS function primarily employs a likelihood-based methodology to infer parentage relationships based on genotype information. This approach allows for a more nuanced and probabilistic assessment of parentage, taking into account the likelihood of different parent-offspring relationships given the observed genotypes. For our analysis, the function took the offspring, sire, and dam genotypes, along with an error rate of 0.05, as inputs. Further analyses were conducted using the full set of 15,615 SNPs, but with varying error rates of 0.01 and 0.025 to assess the sensitivity of the results to changes in the error rate.

2.3.1 SNPs Subsets Analysis

Additionally, we investigated the impact of marker set size on the accuracy of parentage assignment in a carp population. We used APIS package to assign parentage based on genetic markers. In order to evaluate the assignment power and estimate the number of offspring with at least one missing parent, different subsets of SNPs were selected arbitrarily. The marker subsets included in our analysis consisted of sets 100, 500, 1,000, 5,000, 10,000, and the full set of 15,615 SNPs.

2.3.2 Interval-Based SNP Selection Analysis

In addition to the SNP subsets analysis, we also conducted an interval-based SNP selection analysis. In this analysis, SNPs were selected based on their positions in the common carp reference genome assembly version GCA_000951615.2 (Xu et al., 2014). This genome assembly was derived from 2.8 billion paired end reads, with approximately 82% of these reads passing initial quality filters. The assembly also identified 397,047 putative RAD loci with a mean coverage of 21X at different intervals: every 150.000, 500.000, and 1.000.000 base pairs. This resulted in different sets of SNPs for each selection method. The selection interval of 150,000 base pairs was chosen as they resulted in a selection of 194 SNPs, facilitating a comparative analysis with the chromosome-specific study that was based on chromosomes containing around 200 SNPs. The 500.000 and 1.000.000 base

intervals yielded 59 and 30 SNPs, respectively. The APIS analysis was then performed for each set of SNPs. The results for each set of SNPs were compared to evaluate the impact of SNP selection on parentage assignment.

2.3.3 Chromosome Specific Analysis

We performed an analysis specific to chromosomes with the intention of investigate the impact of SNP selection on parentage assignment. For this, SNPs were selected based on their position on the chromosome. We ran the APIS function for each chromosome with at least 200 SNPs (LN590700.1, LN590703.1, LN590718.1, LN590685.1, LN590687.1, LN590684.1, LN590695.1, LN590678.1, LN590690.1, LN590696.1, LN590692.1, LN590686.1, LN590706.1, LN590683.1, LN590711.1, LN590688.1) and observed the assignment power. The different chromosomes are referenced by unique code (e.g., LN590700.1), which are identifiers for specific versions of the chromosome data in genetic databases. The number of SNPs per chromosome ranged from 204 to 275. The APIS analysis was performed for each chromosome separately. This allowed us to compare the assignment power of each chromosome to the assignment power of a randomly selected set of 200 SNPs.

2.3.4 Random SNP Selection

In addition, we also conducted an analysis using a randomly selected set of 200 SNPs from the original dataset. The reason for this was to compare to the chromosome-specific analysis as well as comparing to the interval-based SNP selection every 150.000 base pairs, which accounted for 194 SNPs, allowing us to observe the impact of SNP selection on parentage assignment. The APIS function was run on this randomly selected set of SNPs, and the assignment power was compared to that of the chromosome-specific analyses and interval-based selection.

3. Results

3.1 APIS Analysis

The parentage assignment analyses were conducted using various levels of marker number, starting from 100 and extending up to 15,615 SNPs. Throughout these analyses, we observed a consistent improvement in the parentage assignment results of the carp population as the number of markers used increased.

The theoretical assignment power is a measure of how well a set of genetic markers can distinguish between potential parents and correctly assign offspring to their true parents. It is the probability that a randomly chosen non-parent will be excluded as a potential parent based on the genetic markers used. A theoretical assignment power of 100% means that the markers used are highly effective in distinguishing between potential parents and can accurately assign offspring to their true parents.

The initial analysis conducted with 100 markers estimated that between 84 and 88 offspring had at least one missing parent, with a theoretical assignment power of 99.973% and an actual assignment rate of 71.673%. However, as the marker number increased to 500, we observed a substantial decrease in the estimation of offspring with missing parents, reducing the number to 24. Simultaneously, the assignment rate reached 100%, marking a significant improvement in the accuracy of parentage assignments. It is important to note that the 100% assignment rate indicates that each offspring was assigned to at least one parent, not necessarily both.

Further increasing the marker number to 1,000 led to a further reduction in the estimated number of offspring with missing parents, falling to 18. This increment in markers did not affect the assignment rate, which remained at 100%.

When we increased the marker number to 5,000, we observed an even further reduction in the number of offspring with at least one missing parent, with the number dropping to 10. The parentage assignment power consistently remained at 100%, indicating that the larger set of markers did not negatively impact the robustness of the analysis.

In the analysis with 10,000 markers, we observed that the estimated number of offspring with at least one missing parent fell to a range around 8. Both the assignment rate and assignment power consistently remained at 100%, underscoring the high accuracy of the analysis even with a substantial number of markers.

When using the complete set of 15,615 SNPs, the estimated number of offspring with at least one missing parent remained at 8, indicating the continued precision of the analysis even with the increased number of markers. Both the assignment rate and the assignment power reached 100% with this extensive marker set, further indicating the robustness and suitability of these markers for parentage assignment in this population. The assignment error rate accepted during this analysis was 0.05, showing that our analysis maintained high assignment accuracy even while accommodating an error rate of up to 5%.

The parentage assignment analyses were performed considering three different error thresholds: 0.01, 0.025, and 0.05. The set of markers used had a theoretical assignment power of 100% in all three scenarios, indicating their potential to accurately assign offspring to their respective parents. Across all thresholds, the estimated number of offspring with at least one missing parent remained constant at eight.

The theoretical assignment power of the marker set was 100% for all three error rates, indicating that the chosen SNP markers were highly informative and capable of distinguishing between potential parents.

The assignment rate, which measures the proportion of offspring that were successfully assigned to a pair of parents, was also 100% for all three error rates. This high assignment rate demonstrates the effectiveness of the APIS package in accurately assigning parentage based on the SNP genotypes.

To further display the patterns and trends observed in these analyses, the following plots have been created to help visualization and understanding of the parentage assignment results, especially in what concerns Delta values, Mendelian transmission probabilities, and mismatches. In this context, Delta values represent the differences probability scores between the most likely parent-pair to the second most likely. In the plots, the red color relates to the most likely parent-pair and blue is attributed to the second most likely parent-pair. The higher the Delta value, the greater the confidence in the parentage assignment. The Average Mendelian Transmission Probability evaluates the likelihood that an offspring's genotype aligns with Mendelian inheritance from their potential parents. Higher probabilities indicate a better match and increased likelihood that the individuals are true parent-offspring pairs. Mismatches refer to genotypic discrepancies between offspring and potential parents, meaning the number of SNPs that do not match between the offspring and the potential parent, given Mendelian inheritance laws. Fewer mismatches suggest a better match between the genotype of the offspring and the assigned parent, however, some mismatches are expected due to mutations and genotyping errors.

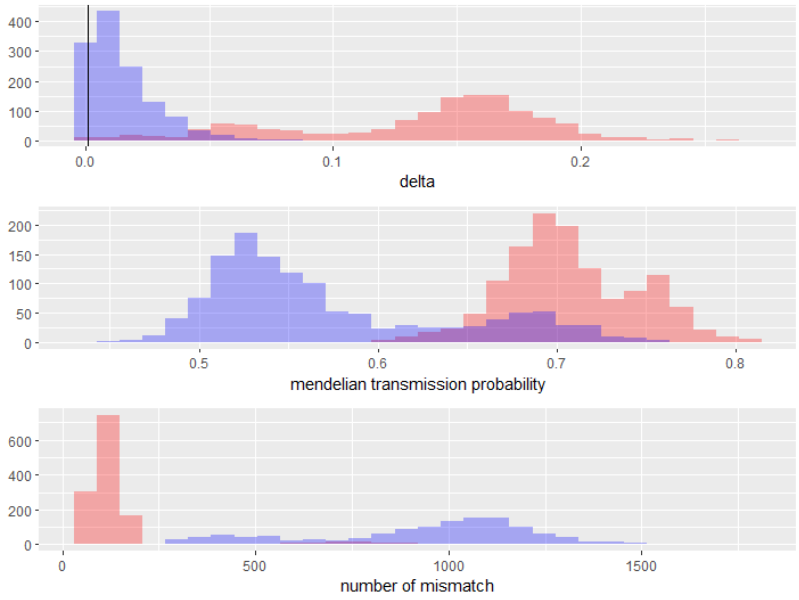


Figure 1. APIS outputs for 15.615 SNPs analysis. The top plot shows a moderate delta score for the most likely-parent pair, ranging from 0.1 to 0.2, with most cases at 0.15. The middle plot indicates that the probability score for the most likely parent-pair (in red) ranged from 0.6 to 0.8, with most cases falling into the score of 0.7. The second most likely parent-pairs ranged from 0.5 to 0.75, however the most frequent probability was observed between score 0.5 and 0.55. The bottom plot reveal the number of mismatches observed for the two most likely parent pairs. It shows a significant difference on the number of SNPs that did not match between the two most probable parent-pairs. Despite the probability scores observed in the middle plot, mismatches are still detected, suggesting some genotyping errors or mutations.

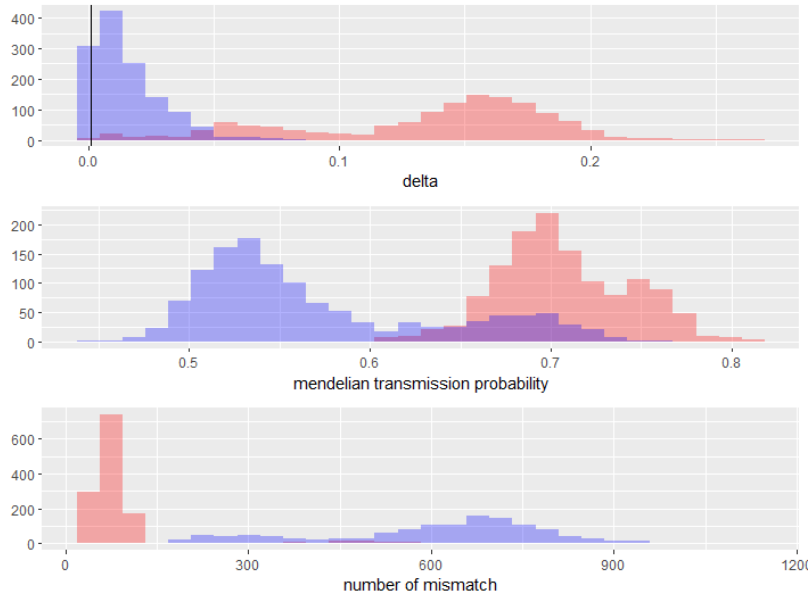


Figure 2. APIS outputs for 10.000 SNPs analysis. When lowering the number of SNPs taken into analysis, we observe that the delta scores and the Mendelian transmission probability remain similar to that of the analysis using 15.615, indicating that the 10.000 SNPs subset is still powerful in assigning parents accurately. The number of mismatches diminishes as the number of SNPs taken

into account also diminishes, which is expected as the number of available information is also reduced.

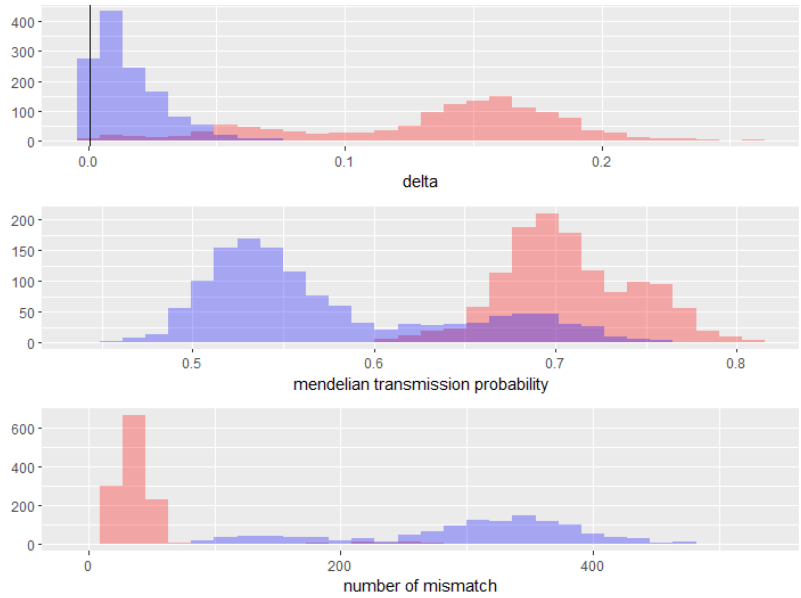


Figure 3. APIS outputs for 5.000 SNPs analysis. Despite the reduction of SNPs APIS yielded consistent results with the two previous analyses, suggesting that this density of markers still performs well. The probability scores between the two most likely parents slowly come closer together. The number of mismatches further decreases as the number of SNPs is reduced.

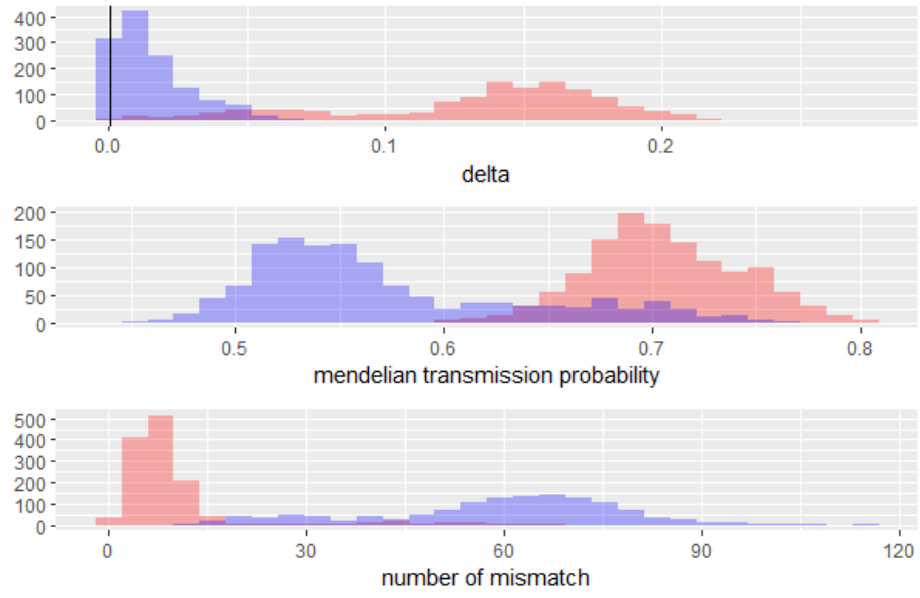


Figure 4. APIS outputs for 1.000 SNPs analysis. As we further reduce the density of SNPs, the more the probability scores intertwine with one another, suggesting a reduction in APIS' confidence in assigning parents.

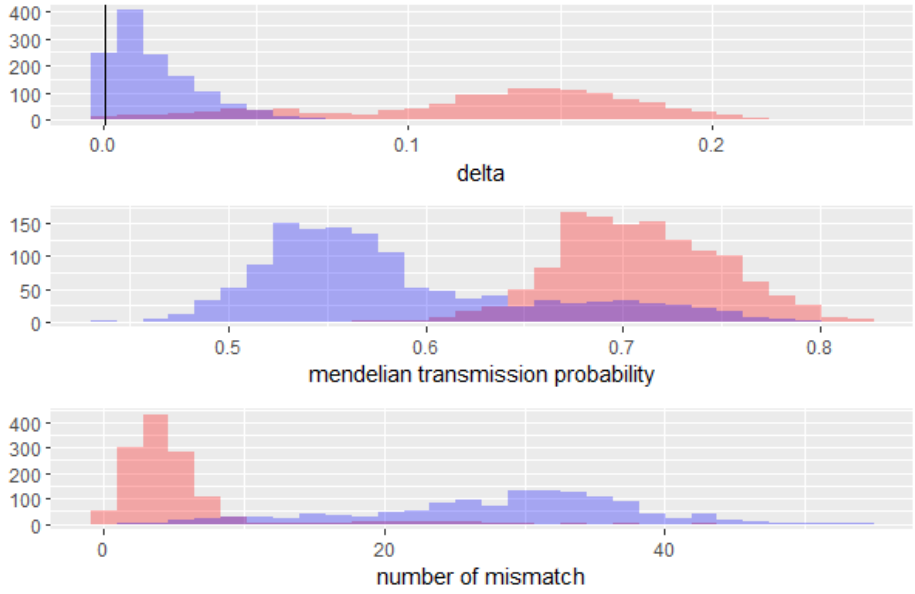


Figure 5. APIS outputs for 500 SNPs analysis. Similar delta distribution as other subsets. Probability scores peak around 0.7, showing a consistent high level of confidence in assignments.

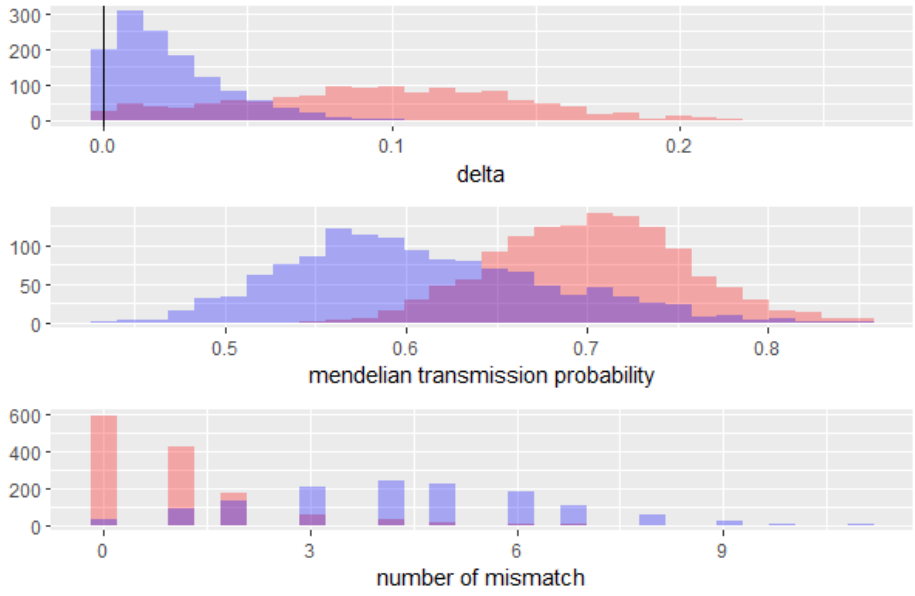


Figure 6. APIS outputs for 100 SNPs analysis. As the number of SNPs available reaches its minimal point, APIS still shows ability to distinguish parents, however as indicated by the top two plot, the values attributed to the most likely parent-pair (in red) and those to the second most likely parent (in blue), APIS' confidence is reduced as fewer information (SNPs) is available. This is consistent with the number of mismatches of the top two most probable parents intertwining.

3.2 Comparison with Exclusion Method

In comparison, the exclusion method successfully assigned parents to the 1285 offspring individuals, but 74 of those were assigned to only one parent (either sre or dam). The APIS algorithm consistently achieved a 100% assignment rate with SNP subsets of 500, 1,000, 5,000, and 10,000, indicating a complete and accurate assignment of parentage. This trend was also observed when using the entire dataset of 15,615 SNPs, confirming the high accuracy and completeness of the APIS algorithm in assigning both sire and dam for all individuals.

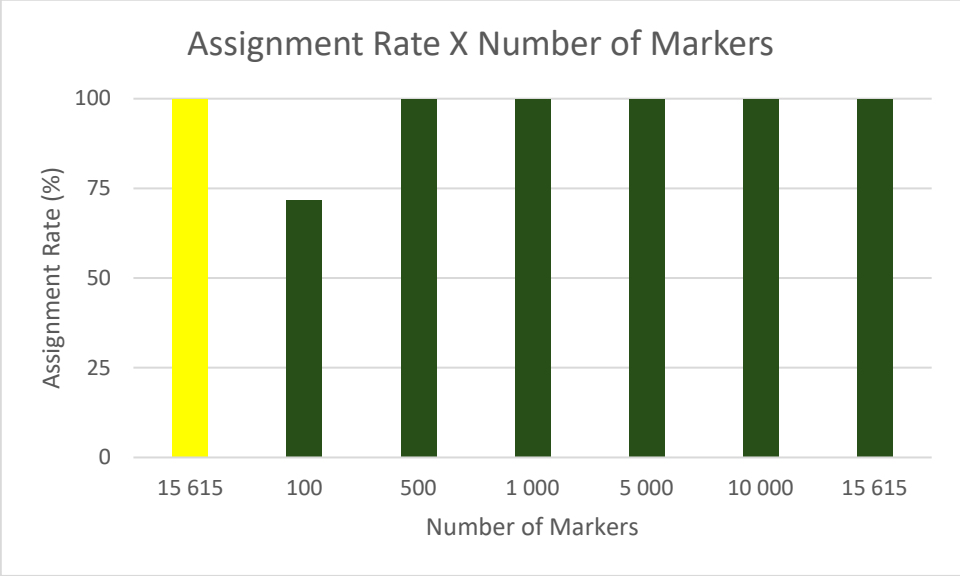


Figure 7. Assignment Rate for different marker subsets. The yellow colour indicates that an Exclusion method was used for the analysis, and the green colour indicate a Likelihood method.

3.3 SNPs Subsets Analysis

The parentage assignment analyses were conducted based on various levels of marker number: 100, 500, 1,000, 5,000, and 10,000 markers. The parentage assignment results of the carp population improved as the number of markers increased.

In the initial analysis with 100 markers, the estimated number of offspring with at least one missing parent was between 84 and 88. The theoretical assignment power of the marker set was 99.973%, and the assignment rate was 71.673%. As the marker number increased to 500, the estimation for offspring with at least one missing parents dropped to 24, but the assignment rate improved, reaching 100%. This suggests that while not all offspring were assigned to both parents, they were

successfully assigned to at least one parent. With 1,000 markers, the number of offspring with missing parents reduced further to 18. This increment in markers retained the assignment rate at 100%. When the marker number was escalated to 5,000, the estimated number of offspring with at least one missing parent was reduced to 10. The parentage assignment power remained at 100%. In the final analysis with 10,000 markers, the estimated number of offspring with at least one missing parent was between 8, with the assignment rate and assignment power remaining constant at 100%.

In each analysis, the Mendelian probability differed for each offspring and parent pair, suggesting varying degrees of accuracy in the parentage assignments.

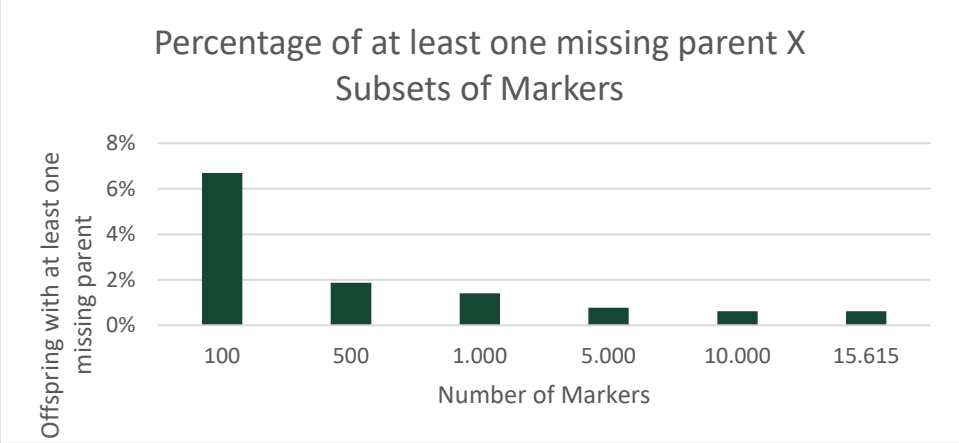


Figure 8. Percentage of at least one missing parent across different Markers Subsets.

3.4 Interval-Based SNP Selection Analysis

In the analysis with SNPs selected every 150,000 base pairs, the estimated number of offspring with at least one missing parent was 52. The theoretical assignment power of the marker set was 99.999%, and the assignment rate was 100%. The assignment error rate accepted during this analysis was 0.05.

In the analysis with SNPs selected every 500,000 base pairs, the estimated number of offspring with at least one missing parent was between 524 and 654. The theoretical assignment power of the marker set was 93.907%, and the assignment rate was 27.548%. The assignment error rate accepted during this analysis was 0.05.

In the analysis with SNPs selected every 1,000,000 base pairs, the estimated number of offspring with at least one missing parent was between 850 and 1204. The theoretical assignment power of the marker set was 13.854%, and the assignment rate was 9.4163%. The assignment error rate accepted during this analysis was 0.05.

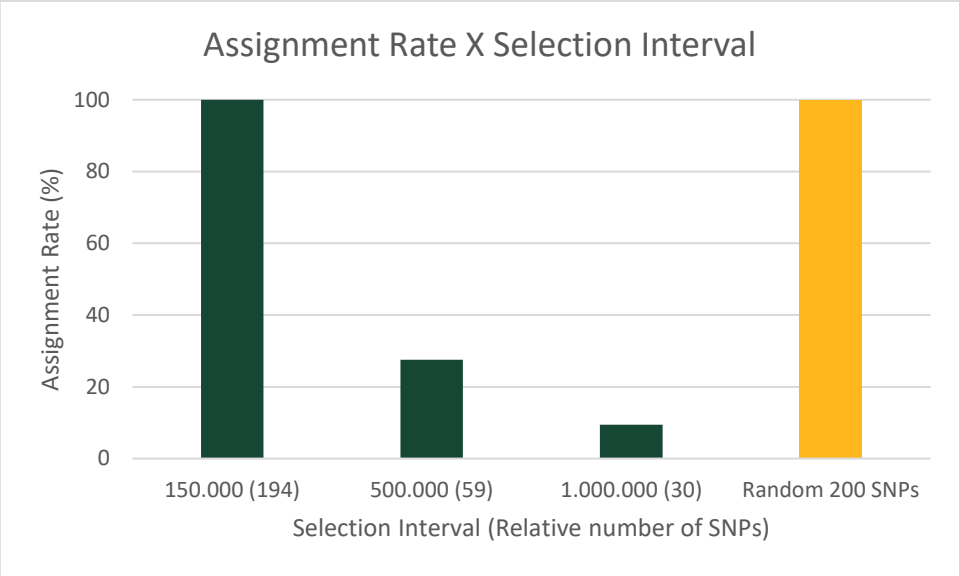


Figure 9. Assignment Rate for different Intervals of SNPs selection. The green bar indicates a selection of SNPs based on regular intervals. The yellow bar indicates that the selection of SNPs was randomized.

3.5 Chromosome-Specific Analysis

The parentage assignment analyses were conducted using the SNPs located on specific chromosomes. The estimated number of offspring with at least one missing parent varied between chromosomes, ranging from 84 to 500. The assignment rate also varied, ranging from 45.992% to 77.743%. Despite these variations, the theoretical assignment power of the marker set was consistently 99.999% for all chromosomes, indicating the robustness of these markers for parentage assignment in this population. The assignment error rate accepted during this analysis was 0.05.

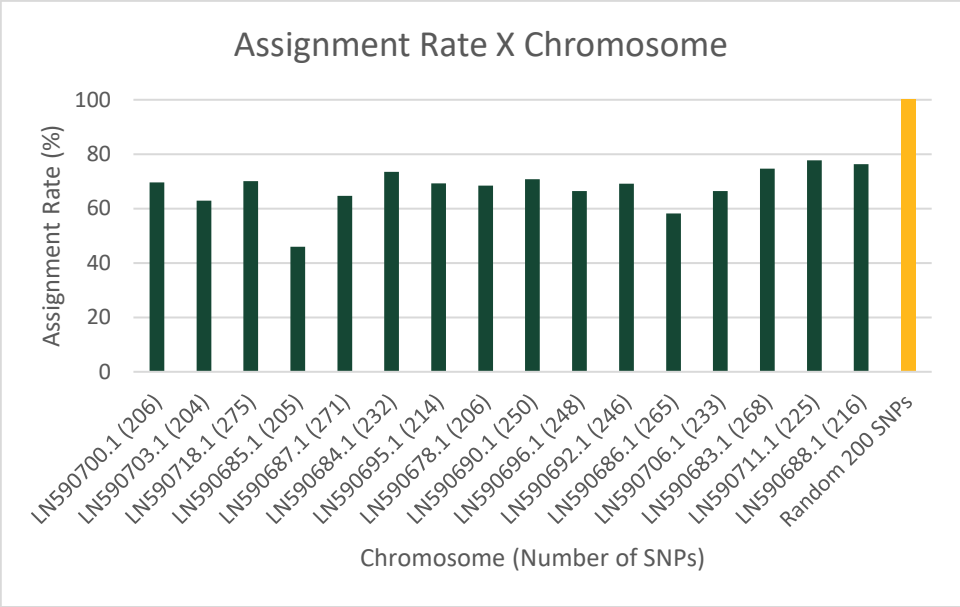


Figure 10. Assignment Rate for different chromosomal regions. The green colour indicate that the selection of SNPs was based on their genetic regions. The yellow colour indicate a random SNP selection, regardless of their position.

3.6 Random SNP Selection Analysis

In the analysis using 200 randomly selected SNPs, the estimated number of offspring with at least one missing parent was 44. The assignment rate reached 100%, indicating an improvement in the accuracy of parentage assignments compared to the chromosome-specific analysis. The theoretical assignment power of the marker set was 99.999%, and the assignment error rate accepted was 0.05.

4. Discussion

4.1 SNPs Subsets Analysis

Our results underscore the significant role of the quantity of SNP markers in parentage assignment within carp populations, consistent with previous studies that have highlighted the importance of marker density in accurate parentage assignment (Vandeputte, M., & Haffray, P., 2014). As the number of markers increased, we observed a consistent improvement in the precision and reliability of parent-offspring relationship identification. However, our findings also indicate that high assignment power can be achieved with varying marker quantities. This suggests that even when using smaller marker sets, satisfactory results can still be obtained. Aquaculture operations with budget limitations can benefit from the flexibility in marker quantity, since it allows for cost-effective genotyping without significantly affecting the accuracy (Griot et al., 2020). Although smaller subsets may result in a slightly higher number of offspring with unidentified parents, the overall accuracy remains satisfactory.

When evaluating the full marker subset of 15,615 SNPs, the consistent estimate of eight offspring with at least one missing parent underscores the continued precision of the analysis as the number of markers increased. The assignment power and rate remained constant at 100%, indicating the robustness of these markers for parentage assignment, even with an accepted assignment error rate of 0.05.

The plots generated by the APIS function indicate how Delta values from the most likely parent pair, across SNP subsets indicate a consistent genetic differentiation, suggesting that even limited markers can obtain individual genetic distinctiveness effectively. This means that the selected SNPs are informative enough to distinguish between individuals, which is essential for accurate parentage analysis. On the other hand, Delta of the second most likely parent pair values closer to 0.0 across subsets hint at potential inbreeding or a reduced genetic pool. This lower value suggests a lack of genetic variability among the SNPs, which could be due to mating between closely related individuals (inbreeding) or a small number of ancestors within the population. In terms of Mendelian transmission probabilities the results consistently showed high probabilities of genetic relatedness between offspring and potential parents across all SNP subsets, indicating a high marker robustness.

As SNP density rose, so did the number of mismatches observed. While more markers yield better genetic data, they also amplify potential mismatches,

possibly due to genotyping errors or genuine discrepancies. The steady count of offspring with at least one unidentified parent, even at peak SNP density, indicates challenges in certain parent-offspring identifications.

However, it is important to note that while the accuracy of assignment improves with the use of more markers, it does not necessarily imply that using the maximum number of markers is always the most efficient strategy. A balance must be struck between the cost and time of genotyping an increased number of markers and the improvement in parentage assignment accuracy. Furthermore, the presence of some level of mismatching in all datasets points to possible genotyping errors or mutations, indicating a fine balance between increased genomic resolution and the accepted level of assignment error. This is a limitation of our study and future research could focus on developing methods to minimize these errors. For instance, Nguyen et al. (2018) demonstrated the potential of high-throughput genotyping to identify SNPs significantly associated with economically important traits in Yellowtail Kingfish, highlighting the potential benefits of investing in detailed genotyping. This highlights the advantages and disadvantages we observed in our study between the cost of genotyping and the accuracy of parentage assignment. Additionally, Hollenbeck and Johnston (2018) discussed the genetic challenges in mollusks, such as high mutation load, which can complicate selective breeding efforts in aquaculture. These findings underscore the universal challenge of achieving a balance between genomic resolution and assignment error, further validating the importance of our study's focus.

4.2 Error Rates

These results provide insight into how the choice of error threshold can impact parentage assignments in carp. Interestingly, despite varying the error threshold, the estimated number of offspring with at least one missing parent remained constant, and all offspring were assigned to parents (100% assignment rate). This indicates that the marker set used was robust to different levels of accepted error in the assignment process, underlining the reliability and efficiency of these genetic markers in this particular population of carp.

When comparing our results to those by Griot et al.(2020), it indicates the robustness of the APIS software in handling missing parent data and genotyping errors. In their study, APIS was able to maintain a user-set acceptable error rate of 1% or 5%, even when tested on simulated data with high genotyping error rates (1% or 3%) and up to 50% missing sires. Griot et al. (2020) has demonstrated that the adaptability of APIS, under different scenarios, shows the potential of the software in several aquaculture settings, making it a valuable tool not only for researchers but breeders as well. This aligns with the findings of our study, where the error thresholds did not impact the assignment rate or the estimated number of offspring with missing parents, further validating the effectiveness of the APIS package in different scenarios.

It is important to note that while maintaining a low error threshold can increase confidence in the parentage assignments, it may also limit the ability to detect true parent-offspring relationships if there is significant genetic variation or mutation within the population. On the other hand, a higher error threshold could

potentially lead to false parent-offspring assignments. Therefore, choosing an appropriate error threshold is critical and should be based on a balance between assignment accuracy and the genetic characteristics of the population under study.

This study has provided valuable information on the effectiveness of different error thresholds in parentage assignments in carp, and the utility of large-scale genotyping for parentage assignment in this species.

4.3 Interval-Based SNP Selection Analysis

The selection of SNPs at different intervals in the genome is a critical aspect of parentage assignment, as demonstrated by our results. Our analysis confirms Houston et al. (2014) findings in Atlantic salmon that SNP density plays a crucial impact in parentage assignment. In their study, they developed a high-density SNP genotyping array for Atlantic salmon. They emphasized the importance of dense SNP genotyping arrays in studies of the genetic architecture of quantitative traits and in improving the accuracy of selection in breeding programs. In our study, as the interval between selected SNPs increased, we observed a significant decrease in parentage assignment accuracy, which is in line with the idea that a higher density of markers is usually beneficial for precise parentage assignment. This economic impact of genotyping, emphasized by Griot et al (2020), suggests that, while high-density genotyping might provide better detailed information, it may not always be the most cost-effective approach for different aquaculture production systems.

4.4 Chromosome-Specific Analysis

Our chromosome-specific analysis revealed variability in parentage assignment accuracy depending on the specific set of SNPs used. This is consistent with the findings of Liu and Cordes (2004), who discussed the impact of DNA-based genetic markers, including SNPs, in animal genetics. They highlighted the potential to observe and exploit variation in the entire genome using these markers. Our results suggest that in terms of parentage assignment, not all chromosomes are equally informative, which could be because of the distribution of SNPs across the genome, and how the population's genetic make-up is structured. This indicates how crucial it is to choose SNPs based on their position in the genome and how informative they potentially can be. This point is elaborated by Griot et al. (2020) where it is suggested that the selection of SNPs, based on their position in the genome can improve the accuracy of parentage assignment and also balance genotyping costs.

4.5 Random SNP Selection Analysis

This analysis found that a diverse set of SNPs randomly selected across the genome may provide more accurate parentage assignment than the chromosome-

specific analysis. This goes in line to what Kruuk et al. (2008) found in their study, where they analyzed how a large set of genetic marker information can help explain the relationship between environmental conditions and how genes are expressed in an organism. Their article showed the advantages of using genetic data when constructing models that try to explain how an organism evolves over time. Our findings suggest that a diverse and representative set of SNPs can show more of the genetic variation present in a population, and therefore lead to more accurate parentage assignments. This highlights the importance, in parentage assignment studies, of taking the distribution of SNPs throughout the genome into account. In his paper Griot et al. (2020) highlights that diversification in SNP selection has its benefits. He discusses how capturing a wider genetic makeup can lead to a more complete view of the population's genotype, which in turn increases accuracy in parentage assignment.

5. Conclusion

This study has investigated the role and impact of Single Nucleotide Polymorphisms (SNPs) in parentage assignment within carp populations, focusing on the factors that influence the accuracy and reliability of parentage assignments. Our analyses demonstrated that the number of SNP markers utilized significantly impacts the precision of parent-offspring relationship identification. As the number of markers increased, we observed a consistent improvement in the precision and reliability of this process. This finding aligns with previous studies emphasizing the role of marker density in accurate parentage assignments (Vandeputte & Haffray, 2014). Despite the high accuracy across varying marker quantities, we also underlined the necessity of a balance between the cost and time of genotyping a larger number of markers and the improvement in assignment accuracy.

The study examined how error thresholds can influence parentage assignments in carp. Notably, the marker set used was robust to different levels of accepted error, which shows reliability and efficiency of these genetic markers.

Exploration into interval-based SNP selection showed that SNP selection, based on genomic positions, can significantly impact parentage assignment's accuracy. Higher density of SNPs led to improved parentage assignment accuracy, while a lower density resulted in decreased accuracy.

The chromosome-specific SNP selection analyses contributed further to the understanding of how SNP selection affects parentage assignment. It demonstrated that a diverse and representative set of SNPs, selected randomly across the genome, may provide more accurate results than a set of SNPs from specific chromosomes.

In conclusion, this study has increased our understanding of parentage assignment in carp, as well as the role of SNPs in genetic research. While further research is needed to build upon these findings, this study provides a solid foundation for such work and contributes to the ongoing research in this area.

References

Andreotti, V., Chindris, A., Brundu, G., Vallainc, D., Francavilla, M. and García, J. (2017). Bioremediation of aquaculture wastewater from Mugil cephalus (Linnaeus, 1758) with different microalgae species. *Chemistry and Ecology*, 33(8), pp.750–761. doi:https://doi.org/10.1080/02757540.2017.1378351.

Ansah, Y., Frimpong, E. and Hallerman, E. (2014). Genetically-Improved Tilapia Strains in Africa: Potential Benefits and Negative Impacts. *Sustainability*, 6(6), pp.3697–3721. doi:https://doi.org/10.3390/su6063697.

Boyd, C. and McNevin, A. (2014). *Aquaculture, Resource Use, and the Environment*. [online] *The Open Library*. Wiley & Sons, Incorporated, John. Available at: https://worldcat.org/title/895388055

Boyd, C.E. and Massaut, L. (1999). Risks associated with the use of chemicals in pond aquaculture. *Aquacultural Engineering*, 20(2), pp.113–132. doi:https://doi.org/10.1016/s0144-8609(99)00010-2.

Cavalli, L.S., Da Rocha, A.F., De Brito, B.G., De Brito, K.C.T. and Rotta, M.A. (2021). Major Sustainable Development Goals applied to Aquaculture. *Pesquisa Agropecuária Gaúcha*, 27(1), pp.110–126. doi:https://doi.org/10.36812/pag.2021271110-126.

CGIAR. (n.d.). *Genetically Improved Farmed Tilapia (GIFT)*. [online] Available at: https://www.cgiar.org/innovations/genetically-improved-farmed-tilapia-gift/.

Chakraborty, R., Meagher, T. R., & Smouse, P. E. (1988). Parentage analysis with genetic markers in natural populations. I. The expected pro- portion of offspring with unambiguous paternity. Genetics, 118(3), 527.

FAO (2018). *WORLD FISHERIES AND AQUACULTURE*. [online] Available at: https://www.fao.org/3/I9540EN/i9540en.pdf.

FAO (2022). *The state of World fisheries and aquaculture 2022*. [online] www.fao.org. Available at: https://www.fao.org/3/cc0461en/online/cc0461en.html.

FAO (Food and Agriculture Organization of the United Nations). 2011. The State of the World's Land and Water Resources for Food and Agriculture. Rome: FAO

FAO (Food and Agriculture Organization of the United Nations). 2014a. The State of the World Fisheries and Aquaculture 2014. Rome: FAO.

Garcia, A., Tsuruta, S., Gao, G., Palti, Y., Lourenco, D. and Leeds, T. (2023). Genomic selection models substantially improve the accuracy of genetic merit predictions for fillet yield and body weight in rainbow trout using a multi-trait model and multi-generation progeny testing. *Genetics Selection Evolution*, 55(1). doi:https://doi.org/10.1186/s12711-023-00782-6.

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

Glaubitz, J. C., Rhodes, O. E., & Dewoody, J. A. (2003). Prospects for inferring pairwise relationships with single nucleotide poly- morphisms. Molecular Ecology, 12(4), 1039–1047. https://doi. org/10.1046/j.1365-294X.2003.01790.x

Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M. and Toulmin, C. (2010). Food Security: The Challenge of Feeding 9 Billion People. *Science*, 327(5967), pp.812–818. doi:https://doi.org/10.1126/science.1185383.

Grashei, K.E., Ødegård, J. and Meuwissen, T., 2018. Using genomic relationship likelihood for parentage assignment. Genetics Selection Evolution, 50(1).

Griot, R., Allal, F., Brard-Fudulea, S., Morvezen, R., Haffray, P., Phocas, F. and Vandeputte, M. (2020). APIS: An auto-adaptive parentage inference software that tolerates missing parents. *Molecular Ecology Resources*, [online] 20(2), pp.579–590. doi:https://doi.org/10.1111/1755-0998.13103.

Han, W., Mao, Y., Wei, Y., Shang, P. and Zhou, X. (2020). Bioremediation of Aquaculture Wastewater with Algal-Bacterial Biofilm Combined with the Production of Selenium Rich Biofertilizer. *Water*, 12(7), p.2071. doi:https://doi.org/10.3390/w12072071.

Hollenbeck, C.M. and Johnston, I.A. (2018). Genomic Tools and Selective Breeding in Molluscs. Frontiers in Genetics, 9. doi:https://doi.org/10.3389/fgene.2018.00253.

Houston, R. D., Taggart, J. B., Cézard, T., Bekaert, M., Lowe, N. R., Downing, A., ... & Bishop, S. C. (2014). Development and validation of a high density SNP genotyping array for Atlantic salmon (Salmo salar). BMC genomics, 15(1), 90.

Hulata, G. (1995). A review of genetic improvement of the common carp (Cyprinus carpio L.) and other cyprinids by crossbreeding, hybridization and selection. *Aquaculture*, [online] 129(1), pp.143–155. doi:https://doi.org/10.1016/0044-8486(94)00244-I.

Janssen, K. (2019). *The economic optimization of breeding programs in aquaculture*. [online] Available at: https://edepot.wur.nl/472285 [Accessed 12 Jul. 2023].

Kruuk, L.E.B., Slate, J. and Wilson, A.J. (2008). New Answers for Old Questions: The Evolutionary Quantitative Genetics of Wild Animal Populations. Annual Review of Ecology, Evolution, and Systematics, 39(1), pp.525–548. doi:https://doi.org/10.1146/annurev.ecolsys.39.110707.173542.

Lehnert, S.J., Heath, D.D. and Pitcher, T.E. (2012). Sperm trait differences between wild and farmed Chinook salmon (Oncorhynchus tshawytscha). Aquaculture, 344-349, pp.242–247. doi:https://doi.org/10.1016/j.aquaculture.2012.03.007.

Lillehammer, M., Meuwissen, T. H., & Sonesson, A. K. (2013). A low-marker density implementation of genomic selection in aquaculture using within-family genomic breeding values. Genetics Selection Evolution, 45(1), 1-10.

Li, Z., Yu, E., Zhang, K., Gong, W., Xia, Y., Tian, J., Wang, G. and Xie, J. (2020). Water Treatment Effect, Microbial Community Structure, and Metabolic Characteristics in a Field-Scale Aquaculture Wastewater Treatment System. *Frontiers in Microbiology*, 11. doi:https://doi.org/10.3389/fmicb.2020.00930.

Liu, Z.J. and Cordes, J.F. (2004). DNA marker technologies and their applications in aquaculture genetics. Aquaculture, 238(1-4), pp.1–37. doi:https://doi.org/10.1016/j.aquaculture.2004.05.027.

Marshall, T. C., Slate, J., Kruuk, L. E. B., & Pemberton, J. M. (1998). Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology, 7(5), 639–655. https://doi. org/10.1046/j.1365-294x.1998.00374.x

Martinez-Porchas, M. and Martinez-Cordova, L.R. (2012). *World Aquaculture: Environmental Impacts and Troubleshooting Alternatives*. [online] The Scientific World Journal. Available at: https://www.hindawi.com/journals/tswj/2012/389623/.

McKindsey, C.W., Landry, T., O'Beirn, F.X. and Davies, I.M. (2007). Bivalve aquaculture and exotic species: a review of ecological considerations and management issues. *oar.marine.ie*. [online] Available at: https://oar.marine.ie/handle/10793/1228#:~:text=McKindsey%2C%20C.W.%2C%20Landry%2C%20T.%2C%20O%27Beirn%2C%20F.X.%20and%20Davies%2C [Accessed 12 Jul. 2023].

Meuwissen, T.H.E., Hayes, B.J. and Goddard, M.E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics, [online] 157(4), pp.1819–1829. doi:https://doi.org/10.1093/genetics/157.4.1819.

Nasr-Allah, A.M. (2021). FAO_National Strategy on Aquatic Animal Health (NSAAH) Management within the Progressive Management Pathway to improve aquaculture biosecurity (PMP/AB). Second Progress Report - November 2021. *digitalarchive.worldfishcenter.org*. [online] Available at: https://digitalarchive.worldfishcenter.org/handle/20.500.12348/5170 [Accessed 12 Jul. 2023].

Naylor, R., Hindar, K., Fleming, I.A., Goldburg, R., Williams, S., Volpe, J., Whoriskey, F., Eagle, J., Kelso, D. and Mangel, M. (2005). Fugitive Salmon: Assessing the Risks of Escaped Fish from Net-Pen Aquaculture. *BioScience*, 55(5), p.427. doi:https://doi.org/10.1641/0006-3568(2005)055[0427:fsatro]2.0.co;2.

Naylor, R.L., Goldburg, R.J., Primavera, J.H., Kautsky, N., Beveridge, M.C.M., Clay, J., Folke, C., Lubchenco, J., Mooney, H. and Troell, M. (2000). Effect of aquaculture on world fish supplies. *Nature*, 405(6790), pp.1017–1024. doi:https://doi.org/10.1038/35016500.

Nguyen, N.H., Rastas, P.M.A., Premachandra, H.K.A. and Knibb, W. (2018). First High-Density Linkage Map and Single Nucleotide Polymorphisms Significantly Associated With Traits of Economic Importance in Yellowtail Kingfish Seriola lalandi. Frontiers in Genetics, 9. doi:https://doi.org/10.3389/fgene.2018.00127.

Nguyen, N.H., Sonesson, A.K., Houston, R.D. and Moghadam, H. (2022). Editorial: Applications of Modern Genetics and Genomic Technologies to Enhance Aquaculture Breeding. *Frontiers in Genetics*, 13. doi:https://doi.org/10.3389/fgene.2022.898857.

Ozbay, G., Blank, G. and Thunjai, T. (2014). Impacts of Aquaculture on Habitats and Best Management Practices (BMPs). *Sustainable Aquaculture Techniques*. doi:https://doi.org/10.5772/57471.

Palaiokostas, C., Cariou, S., Bestin, A., Jean-Sébastien Bruant, Pierrick Haffray, Morin, T., Joëlle Cabon, François Allal, Vandeputte, M. and Houston, R.D. (2018). Genome-wide association and genomic prediction of resistance to viral nervous necrosis in European sea bass (Dicentrarchus labrax) using RAD sequencing. *Genetics Selection Evolution*, 50(1). doi:https://doi.org/10.1186/s12711-018-0401-2.

Palaiokostas, C., Ferraresso, S., Franch, R., Houston, R. D., & Bargelloni, L. (2016). Genomic prediction of resistance to pasteurellosis in gilthead sea bream (Sparus aurata) using 2b-RAD sequencing. G3: Genes, Genomes, Genetics, 6(11), 3693-3700.

Price, E.O. (1984). Behavioral Aspects of Animal Domestication. *The Quarterly Review of Biology*, 59(1), pp.1–32. doi:https://doi.org/10.1086/413673.

Radanović, A., Sprycha, Y., Jocković, M., Sundt, M., Miladinović, D., Jansen, C. and Horn, R. (2022). KASP Markers Specific for the Fertility Restorer Locus Rf1 and Application for Genetic Purity Testing in Sunflowers (Helianthus annuus L.). *Genes*, [online] 13(3), p.465. doi:https://doi.org/10.3390/genes13030465.

Rawson Jr., M.V., Chen, C., Ji, R., Zhu, M., Wang, D., Wang, L., Yarish, C., Sullivan, J.B., Chopin, T., Carmona,R. 2002. Understanding the interaction of extractive and fed aquaculture using ecosystem modeling. In:Stickney, R.R., McVey, J.P. (Eds.), Responsible Marine Aquaculture. CABI Publishing, Oxon, pp. 263– 2961.

Romero-Soto, I.C., Dia, O., Leyva-Soto, L.A., Drogui, P., Buelna, G., Díaz-Tenorio, L.M., Ulloa-Mercado, R.G. and Gortáres-Moroyoqui, P. (2018). Degradation of Chloramphenicol in Synthetic and Aquaculture Wastewater Using Electrooxidation. *Journal of Environmental Quality*, [online] 47(4), pp.805–811. doi:https://doi.org/10.2134/jeq2017.12.0475.

Sampantamit, T., Ho, L., Van Echelpoel, W., Lachat, C. and Goethals, P. (2020). Links and Trade-Offs between Fisheries and Environmental Protection in Relation to the Sustainable Development Goals in Thailand. *Water*, 12(2), p.399. doi:https://doi.org/10.3390/w12020399.

Singh, A.K., Pathak, A.K. and Lakra, W.S. (2010). Invasion of an Exotic Fish—Common Carp, *Cyprinus Carpio* L. (Actinopterygii: Cypriniformes: Cyprinidae) in

the Ganga River, India and its Impacts. *Acta Ichthyologica Et Piscatoria*, 40(1), pp.11–19. doi:https://doi.org/10.3750/aip2010.40.1.02.

Thodesen, J. and Gjedrem, T. (2006). *Breeding programs on Atlantic salmon in Norway: lessons learned*. [online] *aquadocs.org*. WorldFish Center. Available at: https://aquadocs.org/handle/1834/20390 [Accessed 12 Jul. 2023].

Troell, M., Costa-Pierce, B.A., Stead, S.M., Cottrell, R.S., Brugere, C., Farmery, A.K., Little, D.C., Strand, Å., Pullin, R.S.V., Soto, D., Beveridge, M., Salie, K., Dresdner, J., Moraes-Valenti, P., Blanchard, J.L., James, Rodrigue Yossa, Allison, E.H., Devaney, C. and Barg, U. (2023). Perspectives on aquaculture's contribution to the Sustainable Development Goals for improved human and planetary health. Journal of the World Aquaculture Society, 54(2), pp.251–342. doi:https://doi.org/10.1111/jwas.12946.

Tsai, H. Y., Hamilton, A., Tinch, A. E., Guy, D. R., Bron, J. E., Taggart, J. B., ... & Bishop, S. C. (2016). Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. Genetics Selection Evolution, 48(1), 1-10.

Vallejo, R.L., Leeds, T.D., Gao, G., Parsons, J.E., Martin, K.E., Evenhuis, J.P., Fragomeni, B.O., Wiens, G.D. and Palti, Y. (2017). Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture. *Genetics Selection Evolution*, 49(1). doi:https://doi.org/10.1186/s12711-017-0293-6.

Vandeputte, M., & Haffray, P. (2014). Parentage assignment using genomic markers: a major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. Frontiers in Genetics, 5, 432.

Vandeputte, M. and Prunet, P. (2002). Genetics of adaptation in fish: Effects of domestication, stress resistance and adaptation to the environment. *INRAE Productions Animales*, 15(5), pp.365–371. doi:https://doi.org/10.20870/productions-animales.2002.15.5.3716.

Waite, R., Beveridge, M. and Chaiyawannakarn, N. (2014). *Improving Productivity and Environmental Performance of Aquaculture. The Challenge Program on Water and Food (CPWF); Nile Basin Focal Project: Aquaculture Nile Delta -Water budgeting for Aquaculture Production View project Aquaculture Research 30 (11-12), 901-905 View project*.

Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., Xu, J., Zheng, X., Ren, L., Wang, G., Zhang, Y., Huo, L., Zhao, Z., Cao, D., Lu, C., Li, C., Zhou, Y., Liu, Z., Fan, Z. and Shan, G. (2014). Genome sequence and genetic diversity of the

common carp, Cyprinus carpio. Nature Genetics, 46(11), pp.1212–1219. doi:https://doi.org/10.1038/ng.3098.

Zhang, Z., Lin, Z., Wei, M., Chen, Z., Shen, M., Cao, G., Wang, Y., Zhang, Z. and Zhang, D. (2022). Development of Single Nucleotide Polymorphism and Association Analysis with Growth Traits for Black Porgy (Acanthopagrus schlegelii). *Genes*, 13(11), p.1992. doi:https://doi.org/10.3390/genes13111992.

Popular science summary

In aquaculture and conservation biology, accurately identifying parent-offspring relationships is crucial. Our research focused on carp, using genetic markers known as Single Nucleotide Polymorphisms (SNPs) to determine parentage. The study revealed that the number of these markers significantly influences the accuracy of parent-offspring identification.

We tested a range of marker quantities, from 100 up to 15,615. Our findings indicate that even though increasing the number of markers led to more accurate results, the markers yielded reliable results for parentage assignment already at lower quantities. This is particularly beneficial for aquaculture operations that may have budget constraints.In addition to the number of markers, the study also examined the impact of their genomic locations. We found that a diverse set of markers, spread across different regions of the genome, provided the most accurate parentage assignments. This insight could guide future research and practical applications in selective breeding programs.

We also explored the role of error rates in the assignment process. Interestingly, the study showed that the method was robust even when different levels of error were permitted, highlighting the reliability of these genetic markers for parentage assignment in carp. Understanding the genetic relationships within carp populations has broader implications for both conservation efforts and the aquaculture industry. Our research contributes valuable insights into the optimization of genetic markers for parentage assignment, balancing both accuracy and cost-effectiveness.

Acknowledgements

To my family, I cannot express in words how much your support and love have meant to me during this process. You have stood by me through every step, providing endless encouragement and strength. This accomplishment would not have been possible without your unwavering faith in me.

I would first like to express my deepest gratitude to my supervisor, Dr. Christos Palaiokostas, whose expertise, guidance, and support have been fundamental in the completion of this research. Your teachings have enriched my educational journey immensely.

I am deeply thankful to my dear friend, Daria Ketrzynska. From assisting me with navigating Canvas, to helping me find class locations, to ensuring I never missed an important deadline, your help has been invaluable. Your support from the very beginning has made this journey smoother and more enjoyable, and for that, I am truly grateful.

I would also like to thank my housemates. Living with you has not only been a joy but also a great support during this process. Your understanding and patience during stressful times, our chess games, shared moments of relaxation and laughter, and your support have made our house feel more like home. You've contributed significantly to making this challenging journey an enjoyable experience.

Lastly, to all my friends who have shared in this journey, your constant positivity, companionship, and understanding have been a tremendous source of motivation. Your faith in my abilities and your reminders to take a break and enjoy life have made this experience so much more fulfilling.

Appendix 1

APIS Analysis Script:

```
library(APIS)
carp_geno <- read.table("carp_geno_apis.txt", header = TRUE)

# Extract genotype data for the offspring population (lines 61 to 1345)
offspring_geno <- carp_geno[61:1345,]
rownames(offspring_geno) <- offspring_geno$Id
offspring_geno <- offspring_geno[-1]

# Extract parent genotypes
parent_geno <- carp_geno[1:60, ]

# Set rownames as individual IDs
rownames(parent_geno) <- parent_geno$Id
parent_geno <- parent_geno[-1]

# Extract Sire and Dam from parent_geno
sire_ids <- rownames(parent_geno)[grepl("_M$", rownames(parent_geno))]
dam_ids <- rownames(parent_geno)[grepl("_F$", rownames(parent_geno))]
sire_geno <- parent_geno[sire_ids, ]
dam_geno <- parent_geno[dam_ids, ]

#Convert genotype to character matrices
off <- as.matrix(offspring_geno)
sire <- as.matrix(sire_geno)
dam <- as.matrix(dam_geno)

# Run APIS analysis
results <- APIS(off.genotype = off,
                sire.genotype = sire,
                dam.genotype = dam,
                error = 0.05)
```

Publishing and archiving

Approved students' theses at SLU are published electronically. As a student, you have the copyright to your own work and need to approve the electronic publishing. If you check the box for **YES**, the full text (pdf file) and metadata will be visible and searchable online. If you check the box for **NO**, only the metadata and the abstract will be visible and searchable online. Nevertheless, when the document is uploaded it will still be archived as a digital file. If you are more than one author, the checked box will be applied to all authors. You will find a link to SLU's publishing agreement here:

- https://libanswers.slu.se/en/faq/228318.

☒ YES, I/we hereby give permission to publish the present thesis in accordance with the SLU agreement regarding the transfer of the right to publish a work.

☐ NO, I/we do not give permission to publish the present work. The work will still be archived and its metadata and abstract will be visible and searchable.