Carl von Ossietzky University, Oldenburg

Faculty V – Mathematics & Natural Sciences

AG Ecological Genomics, Prof. Dr. Arne Nolte

Master programme:

Marine Umweltwissenschaften

Title:

*"*The establishment and application of a genotyping assay to study the sex ratios of pike populations from Rügen*"*.

Submitted by:

Viktor de Oliveira

Matrikelnr: 5565097

Submitted to:

Supervising Consultant:

Prof. Dr. Arne Nolte

arne.nolte@uni-oldenburg.de

Second Consultant:

Dr. Stefan Dennenmoser

stefan.dennenmoser@uni-oldenburg.de

Oldenburg, 14.08.2023

Eidesstattliche Erklärung

Ich versichere, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und die allgemeinen Prinzipien

wissenschaftlicher Arbeit und Veröffentlichungen, wie sie in den Leitlinien guter wissenschaftlicher Praxis der Carl von Ossietzky Universität Oldenburg festgelegt sind, befolgt habe.

14.08.2023 Victor de Oliveira
Ort, Datum Unterschrift

Acknowledgments

I sincerely want to thank the core team of the AG Ecological Genomics for making this thesis possible. Many thanks to Malte Dittman, who showed me all the laboratory procedures from scratch and shared his expertise while being incomparable polite, funny, and professional at the same time. Many thanks to Stefan Dennenmoser, who introduced me to the bioinformatic components of this work, who was always consultable, extremely constructive, and helpful, no matter what the problem was. Many thanks to Arne Nolte of course, who always had the ability to see the bigger picture in things when I didn't. Without his expertise and guidance this thesis would not exist. Many thanks also to Robert Arlinghaus, for letting me join the field trips. Many thanks to Félicie Dhellemmes and Timo Rittweg for the amazing field experience and sampling trip, that served as a base for this project. And finally, many thanks to the numerous pike which unwillingly became part of this work, may your contribution result in something useful.

Abstract

The Northern Pike (Esox lucius) has been at the forefront of research focused on sex determination systems in recent years. The characterization of a new master sex determining (MSD) gene, and the proceeding breakdown of genetic variations among pike populations, resulted in a by this time impressive state of knowledge in regard to this species' genome. This study aimed for the establishment of a genotyping assay to verify the sex determination system in Northern Pike populations from Rügen, based on the existing knowledge of the involved genes. A straightforward PCR assay was established based on a pool of confirmed male and female individuals, and successfully verified the presence of the presumed sex determination system in these populations. The gained knowledge was further used to establish a marker for Next Generation Sequencing (NGS), enabling an *in-silico* sexing approach. External sex determination in Northern Pike, primarily based on morphological differences and the examination of the urogenital opening, has been found to be unreliable. This methods are particularly inconsistent during the spawning season and in earlier life stages. The implementation of a genotype-based sex determination assay, once established, offers numerous advantages, including its consistency, reliability, and applicability across various developmental stages and seasonal variations. The established marker allowed for the comparison of genotyped and externally sexed individuals and revealed a discrepancy in the external sex determination in approximately 9,2 % of the individuals. The majority of the incorrectly sexed individuals were sampled during spawning season, underlining the limitations of external sexing during that period. On one side the findings show the potential of straight forward non-invasive genotyping assays for sex determination, on the other side they revealed the challenges in obtaining unbiased estimates of sex ratios due to selective sampling methods and sex specific growth.

Content

List of Figures

List of Abbreviation

GSD - Genetic sex Determination

SD - Sex Determination

ESD - Environmental Sex Determination

SNP - Single Nucleotide Polymorphism

NGS - Next Generation Sequencing

IBD - Isolation by Distance

IBR - Isolation by Resistance

IBE - Isolation by Environment

PSU - Practical Salinity Units

PCR - Polymerase Chain Reaction

LG - Linkage Group

amh- Anti-Müllerian Hormone

amhby - Y-specific anti-Müllerian hormone paralog

1. Introduction

1.1 Sex determination

Comprehensive biological studies within the animal kingdom are unconceivable without the consideration of the sex of an individual. On one side, an individual's characteristics, traits and behaviour pattern can differ between the sex, on the other side the relative abundance of males and females within a population significantly affects its overall success (Johnson, 2010a; Ancona et al., 2017). Sex determination occurs during early development when specific signals are detected or not detected by gonadal cells, leading to the activation of pathways that direct the development of male or female physiology. This signal can originate from genetic factors, environmental factors, or a combination of both (Johnson, 2019). The underlying male or female phenotypes are controlled by different sex determination (SD) mechanisms that can be divided into genetic sex determination (GSD) and environmental sex determination (ESD) (Heule et al., 2014). The most prominent model for sex determination is the genetic sex determination (GSD), in which the genotype determines the male and female phenotype. Here, a specific gene expression initiates a sex differentiation cascade, leading to the development of either testes or ovaries in the embryo (Suzuki and Iwasa, 1980; Valenzuela and Lance, 2004). The gene that initiates this cascade of events leading to the differentiation towards one sex or another is called master sex determiner (MSD). In most mammals for instance, a male specific Y chromosome contains the gene that initiates the differentiation towards the male phenotype, directs the development of testes and suppresses the development of ovaries, making it the MSD in this example (Johnson, 2019; Pan et al., 2021). On the other hand, in many other vertebrates, environmental conditions can be the primary sex determination factor. Environmental sex determination (ESD) is defined as a condition in which environmental influences primary determine the sex of in individual which has been exposed to the latter during a critical period of the embryogenesis (Valenzuela and Lance, 2004). Environmental factors that have been shown to influence the sex determination in a variety of species are for instance: Temperature, stress, and pH. However, most pronounced is the effect of temperature on the sex differentiation during embryogenesis, due to its effects on enzymatic activities. The correlation of temperature and sex has been shown in several reptiles, amphibians, and fish. While breeding experiments with crocodiles and turtles (Johnson, 2010; Jensen et al., 2018) lead to predominantly female

hatchlings in warmer temperatures, the opposite was the case for fish species like the goldfish (Devlin and Nagahama, 2002; Goto-Kazeto et al., 2006).

1.2 Sex determination in fishes

Because of its apparent association with reproduction and the Darwinian fitness, sex determination was traditionally supposed to be a rather conserved trait, but research on teleost fish within the last decades revealed that especially this clade displays an exceptional variety of sex determination pathways (Avise and Mank, 2009). On one side, teleost fish show both genetic and environmental sex determination, on the other side they exhibit different mechanisms within their sex determination and reproduction systems. Alongside several types of monofactorial and polygenic systems, they show for vertebrates rather exceptional reproduction strategies, including gonochorism, different forms of hermaphroditism, and unisexuality, that respectively originated numerous times independently within this clade (Devlin and Nagahama, 2002; Avise and Mank, 2009; Martinez et al., 2014; Pan et al., 2021a; Pla et al., 2022). Teleost fish constitute the most species-rich clade of all vertebrates, and their sex determination systems show an exceptional lability, resulting in relatively rapid turn-over processes within different sex determination modes, making them a perfect model to study the evolution of sex determining adaptations (Avise and Mank, 2009; Heule et al., 2014). As comprehensively reviewed by Devlin and Nagahama (2002**)** and Heule et al. (2014), the deciphering of the entirety of the sex determination systems of teleost fishes enabled crucial insights into the plasticity of the sex determination processes in vertebrates. Within the last decade the research of teleost fish revealed several new master regulators that are involved in the sexual development, further pointing towards the high plasticity of their sex determination systems. However, most scientific attention in regard to sex determination in fish has been focused on the GSD systems involving sex chromosomes (Heule et al., 2014; Fowler and Buonaccorsi, 2016). Sex chromosomes in fish come in two forms, either they are heteromorphic, and are distinguishable cytologically, or they are homomorphic, meaning they appear identical (Heule et al., 2014). Either way this results in two sexes, one heterogametic and one homogametic. The heterogametic sex produces two types of gametes since it has two different sex chromosomes. The homogametic sex on the other hand possesses two copies of the same sex chromosome and therefore produces only one type of gamete. In case the male is the heterogamic and the female the homogametic sex, this constellation is referred to as an XX-XY system. Conversely, if the female is the heterogamic and the male the homogametic sex, it is referred to as ZZ-ZW (Avise and Mank, 2009; Heule et al., 2014). Teleost fish show both

modes of heterogamety which can be observed in even closely related species (Heule et al., 2014; Fowler and Buonaccorsi, 2016). For a detailed review of the polygenetic mechanisms in regard to sex determination in different fish and their reproduction strategies, see Heule et al. (2014) and Avise and Mank (2009).

1.3 Sex determination loci

Unravelling the genes or genomic regions which are involved in sex determination can be complex process and often relies on intensive sequencing to eventually identify potential genetic marker. While the identification of sex chromosomes via karyotyping is possible in some species, this method is less feasible for the majority of fish species, since they possess homomorphic sex chromosomes (Devlin and Nagahama, 2002; Matsuda et al., 2002). Instead, researchers can use linkage studies to observe the inheritance patterns of potentially sex-linked traits to point out sex determining loci and narrow down potential specific regions (Devlin and Nagahama, 2002; Matsuda et al., 2002; Johnson, 2019). During genome assembling, researchers can apply recombination analysis on sequence data from related individuals of both sexes to determine how closely genetic markers are linked. This helps create genome maps, representing the position of sequences along the chromosome. If chromosomes with low recombination can be found, those are potential candidates for sex chromosomes (Matsuda et al., 2002; Johnson, 2019). In particular the development in the Next-Generation-Sequencing (NGS) techniques allows for cost-effective whole genome data collection from many individuals, enabling the identification of sex-specific single nucleotide polymorphisms (SNPs). By comparing these SNPs, researchers can pinpoint regions associated with sex determination. Identifying sex-specific heterozygosity, where heterozygous SNPs occur exclusively in males or females at certain loci, is crucial for detecting a GSD system. These techniques offer valuable insights into the genetic basis of sex determination in various species (Johnson, 2019; Tørresen et al., 2017). In fish, at least eight different master sex-determining (MSD) genes are known to date. A growing number of MSD that have been observed are variants of the amh gene (anti-Müllerian hormone) (Johnson, 2019; Myosho et al., 2012; Pan et al., 2021). In mammals, amh initialises the regression of Müllerian ducts in males during embryogenesis and is known to influence the sexual development in both males and females. Although fish lack Müllerian ducts, studies have shown that amh in fish inhibits gonadal germ cell proliferation and steroidogenesis (Johnson, 2019; Healy and Schulte, 2015). The last exon of the amh gene encodes the transforming growth factor beta (TGF-β), which constitutes a well-studied growth factor that can stimulate or inhibit cell proliferation. It has been shown that other

sex determining genes in fish also have connections to the TGF-β pathway, underscoring the significance of this signalling cascade and the involvement of amh in sex determination within fish (Johnson, 2019; Myosho et al., 2012; Pan et al., 2021). Recent advances of fish genetics and genomics showed that the plasticity derives from the dynamic genome of this clade and unravelled the importance of gene duplication events and the implications of turn-over processes within sex determination systems (Avise and Mank, 2009). The rapid turnover of sex chromosomes in teleost's allows researchers to study different stages of differentiation in sex chromosome pairs. Recent studies focused on the identification of new master sex determining (MSD) genes in fish gave further insides into the turnover of sex determination systems and the formation of sex chromosomes. These new MSD genes are thought to arise through either gene duplication and sub- or neo-functionalization or allelic diversification. Teleosts are the only clade where all of those mechanisms have been found so far, making them interesting study objects in terms of the evolution of sex determination systems (Johnson et al., 2020; Pan et al., 2021). Besides the numerous research opportunities, unravelling the reproductive biology, especially of species that are harvested extensively in the wild, is imperative to enable efficient and contemporary management strategies and to predict potential environmental or anthropogenic impacts (Devlin and Nagahama, 2002).

1.4 Sex determination in *Esox lucius*

The mentioned rapid evolution of sex determination mechanisms in teleost fish enables many opportunities to study the formation of sex chromosomes and the underlying mechanisms that lead to the development of new master sex determining (MSD) genes (Pan et al., 2019, 2021). In recent years particularly one species was in the focus of interest in terms of studies regarding sex determination. The characterization of a new MSD in the Northern Pike (*Esox lucius*) and the proceeding breakdown of the genetic variations between pike populations resulted in a by this time impressive state of knowledge in regard to this species' genome (Johnson, 2019; Nordahl et al., 2019; Pan et al., 2019; Johnson et al., 2020; Pan et al., 2021; Eschbach et al., 2021; Möller et al., 2021; Sunde et al., 2022). The Northern pike (*Esox lucius*) is a large and long-lived keystone predatory teleost species found in freshwater and brackish coastal water systems in North America and Eurasia. The species is confronted with human-induced disturbances in most areas of its native range. On the one hand this led to its constant decline over the last decades, on the other hand the anthropogenic influences can affect the local and regional genetic diversity with unpredictable outcomes, making conservation imperative form ecological and genetical point of view (Eschbach et al., 2021; Pan et al., 2021). In recent years

Northern Pike has further gained significance as a model organism in ecology and conservation due to its crucial role as a top predator, influencing the composition of local fish communities (Pan et al., 2019, 2021). Additionally, it holds economic importance as a valuable resource for food and sport fishing. As a result, recent efforts, especially by Pan et al. (2019, 2021) have focused on generating as much genomic information for *E. lucius* as possible, including a comprehensive whole genome assembly with chromosome anchoring and a tissue-specific transcriptomes. These genomic basis provides valuable tools for further research on this species' genetics, and will complement conservation strategies (Pan et al., 2019; Johnson et al., 2020; Pan et al., 2021). Pan et al. (2019) characterized a male-specific duplicate of the anti-Müllerian hormone (amh) gene, which they named amhby, as the Master Sex Determining (MSD) gene in the Northern Pike. Among other analysis they conducted RAD-Sequencing on a family panel, which enabled them to identify Linkage Group (LG) 24 as the sex chromosome and pinpointed the location of the sex locus in the sub-telomeric region. They further determined that amhby was located on the Y-specific sequence. They also examined the expression patterns of the amhby gene in various tissues of the Northern Pike. By analysing tissue-specific transcriptomes, they found that amhby exhibited a testis-specific expression pattern. This provided strong indications that amhby played a crucial role in male sex determination. Further phylogenetic and synteny analyses indicated that the duplication of the amh gene occurred around 40 million years ago, and that amhby was subsequently translocated to the Y chromosome. Their findings eventually confirmed that Northern Pike utilizes a XX-XY male heterozygous sex determination system, as previous studies suggested. Those findings enable accurate sex determination in Northern Pike which in turn offers numerous benefits for both scientific research and practical applications (Johnson, 2019; Pan et al., 2019, 2021).

1.5 Benefits and application of accurate sex determination in *Esox lucius*

By employing reliable sex determination techniques, researchers can gain valuable insights into the spatial and temporal population structure with higher resolution. Sex ratios within populations play a crucial role in population genetics (Fowler and Buonaccorsi, 2016; Catchen, 2017; Johnson, 2019). Incorporating sex-specific data in genetic studies ensures more accurate conclusions, avoiding potential biases caused by overlooking sex-related differences in the population. These ratios can be influenced by environmental factors and selective mortality, making them important markers for assessing the overall condition of a population (Kelly et al., 2014). Thus, studying sex determination in Northern Pike enables a deeper understanding of the species' population structure and aids its conservation and management. Traditional

morphological sex determination methods have proven to be unreliable, particularly during the spawning season and for juvenile individuals (Casselman, 1974, 1975; Johnson, 2019). The use of genetic approaches for sex identification enables non-invasive and precise assessments, unaffected by external factors (Fowler and Buonaccorsi, 2016; Catchen, 2017; Johnson, 2019).

1.6 A genotyping assay for Northern Pike from Rügen

Recent studies indicate that there are at least two sex determination mechanisms in Northern Pike (Johnson, 2019; Pan et al., 2019, 2021). The North American lineages, which lost amhby as their MSD, have been separated from the European lineages only since the last glaciation (Pan et al., 2021). Since these are relatively small timescales, evolutionary spoken, this development is often termed as a "recent event". Besides the apparent advantages of having a straight forward sex determination genotyping assay at hand, it is worth testing if diverging pike populations exposed to different pattern of isolation (geographic isolation (allopatry, isolation by distance (IBD), resistance (IBR), and environment (IBE)) still exhibit the presumed sex determination system (Johnson, 2019; Sunde et al., 2022). The differentiation of pike populations into several different cluster along a salinity gradient has been proven for Northern Pike populations of the Baltic Sea (Nordahl et al., 2019; Eschbach et al., 2021; Möller et al., 2021; Sunde et al., 2022). As part of a long-term survey focused on pike populations within the waterbodies around the German island Rügen, the genomic counterpart included the verification of the assumed SD system with amhby as MSD in the local populations. The Boddenhecht Project is an interdisciplinary survey focusing on Bodden waters around Rügen. Its goal is to gather reliable data on pike development and fishery science. The approach combines fishery science, fish biology, and socio-economic aspects to achieve sustainable outcomes (www.ifishman.de). The main goal of the genomic counterpart of this project was the genetic confirmation of the mentioned diverging *E. lucius* populations into different cluster along a salinity gradient, and the implementation of the findings into the assessment of the overall condition of the local pike populations to aid conservation and management strategies. Anyways, this work in particular aimed to validate the sex determination system of the *E. lucius* populations from numerous fresh and brackish waterbodies around Rügen. The main goal was to establish and apply a genotyping assay to study the sex ratios of those populations. For this purpose, a pool of confirmed male and female pike were sampled to test the genetic marker which has been bioinformatically designed earlier. As a result, two different assay approaches have been established and successfully confirmed the assumed sex determination system based on the MSD (amhby) and the autosomal counterpart (amh). On one hand, a PCR based assay

that utilized a length polymorphism in amh and amhby was tested for straight forward sex determination and has been established based on the pools of confirmed male and female individuals. On the other hand, the gained knowledge was used to implement a NGS marker into the genotyping marker panel, enabling an *in-silico* sexing approach that can be utilized for future studies.

1.7 Research motivation and hypothesis

Based on the habitats and isolation pattern observed in the Rügen pike populations, it is hypothesized that these lineages have not experienced such an extensive separation from the rest of the European lineages compared to the North American lineages. Therefore, it is expected that the known sex determination system, with amhby as MSD, is still in place for the Rügen population cluster. If there was any substantial evidence for alternative sex determination in the Rügen pike populations, it can indicate a more pronounced and earlier differentiation of this population cluster than anticipated, and indicate an alternative SD system, potentially with another MSD (Möller et al., 2021; Sunde et al., 2022). Further, it is assumed that the establishment of a genotyping assay for sex determination using SNP's and indels in amh and amhby is feasible, if these genes are present and if amhby remains the MSD. It is assumed that amh can be detected in both males and females, while only males will exhibit amhby. Given the assumption of two copies of amh – the autosomal one on LG08 and the duplication containing the Y-specific amhby on LG24 – the expected coverage ratio of amh to amhby is around 3:1 (Pan et al., 2019). Utilizing the gained knowledge in regard to the actual proportion of males in females within the populations, estimations in regard to the sex ratios can be inferred. Observations of North American Northern Pike populations have shown both male and female-biased sex ratios (Huffman et al., 2014; Johnson, 2019). Although pike sex ratios are considered to be skewed, the availability of pike shows clear seasonal trends that are related to sex specific activity pattern and are independent of locality and method of capture (Casselman, 1975). Casselman, (1975) observed similar seasonal pattern of availability in regard to the total number of males and females sampled in fresh and brackish water locations. He reported that the sexes show biannual peaks of availability. The observed annual behavioural differences indicated by the frequency of occurrences (catches), can be aligned with the supposed annual pattern regarding migration and behavioural characteristics related to growth and spawning, observed in other studies. This analysis can be applied for the pike populations from Rügen. The combination of an accurate sex determination marker, and the corrected inferred depiction of the sex ratios within the sampled pool can potentially portray the seasonal

availability pattern of pike more reliable. Most importantly, the analysis of the sex ratios might indicate a sex ratio shift towards female dominance as referred to by numerous studies regarding pike populations (Han et al., 2018; Johnson 2019; Johnson et al., 2020; Kelly et al., 2014; Leopold, Mark & Farrell, 2012). Laboratory experiments in North America to rear Northern Pike resulted in a highly female-biased sex ratios. These findings implicate that environmental and potentially anthropogenic factors play a significant role in determining the sex of Northern Pike, since a sex determination mechanism that purely relies on genetic factors would result in a ratio close to 1:1 (Johnson, 2019). Also, Leopold and Farrell (2012) compared sex ratios of Northern Pike between the St. Lawrence River and inland New York waterbodies. Historically, in the St. Lawrence River, sex ratios of Northern Pike have been nearly 1:1, but in recent decades, sex ratios have skewed heavily towards females. Previous studies have yet failed to determine the geographic extent of the skewed sex ratios and the potential implications of female dominated populations on the reproductive health of local populations (Leopold and Farrell, 2012). Under consideration of the expected annual activity pattern, it is expected that the sex ratios except in the spawning season show a female bias, since their increased activity throughout the year makes them more susceptible for most catching methods (Casselman, 1975, Johnson 2019). If this is true, a significant deviation from a 1:1 ratio towards females during the spawning season, if to be proven, can be consulted to indicate a female bias and a sex ratio shift towards female dominance. The potential implications of such a bias, gains further importance under consideration of the possible correlations noted by Johnson (2019). He referred to several examples of species that have shown significant sex ratio shifts towards female dominance in reaction to environmental changes (e.g., turtles, and other reptiles). Considering the effects of the climate change on temperature and other water parameters, the knowledge of such a correlation would be crucial for conservation management. Briefly, an accurate depiction of sex ratios is essential for ecological evaluations, hence management approaches and conservation.

1.8 Summary of research questions

Can the existing knowledge regarding sex determination and the involved genes in *Esox lucius*, applied to the Rügen populations, be used to establish a straightforward, and accurate genotyping assay based on genetic marker?

Do the Northern Pike populations from Rügen show amh and amhby with the presumed sequence structures and in presumed quantities in males and females? Is amhby still the MSD in these populations? Is there any evidence for alternative or atypical sex determination?

Can this knowledge and the resulting marker provide a significant contributions to increase the resolution of ecological and population genetic studies regarding pike populations from Rügen?

2. Methods

2.1 Study area and sampling

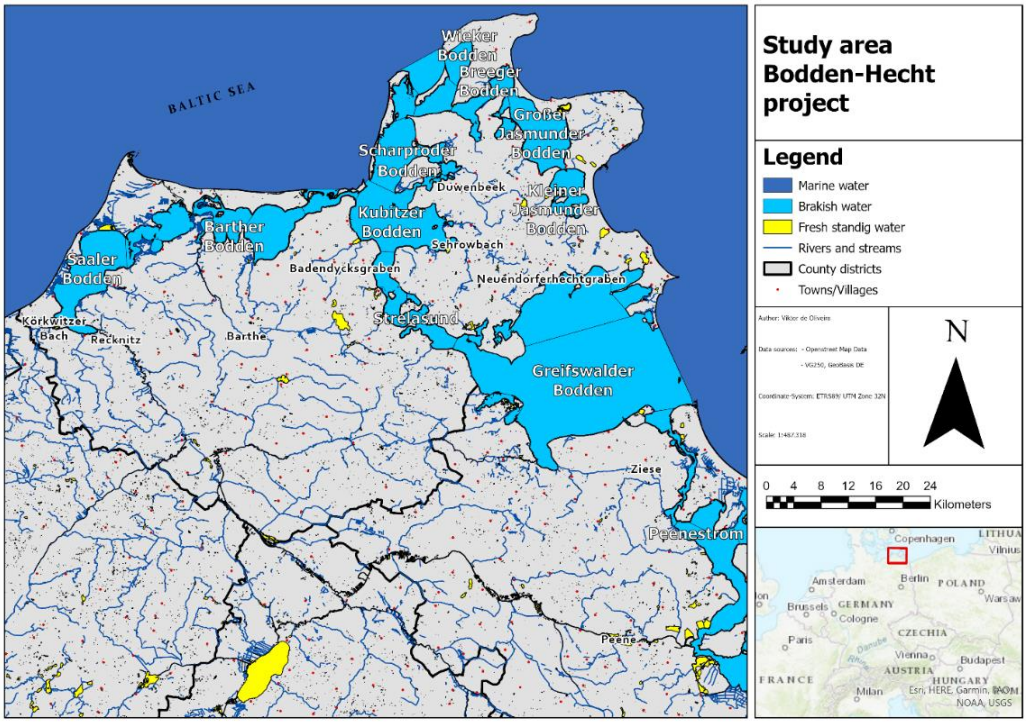


Figure 1: Study area around Rügen. Showing the different water bodies (color coded) and landscapes (© Viktor de Oliveira)

The *E. lucius* samples used in this study originate from numerous water bodies around the German island Rügen. Rügen, Germanys largest island, is located off the Pomeranian coast in the Baltic Sea and is part of the Vorpommern-Rügen district in the state of Mecklenburg-Vorpommern (Figure 1). The Baltic Sea is semi-enclosed postglacial sea with a surface of around 415,000 km² which is characterized by a keen salinity gradient from marine salinity (35 Practical Salinity Units, PSU) within the area of the entrance in the west to almost freshwater (< 2 PSU) in most parts northeast (Arlinghaus et al., 2023). The island is known for its diverse and beautiful landscapes, and open bays, known as Bodden (Frenzel and Oertel, 2002). The Rügen area is shows diverse coastal landscapes, including steep coasts, flat hilly bays, sandy beaches, and vast reed areas along the Bodden. The Bodden are shallow, nutrient-rich brackish coastal waters separated from the sea by islands or tongues land. The water level in the Bodden is influenced by prevailing wind and current conditions, affecting the exchange of water between the Baltic Sea and the Bodden waters, and thus salinity that oscillates around 8-9 PSU. The Bodden waters are in average around 8 meters deep and partly rich an in macrophytes that

offer ideal conditions as habitats for many fish species and serve as spawning grounds. (Frenzel and Oertel, 2002; Nordahl et al., 2019; Möller et al., 2021). The productive Bodden ecosystems are especially known for their pike stocks, where individuals can grow to remarkable sizes. This, and the presence of other popular saltwater-tolerant freshwater predatory fish species like the pikeperch and perch, makes the Bodden area around Rügen attractive for local anglers, fishing tourists, and commercial fisheries. The sampling of the Individuals for the establishment of the sex marker was conducted via electro fishing from a boat during the spawning season in April 2022. The aim was to sample for a decent pool of male and female individuals with confirmed sex status. The sex was confirmed by either dissection and internal gonad examination or by in situ gonad inspection, whereas the ripe status (running eggs or running ripe) was taken as sufficient for the sex confirmation.

2.2 Marker design

Based on the rich dataset from Pan et al. (2019,2021) there was an attempt in our lab to utilize the provided sequences to implement a sexing marker for the populations from Rügen. The first attempts resulted in suboptimal results for the PCR assay and gel electrophoresis. Consequently, it was decided to reconstruct the sequences for amh and amhby based on the present dataset in form of PoolSeq reads from two populations from Rügen. A conserved region on LG 08 from the Pan et al. 2019 dataset served as the foundational seed for the search algorithm. The next phase was implemented within the UNIX environment of the High-Performance-Computer (HPC) cluster of the Carl von Ossietzky Universität, Oldenburg. It involved employing scripts, to search for the part of the conserved sequence within the present reads (Appendix 1. Fig. 1) These commands effectively identified all the reads containing the searched sequence, giving a list of potential matches. After the identification of these matches, the next step involved using the Basic Local Alignment Search Tool (BLAST) of the National Centre for Biotechnology Information (NCBI) to analyse the matches for their consensus with the gene of interest. Once a match was identified, the succeeding process involved the selection of a part of the sequence adjacent to it, either to the left or right, depending on the desired direction along the gene. This part was then used again as the starting point for the search algorithm. This entire process was executed on older PoolSeq reads from the Boddenhecht Project: The Sehrowbach pool, which exhibited a slight male bias, and the Grosser Jasmunder Bodden pool, with a slight female bias. This iterative process led to the successful reconstruction of the two genes, as present in the populations from Rügen and in accord with the assembles from NCBI, with reasonable certainty. Post reconstruction an alignment was compiled based on the consensus sequences of

amh and amhby. The alignment was performed with the ClustalW tool from UGENE. The underlying objective of the alignment was to pinpoint conserved regions, SNP`s, and length polymorphisms that can be utilized for the strategic placement of primers for the PCR and NGS assay. The chosen primer pair was designed by identifying conserved regions in proximity to areas with insertions or deletions (indels). Owing to the primers' location within conserved regions, a solitary primer pair can simultaneously target the sequence of interest in both genes. A length polymorphism can then be visualized via gel electrophoresis and will reveal a distinct banding pattern. This offers the advantage of visualizing the presence of amh and amhby in one individual, which points up the sex. For the PCR assay two primer pairs were designed and tested. An additional marker was chosen for the NGS assay. The resulting amplicons of amhby and amh differed in length with about 30 bp difference, allowing for visualization in the gel. The result is a banding pattern with two bands for a male (that shows amh and amhby) and one band for a female (showing only amh). For the entire reconstructed sequences, parts of the alignments, the primer regions and primer parameter see Appendix 2. Fig. 2a-d, and Table1.

2.3 Laboratory procedures

Most protocols used for this study were slightly adapted, apart from the producer's protocol and are summarized in condensed form in chapter 2.3.1- 2.4.7. For a detailed description of all laboratory procedures, including all steps, reagents, and further explanations as well as all figures, Tables and pictures see Appendix 3.1 - 4.7. The methods established in this study continue at chapter 2.5.

2.3.1 DNA Extraction

The DNA extraction protocol involved adding dried tissue samples (fin clips stored in ethanol) to a microcentrifuge tube with Homogenization (HOM) Buffer and Proteinase K. After incubation and cooling, RNAse A was added to eliminate RNA contaminants. A Sodium-Chloride solution was added to precipitate proteins and debris. The supernatant, containing DNA, was collected, and mixed with ethanol to precipitate DNA. After centrifugation, the DNA pellet was washed with ethanol twice and air-dried. The DNA was then resuspended in TE Buffer. The DNA extraction was performed on 112 Esox lucius fin clips (Appendix 5, Table 7). The extracted DNA was quantified using a BioSpectrometer for concentration and purity assessment (260/280 ratio). The DNA samples were eventually diluted to a working concentration between 20-100 ng/µl for further analysis.

2.3.2 Polymerase Chain Reaction (PCR)

After DNA extraction and dilution, the DNA was amplified using the Polymerase Chain Reaction (PCR). The QIAGEN Multiplex PCR kit with HotStarTaq DNA Polymerase, MgCl2, dNTPs PCR buffer was used. Q-Solution was added to enhance amplification and prevent potential issues with secondary structures or GC-rich templates (only for primer pair A). The PCR cycle for primer pair A consisted of six steps: Initial denaturation at 95.0 °C for 15 minutes, followed by denaturation at 94.0 °C for 30 seconds, annealing at 62.0 °C for 1 minute and 30 seconds, extension at 72.0 °C for 30 seconds, final extension at 72.0 °C for 10 minutes, and a final hold at 10.0 °C. These steps are repeated 25 times to amplify the DNA target sequences. To determine the ideal annealing temperature of the primers, a gradient PCR was initially performed at different annealing temperatures. The optimal annealing temperature from the gradient PCR was then used to establish a cycling protocol for subsequent PCR reactions. For master mixes and cycling conditions for all PCR's and all primer pairs see (Appendix 3.2, Table 2a-f)

2.3.3 Gel electrophoresis

The Gel of choice was a 3% Low Electroendoosmotic (LE) Agarose Gel with Tris-acetate-EDTA (TAE) buffer. Agarose and TAE buffer were combined in a flask with some water added to compensate the evaporation. The mixture was heated in a microwave until the agarose dissolved. After cooling, it was poured into a casting tray with combs, left to solidify for 20 minutes. Master mixes were prepared for loading the gel, and the gel was run at a constant 125V for 60 minutes for primer pair A (70 for primer pair B). The electric field, which caused the negatively charged DNA fragments to move towards the positive electrode, resulted in distinct bands corresponding to different fragment sizes. Since smaller fragments moved faster than larger ones, this eventually separated the fragments by size. A 100bp equimolar DNA ladder was used as a standard. The visualization of the PCR products via this gel electrophoresis for all samples can be seen in Appendix 3.3 Figure 3a-i.

2.4 Library Preparation

2.4.1 DNA extraction Dneasy 96 Blood & Tissue Kit

For the majority of the pike samples from the Boddenhecht Project, DNA extraction was done using the DNeasy 96 Blood & Tissue Kit. Approximately 20 mg of fin clip tissue was added to each well of a collection microtube plate. A master mix of Proteinase K and ATL Buffer was

added, and the plate was sealed, briefly centrifuged, and incubated at 56°C until the solution cleared. Puffer AL-Ethanol Mix was added, shaken, and centrifuged. The solution was transferred to the DNeasy 96 plate, sealed, and centrifuged. Buffer AW1 and Buffer AW2 were added in separate steps, respectively after further rounds of incubation and centrifugation. Then the DNeasy 96 plate was placed onto the Elution Microtubes RS. Buffer AE was added, incubated, and centrifuged. After quantification, the DNA samples were diluted and prepared for downstream analysis.

2.4.2 PCR 1

The PCR1 protocol began with the preparation of a primer mix, which contains each primer at a concentration of 2µM. The volume required for 72 reactions is 28µL for each primer, leading to a total volume of 2016µL, with no requirement for the addition of molecular grade water. Cycles and master mix preparation shown in 4.2 Table 4.

2.4.3 DNA purification – ExoSAP-IT and Beads clean-up

Following PCR1, the amplified DNA samples underwent an enzymatic clean-up using Exonuclease-Shrimp Alkaline Phosphatase (ExoSAP-IT). Approximately 2µL of ExoSAP-IT reagent was added per 10µL of PCR1 product. After incubating at 37°C for 15 minutes, and then at 80°C for 15 minutes, the ExoSAP-IT to allow the degradation of leftover primers and nucleotides. Next, the PCR products were purified using AMPure XP-Beads. After acclimating the beads to room temperature, a 1.6x volume of beads solution was added to each well. After incubation and magnetic separation, 5µL of supernatant was retained, and the rest of the solution was discarded. The beads were washed with 70% ethanol twice. After evaporation, 40µL of molecular grade water was added to each well, followed by vortexing and centrifugation. The eluate, approximately 36µL, was transferred to a new 96-well plate.

2.4.4 Quality check via Phor Agarose Gel Electrophoresis

After each PCR and subsequent clean-up, a quality check was performed using 4% Phor agarose gel electrophoresis to assess the DNA samples' quality and quantity. The process is similar to the described 3% agarose gel. To set up a 4% Phor-agarose gel, TAE buffer cooled to 4°C is used. Agarose is added to the buffer while stirring at approximately 300 rpm. The mixture is incubated at room temperature for 10 minutes, with added water to account for evaporation. The solution is heated in a microwave until agarose dissolves (regular weighing prevents excessive evaporation). After heating, the gel is briefly cooled, poured into a tray with combs, and cooled at room temperature for 10 minutes before refrigeration for 30 minutes. The gel is

then ready for electrophoresis, run at 167.5 V for 70 minutes. PCR products are loaded into wells, and an electric field separates DNA fragments by size. A 100bp equimolar DNA ladder is used as a size reference. After the run, the gel is visualized under UV light, and the DNA fragments' positions relative to the ladder fragments determine their sizes. These quality checks ensure the DNA fragments' desired size and integrity throughout the analysis.

2.4.5 PCR 2 – Indexing

Before indexing, the total volume of the multiplex PCR product is approximately 40 µL, which is further subjected to a 1:10 dilution. This dilution results in a total volume of 40 µL per well, providing sufficient material for approximately 19 runs of the PCR2. The indexing process is necessary for the identification of DNA samples after sequencing. Unique barcode sequences are integrated into the DNA samples during PCR2. These barcodes, added to each sample, act as tags that remain attached to the DNA fragments through the sequencing process. In downstream analyses, the barcodes enable the identification of the origin of the sequenced DNA, allowing for the differentiation of sequences that belong to different samples. The used indexing primer and the preparation of the master mixes is shown in Appendix 4.5 Table 5a-b.

2.4.6 Pooling – Final Beads clean-up and size selection (~200bp)

For the pooling, 3µL from every well across all plates were brought together in a 12-well strip using a 12-channel pipette. The content in the strip was mixed by pipetting and then transferred into a 2ml tube. The strips were briefly centrifuged to ensure full transfer. The pooled samples were vigorously vortexed for 2 minutes and incubated on a rotator for one hour. After homogenization, two aliquots of 100µL each were prepared for final purification. AMPure XP bead solution was added to each aliquot in quantities of 35µL and 55µL, respectively (0.35x and 0.55x of initial sample volume). After incubating for 5 minutes at room temperature, the samples were placed on an Agencourt SPRIPlate Super Magnet Rack for 2 minutes to separate the beads from the solution. The supernatant was carefully aspirated and transferred to a fresh tube, while the beads were discarded. The beads were resuspended and 0.55 times the initial sample volume was added to each sample. After a 5-minute incubation, the plate was placed on the magnet to separate the beads from the solution. The supernatant was discarded, leaving 5µL of supernatant with the beads. Ethanol washing was performed twice, and the plate was dried until all the ethanol evaporated. The tube was removed from the magnetic stand, and either 10mM TRIS with 0.1% Tween-20 or Qiagen EB with 0.1% Tween-20 buffer (100µL) was added to each tube. After incubation and pipetting, the 96-well plate was placed on the magnetic stand for one minute to separate the beads from the DNA solution. Finally, the ~100µL eluate

was transferred to a new tube, completing the bead clean-up and size selection process. The Gel of the library of which the 112 Individuals used for the PCR assay were part of can be seen in Appendix 4.6, Figure 4.

2.4.7 Sequencing - Quality check - Trimming

The DNA samples were then sequenced on an Illumina MiSeq platform with paired-end reads of 2x151 base pairs using the MiSeq Micro Kit V2 (300 Cycles). Raw sequencing data were accessed via MobaXterm on the CARL/EDDY cluster of the Carl von Ossietzky University in Oldenburg and processed using Linux with SLURM as the job scheduling system. The reads were pre-processed using Trimmomatic with the following parameters: ILLUMINACLIP: 2:30:10 (adapter removal) MINLEN: 50 (discarding reads below this length) MAXINFO: 40:0.2 (sliding window trimming based on quality scores). Quality control analysis was conducted before and after trimming using FastQC. The processed reads were then aligned to a reference based on the reconstructed amh and amhby sequences using NextGenMap, a module designed for handling highly polymorphic genomes. The alignment generated a Sequence Alignment Map (SAM-file) containing information about the aligned reads' positions and sequences in relation to the reference genome (Bolger et al., 2014; Sedlazeck et al., 2013). This information was used for the bioinformatical sexing approach. Only the forward reads were used for further analysis.

2.5 *In silico* sex determination

Post sequencing and trimming, the reads were prepared for processing in R by applying straight forward bash pipelines. To obtain the amh and amhby counts based on the SAM files a pipeline involving regular expressions including "grep" and "awk" commands was applied to extract reads mapping to two loci of interest. The "wc -l" command was then used to count the number of reads for each locus, providing the necessary information to get the counts for each locus of interest and to compute the amh:amhby ratio. The in-silico sexing approach was performed in R from there on and started by incorporating a new variable into the dataset - the amh:amhby ratio, which is simply the ratio of amh to amhby counts. In pursuit of genotyping assay for sex determination, the amh:amhby ratio was used to define appropriate thresholds. Based on the expected quantities of amh and amhby within the pikes' genome, a ratio somewhere close to 3 was expected. The ratio range indicative of male individuals derived from the observed minimum and maximum amh:amhby ratios in a pool of individuals with confirmed sex. The calculation of the minimum, maximum and mean ratio as well as the visualisation of different pools of individuals in regard to the coverage helped to narrow down an appropriate threshold

for the genotyping assay. Initially, the ratios were skewed and showed a wide, biologically unsuggestive range. The investigation of different ratio ranges within different coverage pools, especially the exclusion of reads with a coverage <10, resulted in a biologically more suggestive range, and clearly pinpointed out the presence of outliers. This was specifically a small subset of individuals, characterized by sufficient amh coverage and exactly, and uniformly a single count of amhby. Although the distribution of the amh and amhby counts and the ratios made it hard to statistically pin down this phenomenon, their extreme deviation from the mean ratio range and the biologically unsuggestive uniformity of these outliers can be visualized and is shown in Figure 2 (for further explanations regarding the thresholds, and how they were chosen see Appendix 5, Figure 5a)

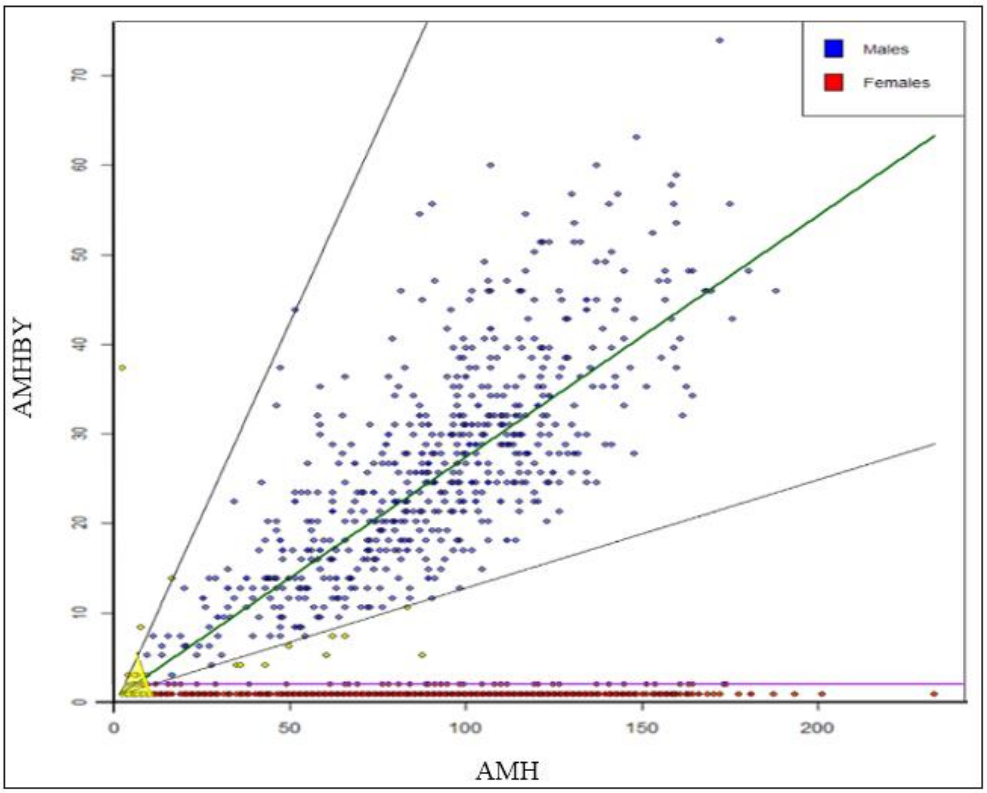


Figure 2: Shows the amh:amhby ratio for all individuals with read data for the target genes in form of a scatter plot. Amhby is on the Y-axis and amh on the X-axis. Each point represents the respective amh:amhby ratio of on individual. The blue dots represent males, the red dots females, and the yellow dots represent individuals that did not pass the genotyping threshold. The red data cloud parallel to the x-axis represents the females and the individuals above, traversed by the purple line, are the outliers (amhby =1), which were also assigned to the females. The continuous dark-green line indicates the mean ratio, the black lines indicate the ratio range in which individuals are assigned to males. Individuals within the area of the yellow triangle close to the origin were also not considered for the genotyping assay and were filtered out. These individuals not only show low read coverage but the diffuse conglomeration close to the origin due to their ratio pattern made the differentiation between male and female unreliable. Again, the ratio range which eventually served as the threshold-range derives from the max and min ratio from a filtered dataset, which was then applied to the unfiltered dataset. Thus, the black lines in this plot represent the actual threshold for the genotyping assay for sex determination.

Figure 2 shows a scatter plot of the amh:amhby ratio of all individuals for which the sequencing and pipelining approach worked, and includes the established thresholds indicated by the two black lines. Further the visualisation was used to implement an extension of the threshold. Individuals that fall into the area highlighted by the yellow triangle were also excluded from the genotyping assay. These individuals not only show low read coverage but the diffuse conglomeration close to the origin due to their ratio pattern made the visual differentiation between male and female unreliable. Based on the defined thresholds, the blue coloured individuals within the min/max range and outside of the yellow triangle were considered as males. All individuals that that fall below the purple line (traversing the outlier individuals) and outside of the yellow triangle were considered as females, including the purple-coloured outlier. For a snipped of the BASH pipelines and R scripts see Appendix 5, Figure 5b.

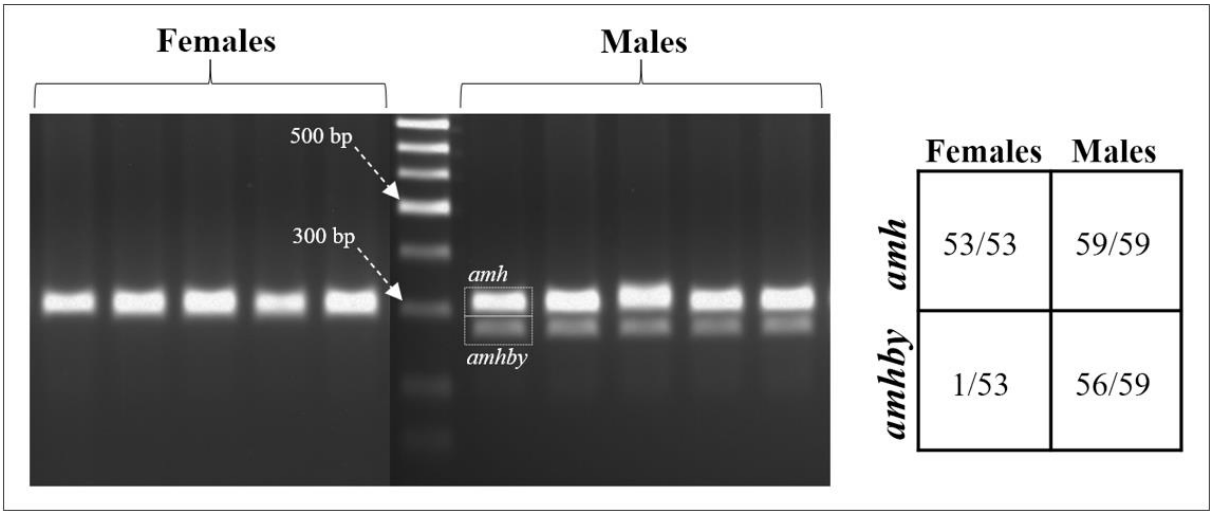
2.6 Statistical approach of the sex ratio assessment

Since this analysis yield in a more reliable determination of the total amount of males and females in the sampled populations, it was decided to examine the corrected sex ratios as well. An analysis of the latter was conducted in different water bodies, clusters and throughout the year by applying Chi-squared statistics to resolve the seasonal and spatial pattern of the sex ratios and infer any coherence with expected behavioural pattern and to check for a ratio shift towards female dominance. The test was used to determine whether the observed sex ratios (female to male) deviate significantly from a 1:1 ratio, on pursuit of any non-random, significant trends or patterns in the sex ratios across different seasons and waterbodies and genotype cluster. In the context of this analysis, a low p-value (< 0.05) suggests that the observed ratios are significantly different from a 1:1 ratio and are therefore not due to random chance. Meanwhile, a high p-value suggests that any observed deviation from the 1:1 ratio could simply be due to random chance, rather than any meaningful pattern or trend. This was applied for each month and season, as well as the different waterbodies, while all fresh water and all brackish water locations were combined respectively for this part of the analysis. Contingency tables were created for different subsets of the available samples, in regard to different times over the year (each month, and season), different waterbodies (brackish and fresh) and the different cluster (purple, red, blue, green, red-blue, green-purple). The cluster represent the genetically pinned down populations that constitute differentiated sub population showing different habitat preferences and life history characteristics. Briefly, they can be differentiated into brackish water residents, freshwater residents, migratory ecotypes as well as some mixed types. To get a more comprehensive picture in regard of the sex ratios, these genotyped fish were consulted

as well. Here, the fish are essentially grouped into 6 cluster, allowing for a differentiated analysis of the sex ratios within these cluster. The cluster consists of two different brackish water cluster (blue and red) one freshwater cluster (green), one anadromous cluster (purple) and two mixed types (redblue, greenpurple). Apart from simply assigning a fish to "fresh" or "brackish", based on the characteristic of the waterbody in which it has been caught, the cluster allow for a differentiation within genetically defined and distinguishable populations.

3. Results

3.1 PCR assay for sex determination



	Females	Males
amh	53/53	59/59
amhby	1/53	56/59

Figure 3: Shows the banding pattern of the amh and amhby amplicons in a 3% agarose gel. Exemplarily five females (left) and five males (right) are shown. For the gels of the 112 confirmed individuals see Appendix 3.3, Figure 3a-i. The table on the right shows the number of individuals from the male and the females pools that showed bands for amh and amhby respectively.

The PCR and gel electrophoresis protocol were used as described above to amplify and visualize the target sequences in 112 individuals with confirmed sex. Since the primer pair A yield in clearer bands, the majority of the samples were amplified using these primers and the corresponding protocol. Figure 3 exemplarily shows the resulting band pattern for five female and five male individuals of confirmed sex. The PCR and subsequent gel electrophoresis resulted in amplicons with the expected length of ~296 bp for amh and ~263 bp for amhby. In total, all individuals (53 females and 59 males) showed bands for amh. The second band, indicating amhby was present in one female and in 56/59 males. Notable here were the results for one confirmed female and three confirmed males in particular, which did not show the expected banding pattern. A full list of all gels and the corresponding individual ID is shown in Appendix 3.3 Figure 3a-i.

3.2 Genetic vs external sex determination

Upon genotyping, the *in-silico* sex determination approach was implemented as described in Methods 2.5 and used to compare the individuals for which the information for both, external sex and genetic sex was at hand. The comparison revealed a discrepancy in the external sex determination in approximately 9,2 % of the individuals. More specifically, there was an overrepresentation of males and an underrepresentation of females in regard to the total counts for each sex based on external sexing. Of in total 1.218 (608 males and 610 females - externally sexed) 15,2 % of the males, and 3,1 % of the females were incorrectly sexed externally. After correcting the misidentified individuals, the actual sex distribution in the sample showed ~12% less males (74 of the externally sexed 608 males) and ~11 % more females (74 of genetic sexed 684 females) compared to the results of the external sexing. The majority of the incorrect sexing occurred during spring and summer, where ~ 70 % of the incorrectly sexed individuals were sampled.

3.3 Results of the sex ratio analysis

3.3.1 Sex ratios within all waterbodies over the year

Utilizing the corrected female and male counts based on the genetic sex determination, there was an attempt to test for non-random annual sex ratio pattern based on the catches of males and females over the year. The figures in this section show the caught male and female individuals in each month in form of bar plots. The statistical part and the ratios themselves are not directly visualized here. Figure 4 shows the catches of males and females in all waterbodies for each month. Biannual peaks of activity indicated by the number of catches can be observed. Particularly during spring and autumn the total number of catches increased, while most months where slightly female dominated. To test if the deviations from a theoretic 1:1 ratio were significant Chi-squared statistics was used. The Pearson's Chi-squared test indicated significant deviations from 1:1 sex ratio in specific months, namely January, June, July, October, November, and December, as their p-values are below the 0.05 significance level. However, during February, March, April, and September, the p-values are greater than 0.05, suggesting the observed deviation is not statistically significant (Appendix 6, Table 6a). The data available for May and August was not sufficient to conduct any reliable statistical test.

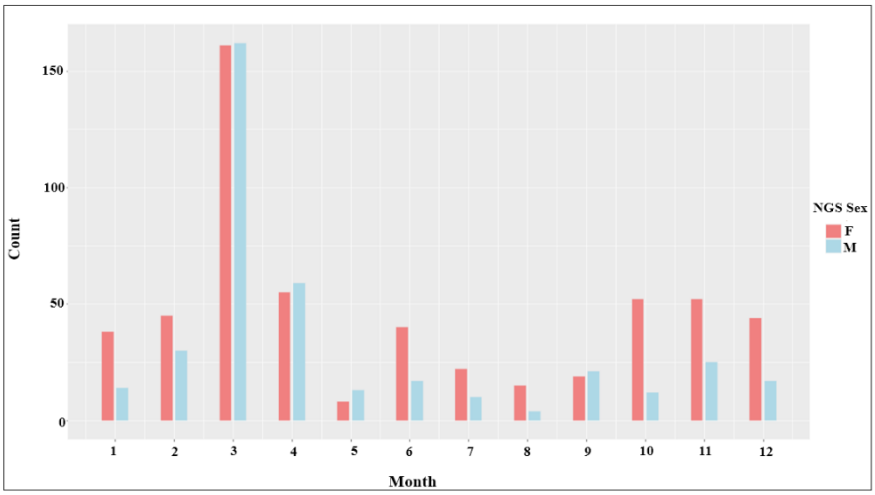


Figure 4: Shows the counts of males and females over the year for all water bodies. Counts are on the Y-axis and the month is on the X-axis. Females are color coded red, and males blue.

3.3.2 Sex ratio within the brackish and freshwater populations over the year

The populations, differentiated into freshwater and brackish water environments, as shown in Figure 5 show as well annual pattern in terms of the number of caught males and females. Again, biannual peaks of activity indicated by the number of catches can be observed, especially in the brackish water populations. Specifically, during spring and autumn the total number of catches increased, whereas more females were caught in most months, except for April. The pool of individuals assigned to the freshwater populations is generally smaller, which is reflected in the figure and makes the annual pattern less clear. Within the brackish water populations, the sex ratio significantly deviates from 1:1 ($p < 0.005$) towards female dominance throughout the year, with an average sex ratio of 1.4. All p-values for all seasons except spring fell well below 0.005 (Appendix 6, Table 6b)

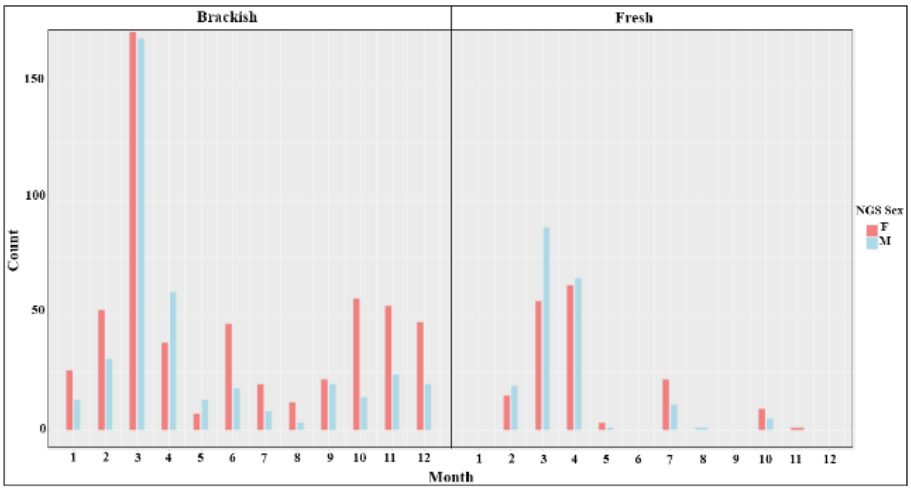


Figure 5: Shows the counts of males and females over the year for brackish and freshwater populations. Counts are on the Y-axis and the month is on the X-axis. Females are color coded red, and males blue.

3.3.3 Sex ratios within the genotype cluster over the year

Figure 6 shows the results for two cluster exemplary .The “blue”, and “redblue” clusters, both linked to brackish water populations show a significant deviation from 1:1 towards females ($p < 0.05$) for all seasons except spring. Spring, however, exhibits significant deviation towards males. The 'Red', 'Green', 'Purple', and 'Greenpurple' clusters were not considered due to insufficient representation in the dataset of this study and where not considered here. For the results of the brackish water associated cluster see Appendix 6, Table 6c. (For the plots for all cluster see Appendix 6, Figure 6)

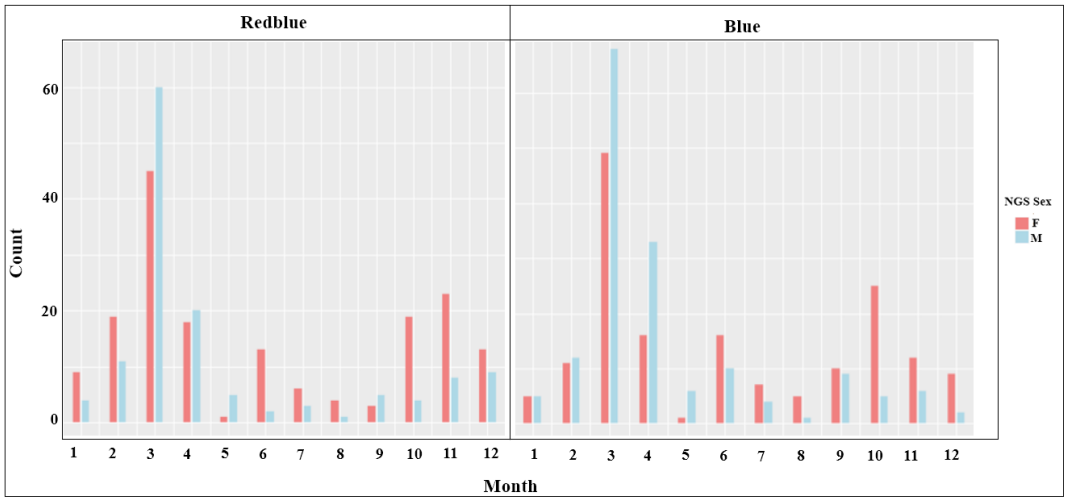


Figure 6: Shows the counts of males and females over the year for the redblue and the blue Cluster. Counts are on the Y-axis and the month is on the X-axis. Females are color coded red, and males blue.

4. Discussion

4.1 PCR and gel electrophoresis assay for sex determination

The results of the PCR-based assay to amplify the target sequences in the Northern Pike individuals showed a high success rate, indicating the reliability of the genetic sex determination based on amh and amhby fragments. The presence of these genes in the studied populations supports the hypothesis that the known sex determination system is still in place for the Rügen population cluster. The primers used for both amh and amhby were designed based on the available genomic data and bioinformatic analyses. The primer pairs were optimized to have appropriate melting temperatures (Tm) and GC content to enhance PCR specificity and efficiency (Appendix 2, Table 1). In this study, primer pair A demonstrated better results, most likely due to the highly conserved region in which the forward primer is located in, as well as the overall shorter target sequences. Shorter target sequences offer advantages over longer ones, since they require less PCR cycles for amplification, reducing the

risk of non-specific amplification and unwanted products. Shorter targets are generally more efficiently amplified, migrate faster in the gel and result in more clear and well-defined bands in gel electrophoresis (Sreedharan et al., 2018). The PCR and subsequent gel electrophoresis resulted in amplicons that showed the expected length of ~296 bp for amh and ~263 bp for amhby. In total, all individuals (53 females and 59 males) showed the desired band for amh, while the second band, indicating amhby was present in one female and in 56/59 males. Several repeats with the outlier individuals based on fin clips from our lab and the samples from Berlin as control, led to the same results. The overall success of the PCR assay is in accord with the findings of Pan et al. (2021). In their PCR assay, the amplification of amhby worked sufficiently for ~ 97% of the males, while it worked for ~ 95% in this study. While exact cause of those outlier males, that do not show the second (amhby) band is still unclear, this phenomenon can most likely be attributed to amplification issues during the PCR reactions. In total three repeats were conducted with all outlier individuals (Appendix 3g-i). Since the amount of DNA and all other basic parameters for successful amplification were given, a conclusive explanation is yet to come. Besides this, particularly one morphological female that showed amhby after the sequencing is out of line. Although alternative sex determination is known to exist in extremely small proportions of fish populations (Kallman, 1983; Schultheis et al., 2009), this has not been observed in Northern pike before. However, the sampling process in the field, the subsequent transition of the fin clips into the sampling tubes and the transportation is prone to human error. The chain of causalities leading to the male genotype in an alleged female is full of uncertainties, making any further discussion of this outlier female to pure speculation. Especially since there are no additional tissue samples of this individual that can be analysed and genotyped.

4.2 NGS *in silico* sex determination

The utilized knowledge based on the reconstructed sequences and the alignments of amh and amhby, and the subsequent primer design enabled the successful sequencing of the majority of the samples, delivering a sufficient coverage for the two genes of interest. One notable observation in this study is the underrepresentation of amhby compared to amh in terms of the coverage. Based on the assumption that amh is present three times within the genome (two times on the autosomal copy on LG08 and once on LG24), while amhby is present just once on LG24, a higher proportion of amh (~3:1) was expected. However, the bioinformatical analyses showed that the average amh:amhby ratio is around 4.5, while it is much higher in numerous individuals that have passed the genotyping thresholds. The GC content of both target

sequences was approximately 50%, which falls within the moderate GC content range that is often associated with decent PCR efficiency and sequencing performance (Chen et al., 2013; Hershberg, 2016). The impact of GC content on PCR efficiency is well-documented in the literature (Chen et al., 2013; Hershberg, 2016). DNA sequences with balanced GC content are more likely to form stable secondary structures, facilitating efficient primer annealing and amplification during PCR. Conversely, regions with extremely high or low GC content may inhibit primer binding, resulting in reduced amplification efficiency. Anyways, considering the moderate TM and even GC content of both target sequences, the underrepresentation of amhby compared to amh is unlikely to be attributed to the GC content and subsequent amplification or sequencing disparities (Chen et al., 2013; Hershberg, 2016). Primer design is another critical factor, potentially influencing PCR efficiency and sequencing coverage. As mentioned before, the primer characteristics are sufficient, thus their characteristics should not affect the amplification and subsequent sequencing to this extent. The primer binding specificity is the same for both targets since the primer were designed to bind on the same conserved sequences in both genes. Suboptimal amplification and sequencing conditions, attributable to the characteristic of the primer or target sequences could explain the discrepancy in the expected amh:amhby ratios. Anyways, considering the given characteristics of the primer and their target sequences, the amplification should theoretically work as expected. There are several possibilities that might explain this phenomenon. One possibility is that the assumption that has been made regarding the presumed amh:amhby ratio has been misguided. Another option is that the chosen sequences for amhby are not as conserved as expected and therefore not as suitable as a male marker as desired. Together with this, the references might be less specific as desired, potentially leading to wrong calls and skewed ratios. However, the sequences have been constantly compared to different assemblies on NCBI during reconstruction, and primer were chosen in regions where they perfectly matched the NCBI versions, underlining the accuracy of the reconstructed sequences and the chosen primer. Finally, bold hypothesis can be made in terms of potential amh duplications, further skewing up the ratios. However, this can neither be confirmed by the gained data in this study, nor is this indicated by any mentions in the literature in this particular context. It is notable here, that the chosen primer for the NGS marker differs from those of the PCR assay and were not tested on the male and female pools via PCR and subsequent gel electrophoresis in the nature of the previous PCR assays. The accuracy of the previous PCR assay and the apparent potential of the amh and amhby for sufficient sex determination were taken as evidence for the accuracy of the NGS marker. To investigate the phenomenon in regard to the underrepresentation of amhby further, future

studies can implement marker that only work for amh or amhby respectively. Especially due to the high similarities of the two genes, shown by Pan et al. (2021), a set of separate marker for amh and amhby might enable a more robust genotyping assay. Overall, the comparison of the PCR-based assay and the in-silico sexing approach using SNP's and indels in amh and amhby demonstrates the efficiency and reliability of the genomic-based sex determination method. Both approaches constitute a non-invasive and consistent method for sex determination, overcoming the limitations of traditional external sex determination methods based on physical characteristics, which can vary with the maturity and reproductive status of the fish. Although the discrepancy in regard to the unexpected amh:amhby ratios cannot be resolved with certainty, the results are sufficient to be utilized for accurate sex determination, which is what this study aimed for. The results of the PCR-based assay to amplify the target sequences in the Northern Pike individuals showed a high success rate, indicating the reliability of the genetic sex determination based on amh and amhby fragments. The research questions centred around the existence of the presumed sex determination system with amhby as the MSD in the studied populations can therefore be answered. The Northern Pike populations from Rügen that have been tested in this study show the presumed sex determination system, as described in contemporary literature (Pan et al., 2019, 2021; Johnson et al., 2020). The utilisation of SNP's an indels within amh and amhby is sufficient to establish a PCR based as well as a NGS and bioinformatics-based sexing assay for those populations. This can be utilized for further studies, by either performing the established PCR assay, or via the implementation of the sex marker into the genotyping marker panel for European Northern Pike.

4.3 Genetic vs external Sex determination

The reliability of external sex determination in Northern Pike is questionable. Morphological differences between males and females, such as body depth and caudal peduncle shape, vary and become more evident in mature fish (Casselman, 1974). The most common method for external sexing, as presented by Casselman (1974), is the examination of the urogenital opening, where females have convoluted pinkish tissue, while males display a transverse, unpigmented slit. Anyways, this method becomes unreliable during the summer and spawning season when males also show pinkish colours in their urogenital areas. Additionally, the urogenital region's morphology changes as the fish matures, especially in their first three years. A more accurate and less invasive sex determination method is needed due to the challenges associated with the examination in the field and variable urogenital region morphology at different life stages and seasons (Casselman, 1975; Johnson, 2019). The comparison of genetic

and external sex determination methods in Northern Pike has revealed a discrepancy in the external sex determination in approximately 9,2 % of the individuals. Specifically, when analysing the misidentified individuals, it was found that 15.2% of the males and 3.1% of the females were incorrectly sexed based on external characteristics. This disparity highlights a higher chance of females to be misidentified externally compared to males. However, after correcting the misidentified individuals using genetic sex determination, the number of males decreased by approximately 12%, and the number of females increased by around 11% compared to the initial external sexing results. Furthermore, the timing of sampling appeared to influence the accuracy of external sex determination, with the majority of the incorrect sexing occurring during the spring and summer (70% of the incorrectly sexed individuals). This is in accord with the changing urogenital characteristics during the spawning season mentioned above and explains the increase in incorrect sexed individuals during this period. Although the genotyping assay for sex determination applied in this study, can in principle be seen as a success, the nature of the dataset makes the overall significance questionable, especially when absolute numbers are used to infer any conclusions. It is worth noting that the pool of individuals tested in this study derives from various sampling methods that are all positive size selective (Robert Arlinghaus, personal communication, 2023). In combination with the sex specific growth (where females grow faster and larger), which is known for pike, this potentially leads to female biased pools, especially in higher length classes. Considering a potentially inherent female biased pool, the formulation of implications that are based on the number of females and males sexed wrong, and the potential effects of the "corrected" proportions of the sexes on the population have to be done with caution. Further, the full analytical potential of this marker was not utilized in this analysis, preventing the demonstration of potentially interesting correlations and implications. Under consideration of sex-specific growth patterns, the implementation of different length classes and the age of the fish, this could have been used to model sexing errors based on these factors and to infer correlations. The data utilized for this study was simply not representative enough and consisted of an extremely random pool of individuals for which the genotyping assay worked. For many of these fishes other relevant parameter were often not at hand (length, date, age), resulting in a small pool of "incorrectly sexed" individuals which was not feasible for further modelling. Especially in regard to the variety of different sized fishes this pool was not representative enough for any further modelling, since the majority of the fishes was over 60 cm. Both of the mentioned limitations can be assessed in future studies, using the sexing approaches presented in this work, to resolve the potential coherences and correlations this study failed to reveal. However,

considering the challenges associated with external sex determination of Northern Pike, the PCR and NGS based approach for sex provides a reliable, non-invasive, and less variable methodology. The implementation of a genotype-based sex determination assay, once established, offers numerous advantages, including its consistency, reliability, and applicability across various developmental stages and seasonal variations. The sex determination via the genotype leverages the genomic differences between males and females, which are relatively constant and uniform once the sex determination mechanisms are resolved. Although the present error rate of the external sexing and the resulting discrepancy in regard to the proportions of the sexes is notable, the assessment of the applicability of the marker is yet to come. Although it can find its application in biological studies focused on behaviour, selection, and population genetics, where the correct genetic sex can be crucial, it is questionable if the genomic sexing is suitable to ever replace the external sexing in the field. From a practical point of view, the external sexing is sufficient enough for most fishery biological studies (Robert Arlinghaus, personal communication, 2023). Conclusively, despite the limitations, the sex determination based on a genotyping assay provides a valuable tool for population genetic studies, conservation biology, and thus fishery management since it can eventually help to resolve the sex distribution within populations more reliable.

4.4 Sex Ratios in the Rügen Northern Pike populations

The increased female activity, indicated by the significant p-values throughout the year, except for the spawning season, is in accord with the expected annual sex specific behavioural pattern that make them more susceptible to most kind of sampling methods (Casselman, 1975; Johnson, 2019). The insignificant skewed ratio towards males during spawning season, indicated by increased male catches show the change in their activity pattern during this period . They arrive at their spawning sites earlier than the females and start foraging intensively in preparation for reproduction. However, since the ratios throughout the year, except for the spawning season significantly deviates from 1:1 in favour of the females, this is sufficient to indicate the seasonal female specific behaviour. They require more sustenance in summer for growth and in winter for the accumulation reproductive products. They intensified activity compared to the males makes them more susceptible to any kind of sampling method (Casselman, 1975; Johnson, 2019). Those at same time skewed but still informative pattern regarding the sex ratios can also be observed when the populations are differentiated into freshwater and brackish water environments (Figure 5). The brackish water populations display a distinct pattern. For the entire year, the sex ratio is significantly deviates from 1:1 towards female dominance ($p <$

0.005) with an average sex ratio of 1.41. Similar to the observed pattern of the total sex ratios over the year, the autumn, summer, and winter seasons reveal significant deviations from a 1:1 ratio in brackish water environments, with all p-values falling below 0.005. The spring season, however, does not show a significant deviation ($p > 0.05$), potentially due to increased sampling during the spawning season and an alignment of the activity pattern of males and females. In freshwater bodies on the other hand, the sex ratios did not significantly deviate from 1 ($p > 0.05$), with an average sex ratio of 0.885 across the year. The spring season, however, is an exception to this trend. In contrast to the observed overall ratios in the spawning season described before, a significant deviation, although rather tight, was noted (p-value of 0.047). This ratio shift towards males during the spring can be attributed to the increased activity of males during the spawning season, as well as the suggested sampling biases during this period. Additionally, the anadromous cluster are assumed to migrate to the freshwater habitats for spawning as well, potentially further skewing the observed ratios. However, a significant deviation towards males is notable in the spawning season, while the ratios throughout the year maintain insignificant (Appendix 6, Table 6a-c). The analysis of the sex ratios within the cluster partly revealed non-random, sex-specific behaviour variations that are reflected in the sex ratios, particularly among brackish-associated clusters. Although the data was not sufficient to provide statistically significant results for all clusters, it is worth noting that the results of the genotyped individuals indicates non-random, sex-specific behaviour variations for the brackish water associated cluster. In summary, the findings show the biannual peaks in pike availability, irrespective of population and location, indicating intrinsic seasonal activity pattern. The faster-growing females require more sustenance in summer for growth and in winter for the accumulation reproductive products increasing their susceptibility to capture due to intensified foraging. In contrast, males, particularly older, slower-growing individuals, show heightened activity in spring for spawning and autumn for reproductive preparation, but remain relatively inactive during other times of the year (Casselman, 1975; Johnson, 2019). This has been reinforced the findings of this analysis which demonstrated significant female dominance during specific months. Despite the challenges involved in obtaining an unbiased estimate of the sex ratio, the analysis showed consistent pattern, and the temporal variations in sex ratios that align with the two biannual peaks of availability that Casselman (1975) observed. However, this analysis leaves open questions regarding the precise mechanisms driving these observed patterns. It is particularly questionable if the observed trends actually represent the population sex ratio or if they are influenced by selective sampling, to an extend that makes any analysis like this obsolete. One of the main challenges in studying the sex ratio of Northern Pike lies in

the sex-specific growth pattern, where females tend to grow faster than males. This phenomenon leads to an overrepresentation of females, especially in higher size classes, since almost all sampling methods are positive size selective (Robert Arlinghaus, personal communication, 2023). Consequently, a dataset based on this is most likely not feasible to accurately determine the actual sex ratio in the population. Firstly, because fish under 50 cm are significantly underrepresented, and secondly because pools of fish in the length category over 80 are potentially female biased. One more thing to consider is the relatively random pool of individuals that are part of the dataset used for this analysis. The individuals that were part of the first sequencing run with the sex marker, might not be representative in terms of the sampling methods and objectives. The exact catching method and objective during the numerous sampling trips of the for the Boddenhecht-Project was not considered in this work. Given this limitation, research questions focused on the sex ratio in the population might not be meaningful due to the random pools, the lack of information regarding sampling methods and the potential overrepresentation of females in the catches. The reliability of the apparent significant seasonal sex ratio trends, that are supposed to reflect the sex specific behavioural characteristics resulting in sex specific susceptibility for catching methods in different seasons and waterbodies, can therefore be questioned. An inherently female biased pool would consequently lead to false positive results in favour of the females especially in studies focused on sex ratio shifts towards females. This further highlights, that it is difficult to infer any reliable depiction of sex ratios in pike based on field sampling methods. To gain meaningful data for this purpose, breeding experiments, where hatchlings are counted, and subsequent projections can be considered in future analysis. Numerous studies report sex ratio shifts, mostly towards females, and highlight unpredictable potential effects on local populations (Han et al., 2018; Johnson 2019; Kelly et al., 2014; Leopold, Mark & Farrell, 2012). The authors emphasize that studies have yet failed to determine the geographic extent of the skewed sex ratios and the potential implications of female dominated populations on the reproductive health of local populations, pointing out the demand of reliable methods to assess sex ratios. Further studies could also incorporate logistic models using length as a predictor variable, to resolve the influence of sampling biases and incorporate the fish size, age, activity levels, and habitat preferences. To establish solid analysis in terms of sex ratio related research questions it is also necessary to adopt a stratified sampling approaches to mitigate biases and improve the representation of sex ratios within the pools (Robert Arlinghaus, personal communication, 2023)

5. Conclusion

The PCR-based assay successfully determined the genetic sex of Northern Pike individuals with a high success rate, validating its reliability. The presence of the amh and amhby genes in the populations supports the persistence of the known sex determination system for the Rügen cluster. Primer pair A, located in a conserved region and targeting shorter sequences, outperformed the other primer pair in specificity and efficiency. The gel electrophoresis confirmed the amplification of the amh and amhby fragments (depending on the sex) with the expected lengths in the majority of the samples. While most individuals displayed expected banding pattern, a few outlier males lacked the amhby band, probably due to PCR amplification issues. The cause of this phenomenon remains unclear and requires further investigation. One notable outlier consisted of a morphological female exhibiting amhby. Anyways, to infer an indication for alternative or atypical sex determination remains farfetched, due to possible human error during sampling and due to the lack of tissue samples from this individual for verification. Overall, this PCR assay serves as a valuable tool for sex determination in Northern Pike populations, with primer pair A showing promising results. Future studies should investigate outlier male genotypes and potential alternative sex determination mechanisms. In summary the assay's success aligns with previous findings and has the potential for further application in genetic studies regarding Northern Pike. Also, the NGS assay successfully determined the genetic sex in Northern Pike, although amhby was underrepresented compared to amh, contrary to expectations. Possible reasons for this disparity include inaccuracies in the assumed ratio, suboptimal primer binding, or unanticipated amh duplications. Nevertheless, both the PCR-based and NGS approaches proved efficient and reliable for sex determination. Considering an assay with amh and amhby specific primer pair could constitute a more robust approach, especially considering the high similarity between the two genes. In summary, the study confirmed the presumed sex determination system with amhby as the MSD in the studied Northern Pike populations. SNP's and indels within amh and amhby are effective markers for PCR and NGS-based sexing assays. These methods offer non-invasive and consistent sex determination, overcoming limitations of traditional methods based on physical characteristics. Future studies can utilize these findings and incorporate the markers into genotyping panels for European Northern Pike populations. The findings again highlighted that external sex determination in Northern Pike is unreliable and prone to misidentification, especially in the spawning season. The comparison between genetic and external sex determination methods revealed a discrepancy of approximately 9.2% in the external sexing results. Correcting the misidentified individuals using genetic sex determination resulted in a decrease in males and

an increase in females compared to the initial external sexing results. The genetic approach offers consistency, reliability, and applicability across different developmental stages and seasonal variations. However, it may not completely replace external sexing in practical field applications, where the latter remains sufficient for most fishery biological studies. Despite its limitations, the genotyping assay provides a valuable tool for studies focused on population genetics and selection, since it offers more reliable and detailed sex distribution information within populations. The attempt to reveal sex ratio pattern over the year based on the relatively random dataset that resulted from the genotyping assay revealed significant differences in female and male activity patterns throughout the year. Despite challenges in obtaining an unbiased estimate of the sex ratio due to sex-specific growth patterns, the analysis indicated non-random, sex-specific behaviour variations in the sex ratios. However, questions remain in regard to the extent to which the observed trends reflect the actual sex ratios in the populations. Considering the effects resulting from the sex specific growth, in combination with size selective sampling methods, the sample pool is most likely not suitable for this kind of analysis. Nonetheless, this study highlights the difficulties in accurately determining sex ratios in Northern Pike populations based on field sampling methods, particularly due to the inherent female bias in catches due to the combination of the sex specific growth and inherently positive size selective sampling methods. Additional research could employ logistic models, considering fish length as a predictor variable, to account for the influence of sampling biases and further incorporate factors such as age, activity levels, and habitat preferences. Further, to ensure robust analysis for sex ratio-related research questions, implementing stratified sampling approaches would be essential (Robert Arlinghaus, personal communication, 2023). These strategies would help mitigate biases and enhance the accurate representation of sex ratios within the sampled populations. Finally, this study contributed to the understanding of sex determination in Northern Pike populations in the Baltic Sea, offering valuable insights into the effectiveness of PCR and NGS-based sexing assays for this species. The identified SNP's and indels within amh and amhby genes serve as promising markers for future sex determination studies in European Northern Pike populations. Further research can explore correlations and implications of this genetic sex determination approach, to unlock its potential to resolve sex-related questions and contributing to the broader field of genomics in fish biology.

6. References

Ancona, S., F. V. Dénes, O. Krüger, T. Székely, and S. R. Beissinger. 2017. Estimating adult sex ratios in nature. Philosophical Transactions of the Royal Society B: Biologica l Sciences 372: 20160313.

Arlinghaus, R., T. Rittweg, F. Dhellemmes, D. Koemle, R. Van Gemert, H. Schubert, D. Niessner, et al. 2023. A synthesis of a coastal northern pike (Esox lucius) fishery and its social-ecological environment in the southern Baltic Sea: Implications for the management of mixed commercial-recreational fisheries. Fisheries Research 263: 106663.

Avise, J. C., and J. E. Mank. 2009. Evolutionary Perspectives on Hermaphroditism in Fishes. Sexual Development 3: 152–163.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illuminasequence data. Bioinformatics, 30(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Casselman, J. M. 1975. Sex Ratios of Northern Pike, Esox lucius Linnaeus. Transactions of the American Fisheries Society 104: 60–63.

Chen, Y.-C., T. Liu, C.-H. Yu, T.-Y. Chiang, and C.-C. Hwang. 2013. Effects of GC Bias in Next-Generation-Sequencing Data on De Novo Genome Assembly Y. Xu [ed.],. PLoS ONE 8: e62856.

Devlin, R. H., and Y. Nagahama. 2002. Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. Aquaculture 208 : 191–364.

Eschbach, E., A. W. Nolte, K. Kohlmann, J. Alós, S. Schöning, and R. Arlinghaus. 2021. Genetic population structure of a top predatory fish (northern pike, Esox lucius) covaries with anthropogenic alteration of freshwater ecosystems. Freshwater Biology 66: 884–901.

Fowler, B. L. S., and V. P. Buonaccorsi. 2016. Genomic characterization of sex-identification markers in Sebastes carnatus and Sebastes chrysomelas rockfishes. Molecular Ecology 25: 2165–2175.

Frenzel, Peter & OERTEL, P. (2002). Die rezenten Ostrakoden und Foraminiferen des Strelasunds. Rostocker Meeresbiologische Beiträge. 11. 23-37.

Goto-Kazeto, R., Y. Abe, K. Masai, E. Yamaha, S. Adachi, and K. Yamauchi. 2006. Temperature-dependent sex differentiation in goldfish: Establishing the temperature-sensitive period and effect of constant and fluctuating water temperatures. Aquaculture 254: 617–624.

Healy, T. M., and P. M. Schulte. 2015. Phenotypic plasticity and divergence in gene expression. Molecular Ecology 24: 3220–3222.

Hershberg, R. 2016. Codon Usage and Translational Selection. Encyclopedia of Evolutionary Biology, 293–298. Elsevier.

Heule, C., W. Salzburger, and A. Böhne. 2014. Genetics of Sexual Development: An Evolutionary Playground for Fish. Genetics 196: 579–591.

Huffman, Kelly & Farrell, John & Whipps, Christopher. (2014). Environmental Determinants of Sex Ratio in St. Lawrence River Northern Pike: Development of a Molecular Sex Identification Tool and Experimentation with Physical and Chemical Variables.

Leopold, Mark & Farrell, John. (2012). Comparing Sex Ratios of Northern Pike Between the St. Lawrence River and Inland New York Waters.

Jensen, M. P., C. D. Allen, T. Eguchi, I. P. Bell, E. L. LaCasella, W. A. Hilton, C. A. M. Hof, and P. H. Dutton. 2018. Environmental Warming and Feminization of One of the Largest Sea Turtle Populations in the World. Current Biology 28: 154-159.e4.

Johnson, H. 2019. Northern Pike of North America: Population Genomics and Sex Determination by.

Johnson, H. A., E. B. Rondeau, D. R. Minkley, J. S. Leong, J. Whitehead, C. A. Despins, B. E. Gowen, et al. 2020. Population genomics of North American northern pike: variation and sex-specific signals from a chromosome-level, long read genome assembly. Genomics.

Kallman, K. D. 1983. The Sex Determining Mechanism of the Poeciliid Fish, Xiphophorus montezumae, and the Genetic Control of the Sexual Maturation Process and Adult Size. Copeia 1983: 755.

Martinez, P., A. M. ViÃ±as, L. SÃ¡nchez, N. DÃ-az, L. Ribas, and F. Piferrer. 2014. Genetic architecture of sex determination in fish: applications to sex ratio control in aquaculture. Frontiers in Genetics 5.

Matsuda, M., Y. Nagahama, A. Shinomiya, T. Sato, C. Matsuda, T. Kobayashi, C. E. Morrey, et al. 2002. DMY is a Y-specific DM-domain gene required for male development in the medaka fish. Nature 417: 559–563.

Möller, S., H. M. Winkler, S. Richter, and R. Bastrop. 2021. Genetic population structure of pike (Esox lucius Linnaeus, 1758) in the brackish lagoons of the southern Baltic Sea. Ecology of Freshwater Fish 30: 140–149.

Myosho, T., H. Otake, H. Masuyama, M. Matsuda, Y. Kuroki, A. Fujiyama, K. Naruse, et al. 2012. Tracing the Emergence of a Novel Sex-Determining Gene in Medaka, Oryzias luzonensis. Genetics 191: 163–170.

Nordahl, O., P. Koch-Schmidt, J. Sunde, Y. Yıldırım, P. Tibblin, A. Forsman, and P. Larsson. 2019. Genetic differentiation between and within ecotypes of pike (ESOX LUCIUS) in the Baltic Sea. Aquatic Conservation: Marine and Freshwater Ecosystems 29: 1923–1935.

Pan, Q., R. Feron, E. Jouanno, H. Darras, A. Herpin, B. Koop, E. Rondeau, et al. 2021a. The rise and fall of the ancient northern pike master sex-determining gene. eLife 10: e62858.

Pan, Q., R. Feron, E. Jouanno, H. Darras, A. Herpin, B. Koop, E. Rondeau, et al. 2021b. The rise and fall of the ancient northern pike master sex-determining gene. eLife 10: e62858.

Pan, Q., R. Feron, A. Yano, R. Guyomard, E. Jouanno, E. Vigouroux, M. Wen, et al. 2019a. Identification of the master sex determining gene in Northern pike (Esox lucius) reveals restricted sex chromosome differentiation C. L. Peichel [ed.],. PLOS Genetics 15: e1008013.

Pan, Q., R. Feron, A. Yano, R. Guyomard, E. Jouanno, E. Vigouroux, M. Wen, et al. 2019b. Identification of the master sex determining gene in Northern pike (Esox lucius) reveals restricted sex chromosome differentiation C. L. Peichel [ed.],. PLOS Genetics 15: e1008013.

Pla, S., C. Benvenuto, I. Capellini, and F. Piferrer. 2022. Switches, stability and reversals in the evolutionary history of sexual systems in fish. Nature Communications 13: 3029.

Sedlazeck, F. J., Rescheneder, P., & von Haeseler, A. (2013). NextGenMap: Fast and accurate read mapping in highly polymorphic genomes. Bioinformatics, 29(21), 2790–2791. https://doi.org/10.1093/bioinformatics/btt468

Schultheis, C., A. Böhne, M. Schartl, J. N. Volff, and D. Galiana-Arnoux. 2009. Sex Determination Diversity and Sex Chromosome Evolution in Poeciliid Fish. Sexual Development 3: 68–77.

Sreedharan, S. P., A. Kumar, and P. Giridhar. 2018. Primer design and amplification efficiencies are crucial for reliability of quantitative PCR studies of caffeine biosynthetic N-methyltransferases in coffee. 3 Biotech 8: 467.

Sunde, J., Y. Yıldırım, P. Tibblin, D. Bekkevold, C. Skov, O. Nordahl, P. Larsson, and A. Forsman. 2022. Drivers of neutral and adaptive differentiation in pike (Esox lucius) populations from contrasting environments. Molecular Ecology 31: 1093–1110.

Suzuki, Y., and Y. Iwasa. 1980. A sex ratio theory of gregarious parasitoids. Population Ecology 22: 366–382.

Tørresen, O. K., B. Star, S. Jentoft, W. B. Reinar, H. Grove, J. R. Miller, B. P. Walenz, et al. 2017. An improved genome assembly uncovers prolific tandem repeats in Atlantic cod. BMC Genomics 18: 95.

Valenzuela, N., and V. Lance eds. . 2004. Temperature-dependent sex determination in vertebrates. Smithsonian Books, Washington, D.C.

7. Appendix

1. Sequence search algorithm

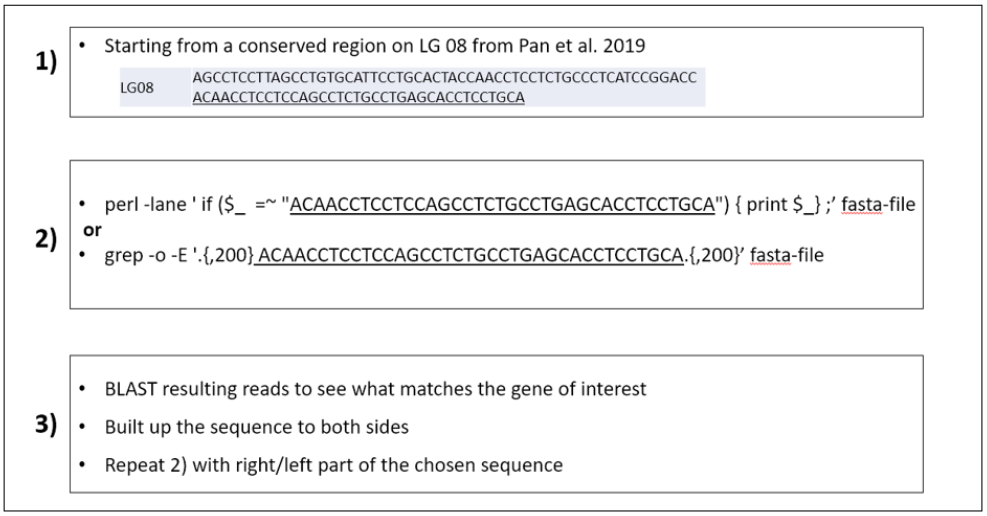


Figure 1: The search algorithm for the reconstruction of amh and amhby sequences based on a conserved sequence from Pan et al. 2019 and the PoolSeq reads of two populations: 1. Sehrowbach: male-biased (40m, 9f, 8 unknown; data from F.Dhellemmes): S2603Nr6. 2. GJB: female-biased (15m, 27f, 8 unknown; data from F.Dhellemmes): S2603Nr4

2. Reconstructed sequences and primer regions

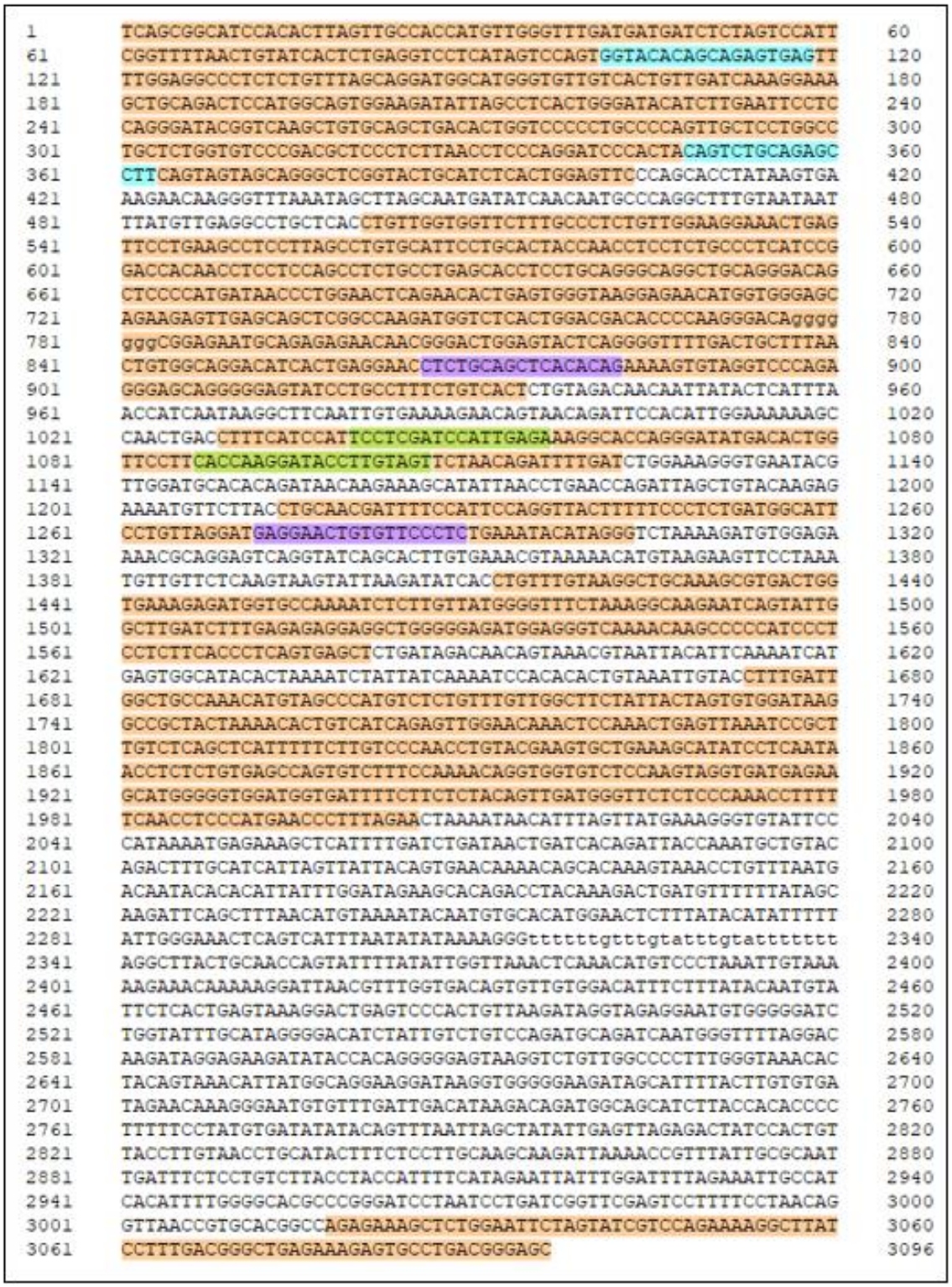


Figure 2a: Showing the reconstructed sequence for amhby. The brownish colour indicates the exons starting from the top with exon 7 and ending on the bottom with exon 1. The blue highlighted sequences represent the primer pair A, the purple ones represents primer pair B used for the PCR assays. The green sequences represent the primer pair chosen for the NGS marker. The sequence was reconstructed based on the comparison to an assembly on GenBank (accession: MK355503.1). The depiction is Reconstructed based on the on NCBI's alignment tools.

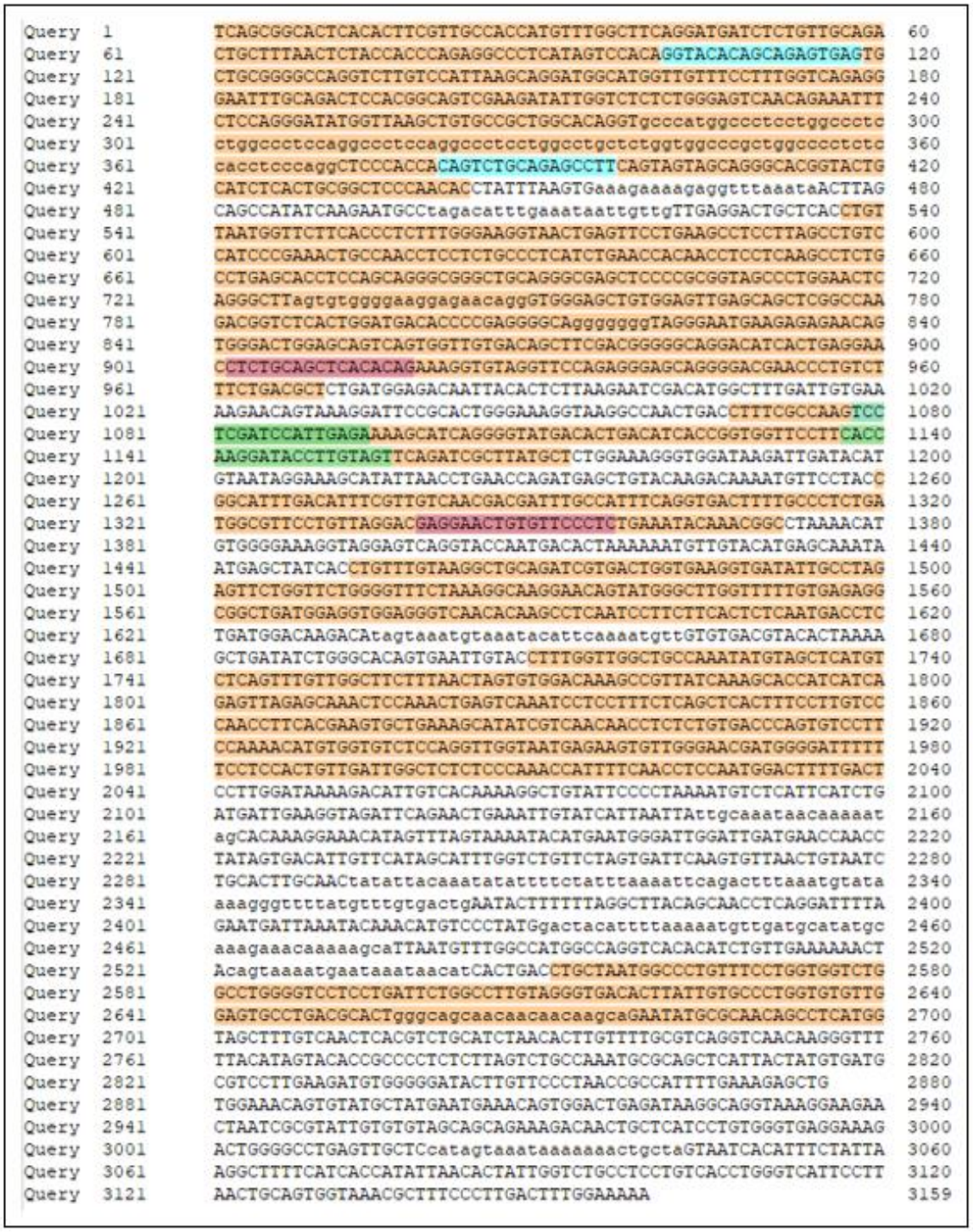


Figure 2b: Showing the reconstructed sequence for amh. The brownish colour indicates the exons starting from the top with exon 7 and ending on the bottom with exon 1. The blue highlighted sequences represent the primer pair A, the red/purple ones represents primer pair B which were both tested for the PCR assays. The green sequences represent the primer pair C chosen for the NGS marker. The sequence was reconstructed based on the comparison to an assembly on GenBank (accession: NC_047576.1). The depiction is Reconstructed based on the on NCBI's alignment tools.

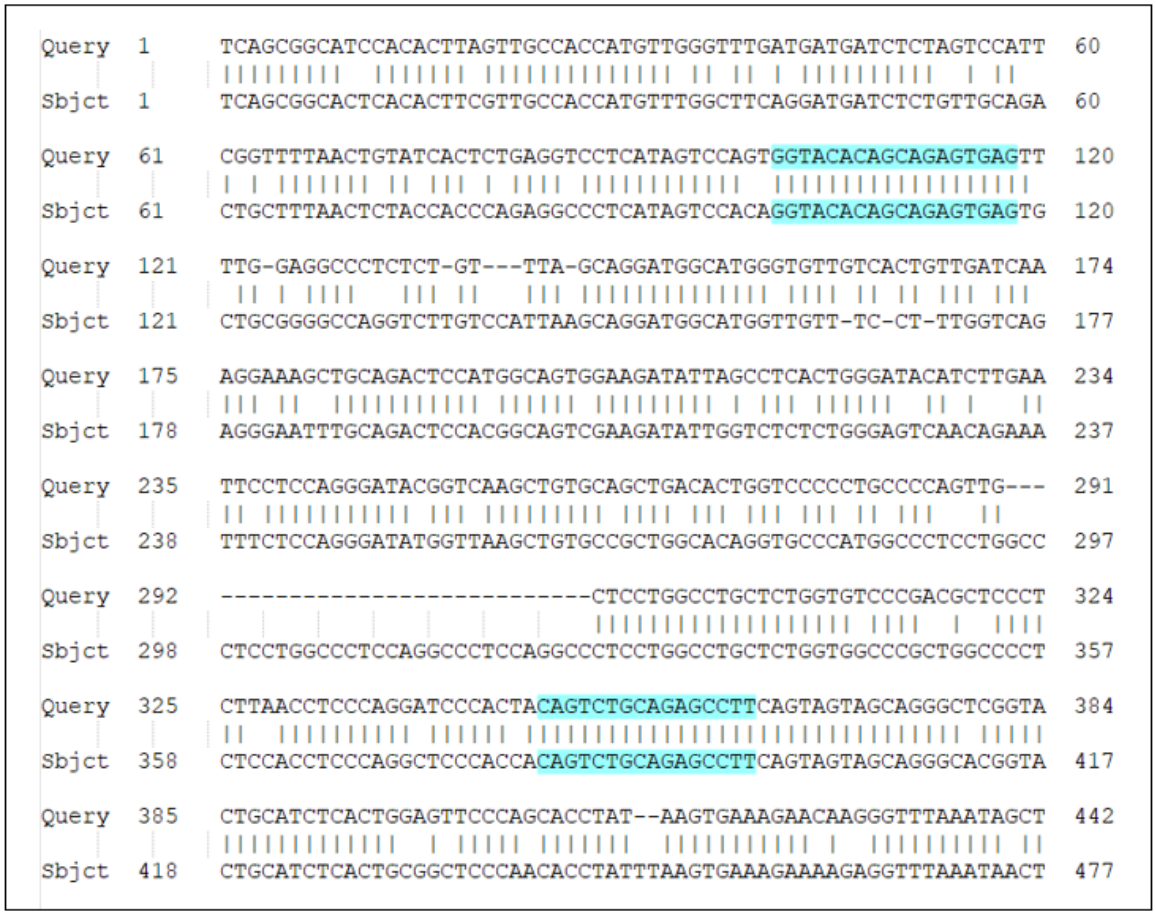


Figure 2c: Showing a part of the alignment of amhby and amh via NCBI's alignment tools. The query is amhby and subject is amh. The blue highlighted sequences represent the primer pair A used for the PCR assays. Primer pair A results in amplicons of ~296 bp for amh and ~263 bp for amhby.

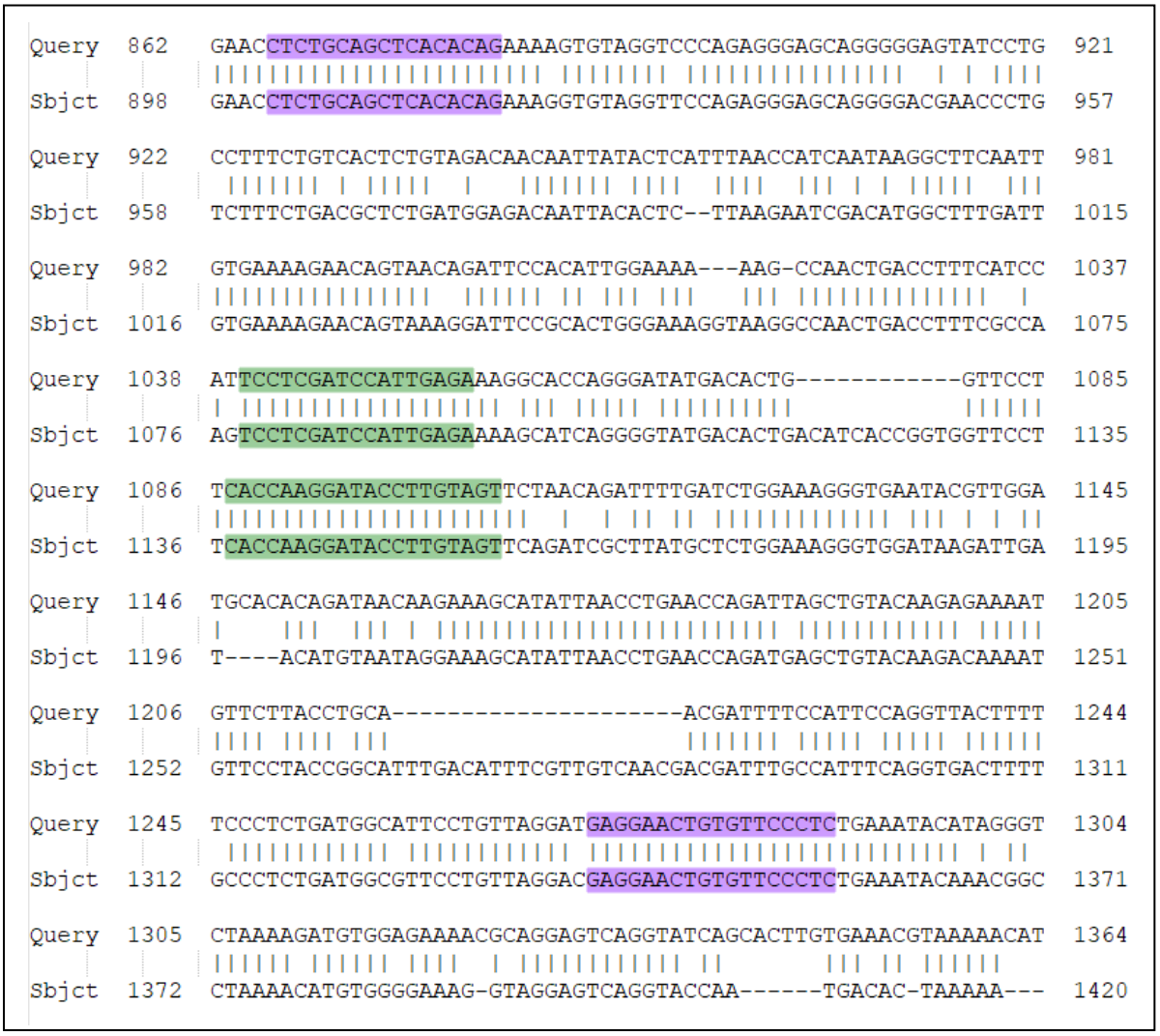


Figure 2d: Showing a part of the alignment of amhby and amh via NCBI's alignment tools. The query is amhby and subject is amh. The purple highlighted sequences represent primer pair B used for the PCR assays. The green highlighted sequences represent the primer pair C for the NGS marker. Primer pair B results in amplicons of ~456 bp for amh and ~424 bp for amhby. The NGS marker was not tested via PCR and gel electrophoresis in the same manner as the pairs A and B, but directly used for downstream analysis.

Table 1: Primer sequences A-C. Pairs A and B were used for the PCR assays and C was used for the NGS assay. The parameters were calculated via ThermoFisher's multiple primer analyser tool (www.thermofisher.com).

Primer	Sequence	Tm°C	CG%	nt	A	T	C	G	Extinction coeff.	Mol. w.	nmol	µg/OD260
fwd_A	ggtacacagcagagtgag	54.5	55.6	18	6.0	2.0	3.0	7.0	188500.0	5597.7	5.3	29.7
rev_A	aaggctctgcagactg	55.2	56.3	16	4.0	3.0	4.0	5.0	153700.0	4906.3	6.5	31.9
fwd_B	ctctgcagctcacacag	57.1	58.8	17	4.0	3.0	7.0	3.0	154300.0	5115.4	6.5	33.2
rev_B	gagggaacacagttcctc	57.6	55.6	18	5.0	3.0	5.0	5.0	176100.0	5508.6	5.7	31.3
fwd_C	tcctcgatccattgaga	58.1	47.1	17	4.0	5.0	5.0	3.0	161700.0	5145.4	6.2	31.8
rev_C	actacaaggtatccttggtg	56.2	45.0	20	5.0	6.0	4.0	5.0	195400.0	6132.1	5.1	31.4

3. Laboratory procedures in detail

3.1 DNA Extraction (salt extraction)

The DNA extraction protocol was initiated by adding tissue samples (fin clips stored in ethanol) that were dried using a paper towel to a 1.5 ml microcentrifuge tube. To facilitate the release of DNA, a total of 300 µl of Homogenization (HOM) Buffer, composed of 80 mM EDTA, 100 mM Tris, and 0.5% SDS, and 3 µl of Proteinase K (at a concentration of 20 mg/ml) were transferred to the tube. This mixture was incubated at 55°C for approximately 1-2 hours or until the tissue was fully dissolved. Upon complete dissolution of the tissue, the tubes were cooled on ice for one minute to reach room temperature. RNAse A, at a concentration of 10 mg/ml, was then added to the cooled mixture in a volume of 10 µl. This RNAse treatment, carried out at room temperature for 30 minutes, served to eliminate any RNA contaminants present in the sample. Subsequently, 300 µl of a 4.5 Molar Sodium-Chloride solution was added to the mixture, which was then thoroughly mixed by inversion. This salt-based DNA extraction process causes proteins and other cell debris to precipitate out of the solution. The mixture was then left on ice for an additional 10 minutes. Post-incubation, a centrifugation step was performed at 16,000 g for 15 minutes, causing further sedimentation of proteins and other debris. Approximately 500 µl of the supernatant, which contained the DNA, was carefully transferred to a new tube, taking care to avoid the precipitate at the bottom. To this DNA-containing solution, an equal volume of pure ethanol (500 µl) was added and thoroughly mixed by inversion. The mixture then incubated at room temperature for 5 minutes, during which the DNA precipitated out of the solution. Another round of centrifugation at maximum speed (16.000g) for 10 minutes resulted in the formation of a visible DNA pellet. The supernatant was discarded, and 500 µl of 70% ethanol was added to the pellet again, followed by thorough mixing and a 5-minute incubation. After the washing step, as much supernatant was removed as possible. The ethanol washing step, followed by centrifugation, aids in the removal of salts and other contaminants, and was repeated twice. Finally, the DNA pellet was allowed to air-dry. It is worth noting that the drying process was most of the time speed up by using a vacuum centrifuge. By suspending the open 1,5 ml tubes, containing the pallet and ethanol rests to a vacuum (without centrifugation), the drying process can be significantly accelerated without harming extracted DNA. Eventually the DNA pallets were resuspended in TE Buffer, which ranged from 20-50 µl, depending on the size of the pellet. This buffer serves to protect the DNA and keep it stable. Gradually, the DNA was extracted in the described manner from 112 *Esox lucius* fin clips and subsequently quantified using a BioSpectrometer. To determine the purity

and quantity of the extracted DNA the c measures the concentration of DNA and assess its purity by providing a 260/280 ratio. The extracted DNA can then be checked by running a gel electrophoresis to identify if the DNA has degraded or if it is intact (not conducted). The DNA samples were then diluted to a working concentration between 20-100 ng/µl in preparation for amplification, gel electrophoresis and library preparation.

3.2 Polymerase Chain Reaction (PCR)

After extraction and dilution, The DNA was amplified via Polymerase Chain Reaction (PCR). The PCR generally consists of repeated cycles of denaturation (separating the DNA strands), annealing (binding of primers to the DNA strands), and extension (synthesis of new DNA strands). In preparation for DNA amplification the samples and master mixes were provided as shown in Table 2a-f. This amplification was carried out with the QIAGEN Multiplex PCR kit containing the QIAGEN Multiplex PCR Master Mix (HotStarTaq DNA Polymerase, MgCl2, dNTPs PCR buffer). After DNA extraction and dilution, the DNA was amplified using the Polymerase Chain Reaction (PCR). The QIAGEN Multiplex PCR kit with HotStarTaq DNA Polymerase, MgCl2, dNTPs PCR buffer was used. Q-Solution was added to enhance amplification and prevent potential issues with secondary structures or GC-rich templates (only for primer pair A). The PCR cycle for primer pair A consisted of six steps: Initial denaturation at 95.0 °C for 15 minutes, followed by denaturation at 94.0 °C for 30 seconds, annealing at 62.0 °C for 1 minute and 30 seconds, extension at 72.0 °C for 30 seconds, final extension at 72.0 °C for 10 minutes, and a final hold at 10.0 °C (for primer pair B, see Table 2c-d). These steps are repeated 25 times to amplify the DNA target. To determine the ideal annealing temperature of the primers, a gradient PCR was initially performed at different annealing temperatures. The optimal annealing temperature from the gradient PCR was then used to establish a cycling protocol for subsequent PCR reactions. The preparation of the master mixes and cycling conditions for both primer pairs are shown in Table 2e-f. The quality and quantity of the amplified DNA can then be assessed via gel electrophoresis.

Table 2a: Master mix for one sample for primer pair A.

Supplies per 1 sample	**µl**
2 x MM	5
Forward primer (2 µM)	1
Reverse primer (2 µM)	1
H2O	1
Q-Solution	1
Template DNA (20 - 100 mg)	1

Table 2b: Cycling conditions for primer pair A.

	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
Time	15:00 min	00:30 min	01:30 min	00:30 min	10:00 min	∞
Temperature	95.0 °C	94.0 °C	57.0 °C	72.0 °C	72.0 °C	10.0 °C
Repeats		28x				

Table 2c: Master mix for one sample for primer pair B

Supplies per 1 sample	**µl**
2 x MM	5
Forward primer (2 µM)	1
Reverse primer (2 µM)	1
H2O	2
Q-Solution	0
Template DNA (20 - 100 mg)	1

Table 2d: Cycling conditions for primer pair B

	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
Time	15:00 min	00:30 min	01:30 min	00:30 min	10:00 min	∞
Temperature	95.0 °C	94.0 °C	62.0 °C	72.0 °C	72.0 °C	10.0 °C
Repeats		25x				

Table 2e: Master mix for 1 sample for the gradient PCR

Supplies per 1 sample	**µl**
2 x MM	5
Forward primer (2 µM)	1
Reverse primer (2 µM)	1
H2O	2
Q-Solution	0
Template DNA (20 - 100 mg)	1

Table 2f: Cycling conditions for gradient PCR.

Cycling conditions Gradient PCR		
95°C	15min	
94°C	30s	
50-65°C	90s	25x
72°C	30s	
72°C	10min	

3.3 Gel Electrophoresis

The first step in the preparation of the gel for DNA electrophoresis involved the casting of a 3% Low Electroendoosmotic (LE) Agarose Gel. It is suggested to ensure the Tris-acetate-EDTA (TAE) buffer was cooled down to 4°C. However, for applications with relatively low voltage applied, room temperature is sufficient. The required amount of LE Agarose was calculated based on the desired volume of the gel, with the concentration set at 3%. In a 250ml Erlenmeyer flask, the calculated quantity of TAE buffer and LE Agarose was combined, with the agarose added first. To account for potential water loss due to the evaporation during the heating process (conducted in a customary microwave), approximately 10g of water was added to the flask. The Erlenmeyer flask was covered, placed in the microwave, and heated at approximately 600W until the agarose was fully dissolved. The weight of the flask was periodically checked throughout the heating process to monitor the loss due to evaporation, using a precision scale. If the weight fell below the original measurement, additional water was added to maintain the desired volume. Following the heating, the flask was cooled for about 60 seconds in a lukewarm water bath to bring down the temperature without causing the agarose to solidify. Once cooled, the agarose solution was poured into a gel casting tray, and the combs were inserted. The gel was then left at room temperature for approximately 20 minutes to ensure appropriate cooling and solidification. Before loading the gel, master mixes containing the loading dye and the DNA were prepared as shown in Table 3a. Once loaded, the gel run at constant voltage of 125V for 60 minutes. The electric field, which caused the negatively charged DNA fragments to move towards the positive electrode, allowed for the migration of the DNA fragments, resulting in distinct bands corresponding to different fragment sizes. Smaller fragments moved faster and thus farther than larger ones, separating the fragments by size. A 100bp equimolar DNA ladder was used as a standard.

Table 3a: Shows the preparation for the Gel electrophoresis for the samples and for the ladder.

MM Probe LD			
	µL	Proben	Total (+15%)
LD6x Red	1	1	1,15
H2O	2	1	2,3
	3Probe+3MM		
Ladder			
	µL	Ladders	Total (+15%)
LD6x Red	1	1	1,15
Ladder	1,5	1	1,725
H2O	3,5	1	4,025

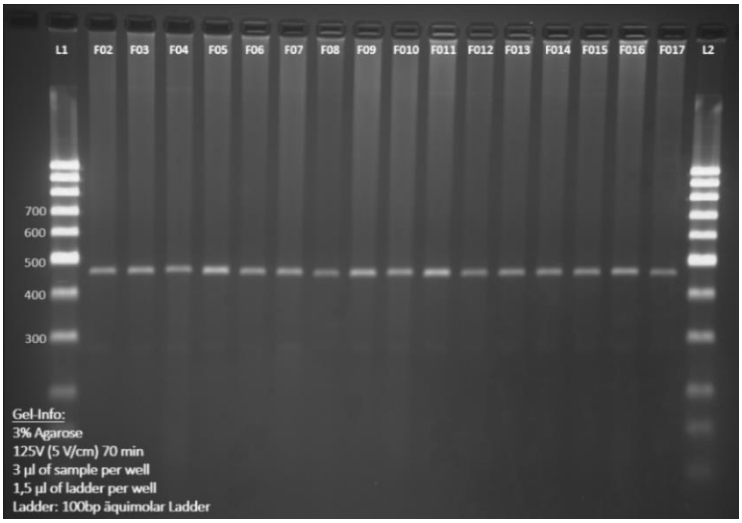


Figure 3a: Showing the 3% agarose gel of the PCR 1 with primer pair B, including the used parameter. It shows 16 confirmed females, flanked by the ladders.

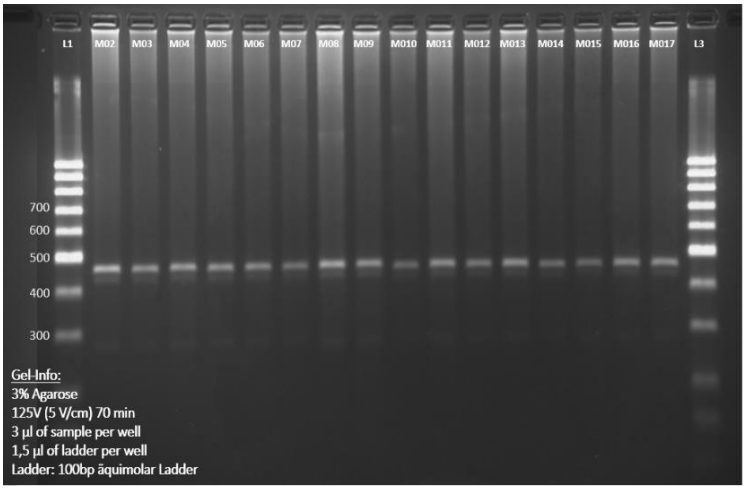


Figure 3b: Showing the 3% agarose gel of the PCR 1 with primer pair B, including the used parameter. It shows 16 confirmed males, flanked by the ladders.

Table 3b: Showing the IDs of the fish and the Gel ID all individuals for PCR1.

Gel_ID-	Fish_ID	Gel_ID	Fish_ID
F02	BH-51246	M02	BH-93061
F03	BH-51397	M03	BH-93111
F04	BH-51614	M04	BH-93148
F05	BH-90546	M05	BH-93165
F06	BH-91534	M06	BH-93166
F07	BH-91548	M07	BH-93167
F08	BH-91549	M08	BH-93170
F09	BH-91765	M09	BH-93174
F010	BH-91789	M10	BH-93175
F011	BH-92990	M11	BH-93176
F012	BH-92991	M12	BH-93177
F013	BH-92993	M13	BH-93179
F014	BH-92996	M14	BH-93184
F015	BH-92998	M15	BH-93186
F016	BH-93002	M16	BH-93188
F017	BH-93013	M17	BH-93191

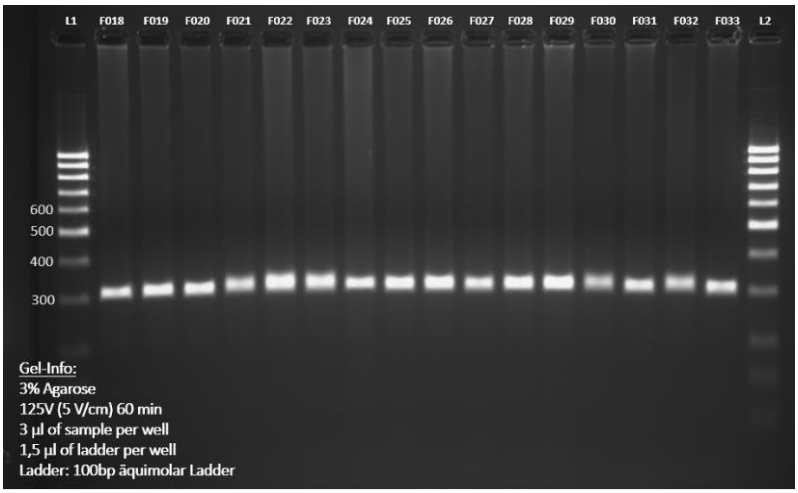


Figure 3c: Showing the 3% agarose gel of the PCR 2 with primer pair A, including the used parameter. It shows 16 confirmed females, flanked by the ladders.

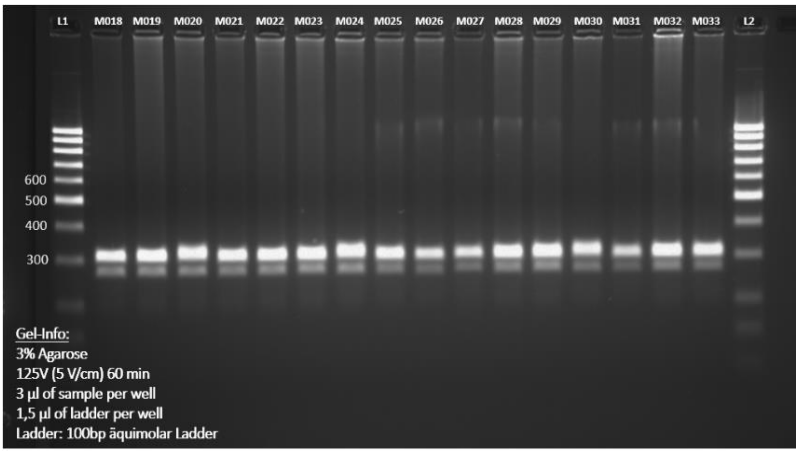


Figure 3d: Showing the 3% agarose gel of the PCR 2 with primer pair A, including the used parameter. It shows 16 confirmed males, flanked by the ladders.

Table 3c: Showing the IDs of the fish and the Gel ID for all individuals for PCR2.

Gel_ID	Fish_ID	Gel_ID	Fish_ID
F018	BH-93003	M018	BH-93162
F019	BH-93005	M019	BH-93171
F020	BH-93006	M020	BH-93182
F021	BH-93007	M021	BH-93183
F022	BH-93015	M022	BH-93185
F023	BH-93021	M023	BH-93187
F024	BH-93039	M024	BH-92994
F025	BH-93041	M025	BH-92995
F026	BH-93044	M026	BH-93034
F027	BH-93045	M027	BH-93051
F028	BH-93046	M028	BH-93056
F029	BH-93057	M029	BH-93059
F030	BH-93074	M030	BH-93060
F031	BH-93084	M031	BH-93088
F032	BH-93099	M032	BH-93093
F033	BH-93108	M033	BH-93097

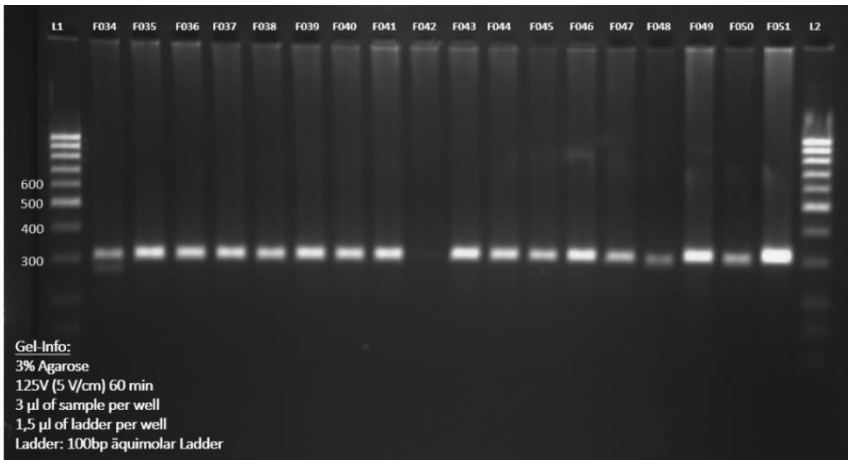


Figure 3e: Showing the 3% agarose gel of the PCR 3 with primer pair A, including the used parameter. It shows 18 confirmed females, flanked by the ladders. Individuals with weak bands or contrary results (deviating from 1 band for female, 2 bands for male) were chosen for repeats.

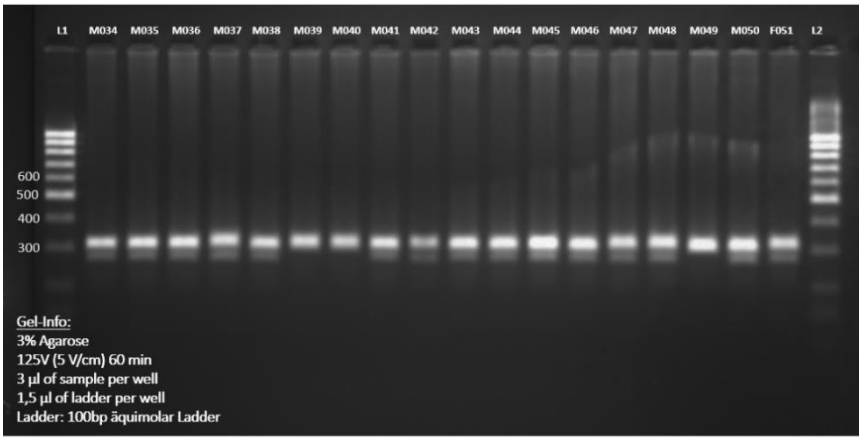


Figure 3f: Showing the 3% agarose gel of the PCR 3 with primer pair A, including the used parameter. It shows 18 confirmed males, flanked by the ladders. Individuals with weak bands or contrary results (deviating from 1 band for female, 2 bands for male) were chosen for repeats.

Table 3d: Showing the IDs of the fish and the Gel ID for all individuals for PCR2.

Gel_ID	Fish_ID	Gel_ID	Fish_ID
F034	BH-93116	M034	BH-93114
F035	BH-93117	M035	BH-93122
F036	BH-93121	M036	BH-93118
F037	BH-93125	M037	BH-93130
F038	BH-93126	M038	BH-93133
F039	BH-93127	M039	BH-93134
F040	BH-93132	M040	BH-93135
F041	BH-93140	M041	BH-93136
F042	BH-93141	M042	BH-93144
F043	BH-93143	M043	BH-93147
F044	BH-93152	M044	BH-93154
F045	BH-93155	M045	BH-93156
F046	BH-93169	M046	BH-93157
F047	BH-93172	M047	BH-93159
F048	BH-93173	M048	BH-93163
F049	BH-93178	M049	BH-93164
F050	BH-93180	M050	BH-93195
F051	BH-93190	M051	BH-93196

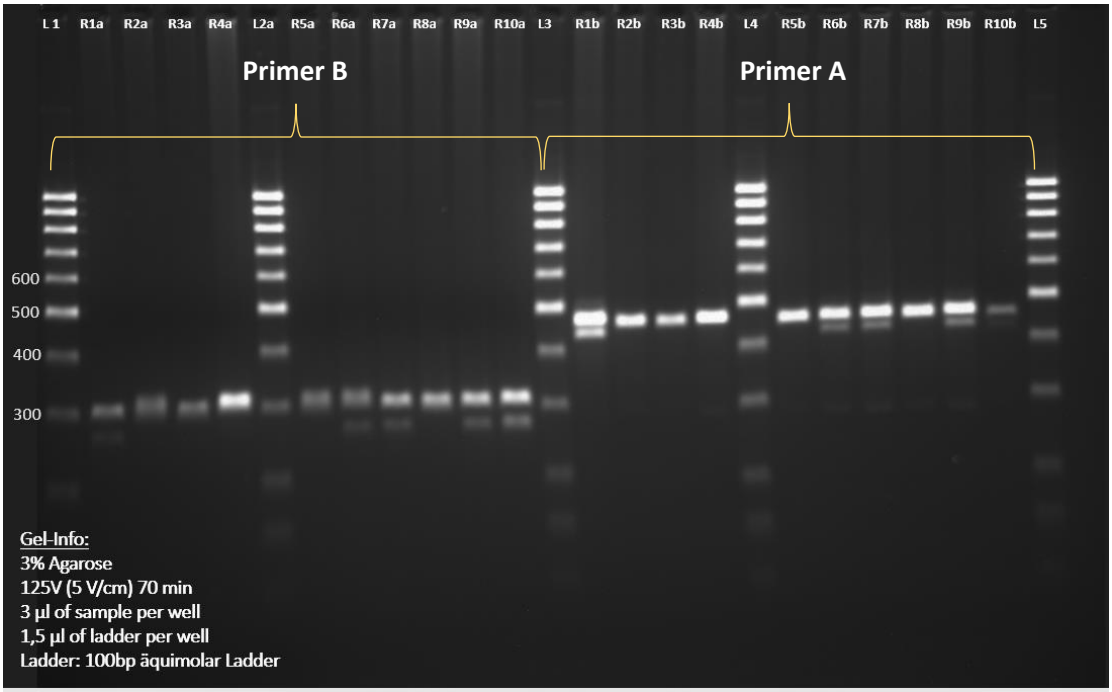
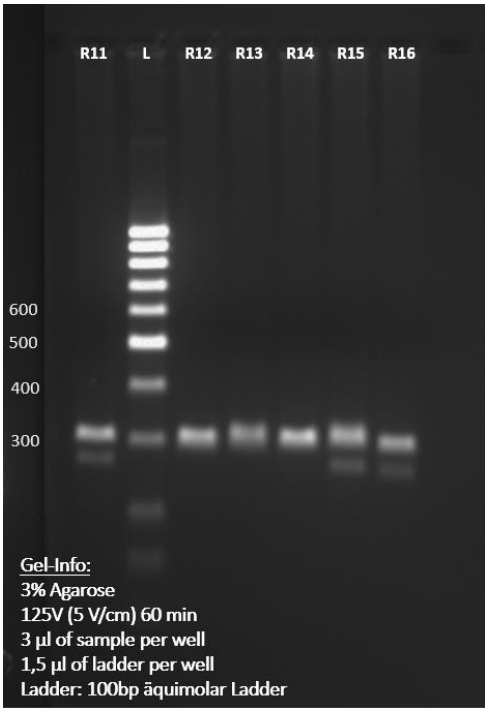


Figure 3g: Showing the 3% agarose gel of the Repeat 1, including the used parameter. The individuals were tested with both primers in one gel. Primer pair B left half of the gel, and primer pair A, right half of the gel. It shows in total 10 individuals, 4 confirmed females (left), and 6 confirmed males (right) flanked by the ladders respectively. Individuals with weak bands or contrary results (deviating from 1 band for female, 2 bands for male) as well as leftover confirmed individuals were chosen for repeats.

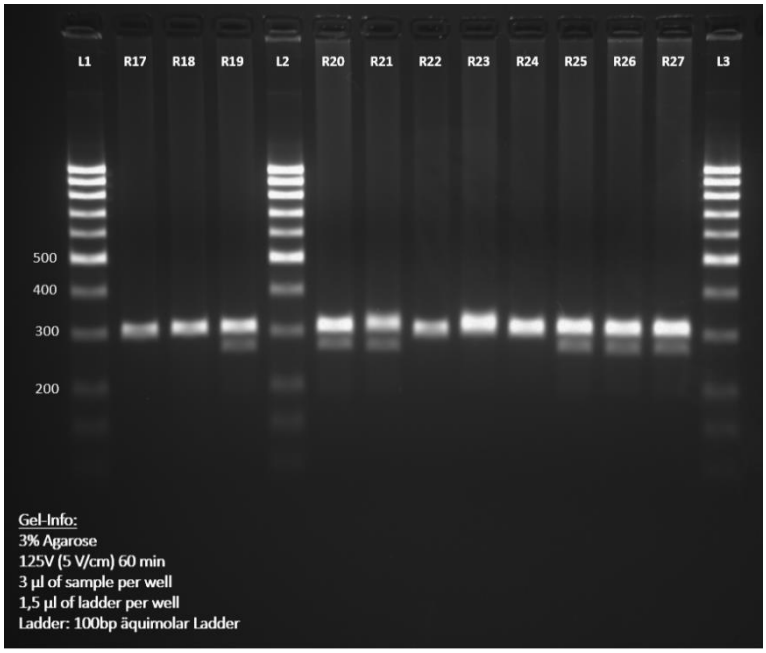
Table 3e: Showing the IDs of the fish and the Gel ID for all individuals for Repeat 1.

Gel_ID	Fish_ID
R1	BH-93116
R2	BH-93141
R3	BH-93173
R4	BH-93180
R5	BH-93134
R6	BH-93160
R7	BH-93107
R8	BH-93164
R9	BH-93100
R10	BH-93135



Gel_ID	Fish_ID
R11	BH-93116
R12	BH-91565
R13	BH-93134
R14	BH-93164
R15	BH-93168
R16	BH-93193

Figure 3h: Showing the 3% agarose gel of the Repeat 2 with primer pair A, including the used parameter and the ID-table (right). It shows 1 confirmed female left next to the ladder and 5 confirmed males right to the ladder. Individuals with weak bands or contrary results (deviating from 1 band for female, 2 bands for male) as well as leftover confirmed individuals were chosen for repeats.



Gel_ID	Fish_ID
R17	BH-93047
R18	BH-93048
R19	BH-93116
R20	BH-93051
R21	BH-93109
R22	BH-91565
R23	BH-93134
R24	BH-93164
R25	BH-93135
R26	BH-93160
R27	BH-93149

Figure 3i: Showing the 3% agarose gel of the Repeat 3 with primer pair A, including the used parameter and the ID-table (right). It shows 3 confirmed females (left), and 8 confirmed males (right).

4 Library Preparation

4.1 DNA extraction Dneasy 96 Blood & Tissue Kit

For the majority of the pike samples from the Boddenhecht-Project the extraction was performed using the DNeasy 96 Blood & Tissue Kit. The kit was slightly adapted for tissue (fin clip) as described in the following. First, approximately 20 mg of tissue were added to each well of a collection microtube plate. A master mix was then prepared by mixing 2ml of Proteinase K with 18ml of Buffer ATL which is then transferred to the plate until each well is supplemented with 200µl. Following this, the rows of the plate are sealed with Collection Microtube Caps, and the plate cover is attached. The mixture is then thoroughly, briefly centrifuged at up to 3000rpm, and subjected to the lysis procedure at 56°C in a thermosphere until the solution is clear. The plate is then covered and shaken vigorously for 15 seconds. After a brief centrifugation, 410µl of a Puffer AL-Ethanol Mix is added to each well. This solution is again shaken vigorously for 15 seconds and briefly centrifuged at up to 3000rpm. Next, the "DNeasy 96 plate" is placed onto the S-Block, and the entire solution (approximately 900µl) from the collection microtube plate is transferred into the DNeasy 96 plate. It is then sealed with an Airpore Tape Sheet and centrifuged for 10 minutes at 6000rpm. Following this, 500µl of Buffer AW1 is added to each sample, sealed again, and centrifuged for 5 minutes at 6000rpm. This step is followed by the addition of 500µl of Buffer AW2 to each sample, which is left unsealed, and then centrifuged for 15 minutes at 6000rpm. The DNeasy 96 plate is then placed onto the Elution Microtubes RS. Subsequently, 100µl of Buffer AE is pipetted onto the membrane of each well, sealed with the Airpore Tape Sheet, and incubated for 1 minute at room temperature. The samples are then centrifuged for 2 minutes at 6000rpm. The process of adding 100µl of Buffer AE to the membrane, sealing, incubating, and centrifuging is then repeated. After this series of steps, the extracted DNA is now present in the Elution Microtubes RS and ready for further downstream analysis. After quantification via the BioSpectrometer, the samples were diluted to a working concentration between 20-100 ng/µl.

4.2 PCR1

The PCR1 protocol begins with the preparation of a primer mix, which contains each primer at a concentration of 2µM. The volume required for 72 reactions is 28µL for each primer, leading to a total volume of 2016µL, with no requirement for the addition of molecular grade water. Cycles and master mix preparation shown in Table 4.

Table 4: Showing the sample preparation and the cycling conditions for PCR1 (library preparation)

Primer mix preperation:
2µM per Primer
40µL per Primer (72 times)
0µL water (mol grad)
Total Volume 2880µL

Cycling conditions			
Initial heat activati	15:00 min	95°C	
Denaturation	00:30 min	94°C	
Annealing	01:30 min	57°C	6x
Extension	01:30 min	72°C	
Denaturation	00:30 min	94°C	
Annealing	01:30 min	74°C	20x
Extension	01:30 min	72°C	
Final extension	10:00 min	72°C	

4.3 DNA purification- ExoSAP-IT and Beads clean-up

Following the initial Polymerase Chain Reaction (PCR1), the amplified DNA samples were subjected to an enzymatic clean-up step using Exonuclease-Shrimp Alkaline Phosphatase (ExoSAP-IT). In this process, approximately 2µL of ExoSAP-IT reagent was added to every 10µL of the PCR1 product. Considering an additional 10% for extra measure, around 17.6µL of the reagent was added to each well of a 12-well-strip from where the 2µL were transferred into each well of the plate. The plates were subsequently placed in a PCR thermocycler, and the samples were incubated at 37°C for 15 minutes to allow the degradation of leftover primers and nucleotides. A subsequent incubation step at 80°C for another 15 minutes served to inactivate the ExoSAP-IT reagent. After the ExoSAP-IT clean-up, the PCR products were purified using AMPure XP-Beads. This bead-based purification system allows for the selective binding of DNA, enabling the removal of contaminants such as primers, nucleotides, and enzymes. In preparation of the beads clean-up procedure, a 12-well strip with 169µL AMPure XP-Beads per well, two 96-well plates with 160µL of freshly prepared 70% Ethanol per well and another 96-well plate with 50µL molecular grade water per well were prepared. Following the ExoSAP treatment, the samples were maintained in a 96 well plate with a volume of 12 µl per well. After the beads acclimate to room temperature, they were resuspended by shaking the bottle vigorously. A 1.6x volume or 19.2 µl of beads solution was added to every well, followed by thorough mixing by pipetting up and down for ten times per row. After a 5-minute incubation at room temperature, the reaction plate was placed on an Agencourt SPRIPlate 96 Super Magnet Plate for 2 minutes to separate the beads from the solution. Upon clearing the solution, 5 µL of supernatant was left behind while the rest of the cleared solution was discarded. This was done carefully, not to disturb the separated magnetic beads and to prevent them from drying. While on the magnetic plate, 150 µl of fresh 70% ethanol was added to each well, incubated for 30

seconds, and then carefully removed. This wash step was performed twice. Once the ethanol was fully evaporated, the plate was removed from the magnetic plate. It is again worth noting that the evaporation processes got speed up via a vacuum centrifuge as described before. Further, the elution procedure was enhanced. After the adding 40 µl of molecular grade water to each well, the plate was sealed again, vortexed, incubated for two minutes and briefly centrifuged. The plate was then placed back on the magnetic rack for two minutes to allow the separation of the beads from the DNA solution. Finally, the eluate, approximately 36 µl, was transferred to a new 96 well plate.

4.4 Quality check via Phor Agarose Gel electrophoresis

Post each PCR and subsequent library clean-up, an analysis using 4% Phor agarose gel electrophoresis was implemented to assess the quality and quantity of the DNA samples. The principles and steps are similar to the described agarose 3% gel. For the setup of a 4% Phor-agarose, the initial requirement is TAE buffer cooled down to 4°C. The procedure starts with adding the TAE buffer into an Erlenmeyer beaker. While stirring the buffer at approximately 300 rpm, agarose is gently added into the beaker to ensure its thorough mixing with the TAE buffer. This mixture is then left on the magnetic stirrer to incubate at room temperature for about 10 minutes, allowing the agarose to soak effectively. To account for potential evaporation in the subsequent heating steps, approximately 10g of water is added to the mixture. The buffer-agarose solution is then heated in a microwave set to about 400 W until the agarose is fully dissolved. Regular weighing of the beaker during the heating process helps monitor evaporation losses. It is crucial to ensure that the weight of the beaker doesn't drop below the weight measured before. Post-heating, the beaker is cooled down in lukewarm water for one-minute seconds. The briefly cooled down gel is poured into a tray, and combs are added. The gel is then cooled down at room temperature for about 10 minutes, before it is transferred to a refrigerator and left there for an additional 30 minutes. The gel is then ready for electrophoresis and is run for 70 minutes at a constant voltage of 167.5 V. The PCR products were then loaded into wells in the gel and subjected to an electric field, which caused the negatively charged DNA fragments to move towards the positive electrode. Smaller fragments moved faster and thus farther than larger ones, separating the fragments by size. For these quality check gels the master mixes were prepared as shown in Figure 4. A 100bp equimolar DNA ladder was used as a standard. The DNA fragments of known sizes serve as a reference for determining the size of the PCR products. The gel was then run at 117V (6.7 V/cm) for 70 minutes. After the run, the gel was visualized under UV light, and the position of the DNA fragments in the sample relative to the DNA ladder fragments was used to determine their sizes. These quality control

checks ensured that the DNA fragments were of the correct size and that the DNA samples maintained their integrity throughout the process.

4.5 PCR 2 – Indexing

Prior the indexing, the total volume of the multiplex PCR product is approximately 40 µL, which is further subjected to a 1:10 dilution to facilitate subsequent processes. This dilution results in a total volume of 40 µL per well, providing sufficient material for approximately 19 runs of the PCR2, also known as PCR2. The indexing process is necessary for the differentiation and identification of DNA samples after sequencing. Unique barcode sequences are integrated into the DNA samples during PCR2. These barcodes, added to each sample, act as tags that remain attached to the DNA fragments through the sequencing process. In downstream analyses, the barcodes enable the identification of the origin of the sequenced DNA, allowing for the differentiation of sequences that belong to different samples. The used indexing primer and the preparation of the master mixes is shown in Table 5a-b

Table 5a: The used indexing primer for the plates that contained the samples of this study.

	Plate VIK1 PCR2 BH-LIB.2											
	1	2	3	4	5	6	7	8	9	10	11	12
A	i5_01	i5_09	i5_17	i5_25	i5_33	i5_41	i5_01	i5_09	i5_17	i5_25	i5_33	i5_41
B	i5_02	i5_10	i5_18	i5_26	i5_34	i5_42	i5_02	i5_10	i5_18	i5_26	i5_34	i5_42
C	i5_03	i5_11	i5_19	i5_27	i5_35	i5_43	i5_03	i5_11	i5_19	i5_27	i5_35	i5_43
D	i5_04	i5_12	i5_20	i5_28	i5_36	i5_44	i5_04	i5_12	i5_20	i5_28	i5_36	i5_44
E	i5_05	i5_13	i5_21	i5_29	i5_37	i5_45	i5_05	i5_13	i5_21	i5_29	i5_37	i5_45
F	i5_06	i5_14	i5_22	i5_30	i5_38	i5_46	i5_06	i5_14	i5_22	i5_30	i5_38	i5_46
G	i5_07	i5_15	i5_23	i5_31	i5_39	i5_47	i5_07	i5_15	i5_23	i5_31	i5_39	i5_47
H	i5_08	i5_16	i5_24	i5_32	i5_40	i5_48	i5_08	i5_16	i5_24	i5_32	i5_40	i5_48
i7:	i7_11						i7_12					

	Plate 43 + EXTRAS PCR2 BH-Lib. 2											
	1	2	3	4	5	6	7	8	9	10	11	12
A	i5_01	i5_09	i5_17	i5_25	i5_33	i5_41	i5_01	i5_09	i5_17	i5_25	i5_33	i5_41
B	i5_02	i5_10	i5_18	i5_26	i5_34	i5_42	i5_02	i5_10	i5_18	i5_26	i5_34	i5_42
C	i5_03	i5_11	i5_19	i5_27	i5_35	i5_43	i5_03	i5_11	i5_19	i5_27	i5_35	i5_43
D	i5_04	i5_12	i5_20	i5_28	i5_36	i5_44	i5_04	i5_12	i5_20	i5_28	i5_36	i5_44
E	i5_05	i5_13	i5_21	i5_29	i5_37	i5_45	i5_05	i5_13	i5_21	i5_29	i5_37	i5_45
F	i5_06	i5_14	i5_22	i5_30	i5_38	i5_46	i5_06	i5_14	i5_22	i5_30	i5_38	i5_46
G	i5_07	i5_15	i5_23	i5_31	i5_39	i5_47	i5_07	i5_15	i5_23	i5_31	i5_39	i5_47
H	i5_08	i5_16	i5_24	i5_32	i5_40	i5_48	i5_08	i5_16	i5_24	i5_32	i5_40	i5_48
i7:	i7_09						i7_10					

Table 5b: Sample preparation and cycling conditions for the indexing (PCR2- library preparation)

Master Mix 1:			
	1rec.	48 rec.+17,5%	
2x MM	5	282	
i7_x (2µM)	1	56,4	
Q-Sol.	1	56,4	
H2O	1	56,4	
Cycling conditions			
Initial heat activa	15:00 min	95°C	
Denaturation	00:30 min	94°C	
Annealing	03:00 min	60°C	6x
Extension	0:15 min	72°C	
Denaturation	00:30 min	94°C	
Annealing	03:00 min	68°C	5x
Extension	0:30 min	72°C	
Final extension	10:00 min	72°C	

4.6 Beads Clean-Up and Size Selection (~200 bp) Post Pooling

From every well across all plates 3 µL was drawn using a 12-channel pipette, which was then dispensed into a strip of 12 wells. Each well's content within the strip was mixed by repeated pipetting. The content of the 12 well strip was then transferred into a single 2ml tube. To ensure the full transfer, the tubes were subjected to a brief centrifugation. The resultant pool of samples was put through a vigorous vortex at maximum speed for 2 minutes and then left to incubate on a rotator set at full speed for one hour. Post homogenization, two aliquots, each 100µL in volume, were prepared for final purification. Based on the initial volume, the quantities of AMPure XP bead solution to be added to each aliquot were deduced to be 35µL and 55µL, respectively (0,35x and 0,55x of initial sample volume). The Beads were acclimatized, resuspended and mixed by pipetting, and then left to incubate for 5 minutes at room temperature for optimal recovery. The next step in this procedure is performed on an Agencourt SPRIPlate Super Magnet Rack (for 1.5 ml Eppendorf tubes) for a span of 2 minutes to allow for the separation of the beads from the solution. Once the solution cleared, the supernatant was carefully aspirated from the reaction plate and transferred into a fresh tube, while the beads were discarded. Subsequently, the beads were once again resuspended by shaking, followed by the addition of 0.55 times the initial sample volume to each sample. Again, sufficient mixing was performed by pipetting up and down ten times. After a 5-minute incubation period at room temperature, the reaction plate was placed again on the magnet plate, allowing for the separation of beads from the solution. Once the solution cleared up, the supernatant was discarded, leaving behind the beads with a residual 5 µL of supernatant. While keeping the plate on the magnetic

stand, each well was treated with 1000µl of fresh 70% ethanol and left to incubate for 30 seconds. Following the incubation period, the ethanol was delicately removed, taking care not to disturb the beads. This washing step was performed twice. Post the second wash, the plate was left to dry until all visible traces of ethanol were evaporated. The tube was then removed from the magnetic stand, and 100 µL of 10mM TRIS with 0.1% Tween-20 or Qiagen EB with 0.1% Tween-20 buffer was added to each tube. After mixing by pipetting and two minutes of incubation, the 96 well plate was positioned on the magnetic stand for one minute to segregate the beads from the DNA solution. Finally, the ~100 µl eluate was transferred to a new tube, marking the end of the bead clean-up and size selection process. Afterwards another final quality check via a 4% Agarose gel is performed as described in 4.4. The Gel for the library of which the 112 confirmed individuals were part of can be seen in Figure 4.

Figure 4: showing the 4% Phor Agarose gel and the quality check of the respective libraries. 2-7 are the libraries after purification, 9-11 are the libraries before the clean up as control.

4.7 Sequencing and Quality Check and Trimming

The DNA samples were subjected to Next-Generation Sequencing (NGS) at the Center for Comparative Genomics (CCG) in Cologne. The sequencing was performed on an Illumina MiSeq platform using the MiSeq Micro Kit V2 (300 Cycles) with a paired-end protocol, producing reads of 2x151 base pairs. Before downstream analysis, the raw sequencing data

needed pre-processing to ensure high-quality results. Post Sequencing, the raw reads were accessed through MobaXterm via the CARL/EDDY cluster of the Carl von Ossietzky University in Oldenburg. Using Linux as operating system and SLURM as a job scheduling system, the reads were processed with established scientific modules for trimming, quality control, and mapping. Trimmomatic, a pipeline-based module (Bolger et al., 2014), was utilized for adapter removal and quality filtering of the NGS reads with the following parameters. The following parameters were employed for Trimmomatic:

ILLUMINACLIP: adapter.fasta: 2:30:10

MINLEN: 50

MAXINFO: 40:0.2

The ILLUMINACLIP parameter specifies the adapter file used for adapter removal, allowing for two mismatches (2), a palindromic mode (30), and a simple mode (10). The MINLEN parameter was set to 50, meaning any read that falls below this length after trimming is discarded. The MAXINFO parameter, set at 40:0.2, indicates that the algorithm will use a sliding window approach to trim bases with quality scores below 40, keeping a read if it still has a mean quality above 0.2. To assess the quality of the reads and evaluate the impact of the trimming process, a quality control analysis was conducted using the FastQC module. This analysis was performed both before and after trimming. FastQC provides comprehensive quality checks and generates graphical and tabular reports summarizing read quality and potential errors. After trimming, the processed reads were aligned to a reference which was based on the reconstructed amh and amhby sequences. The alignment was performed using NextGenMap, a module specifically designed to handle highly polymorphic genomes with significant sequence differences between the target and reference (Sedlazeck et al., 2013). NextGenMap accomplishes this by splitting the reference genome into overlapping k-mers, creating a table of positions, and identifying candidate mapping regions for reads based on k-mer matches. The module then selects the best candidate mapping regions based on alignment scores and proceeds with the full alignment to generate a Sequence Alignment Map (SAM-file). This file contains information about the positions and sequences of the aligned reads in relation to the reference genome (Bolger et al., 2014; Sedlazeck et al., 2013). For this analysis the information of the SAM files were sufficient to establish the bioinformatical sexing approach.

5. Genotyping Scripts and Plot explanations

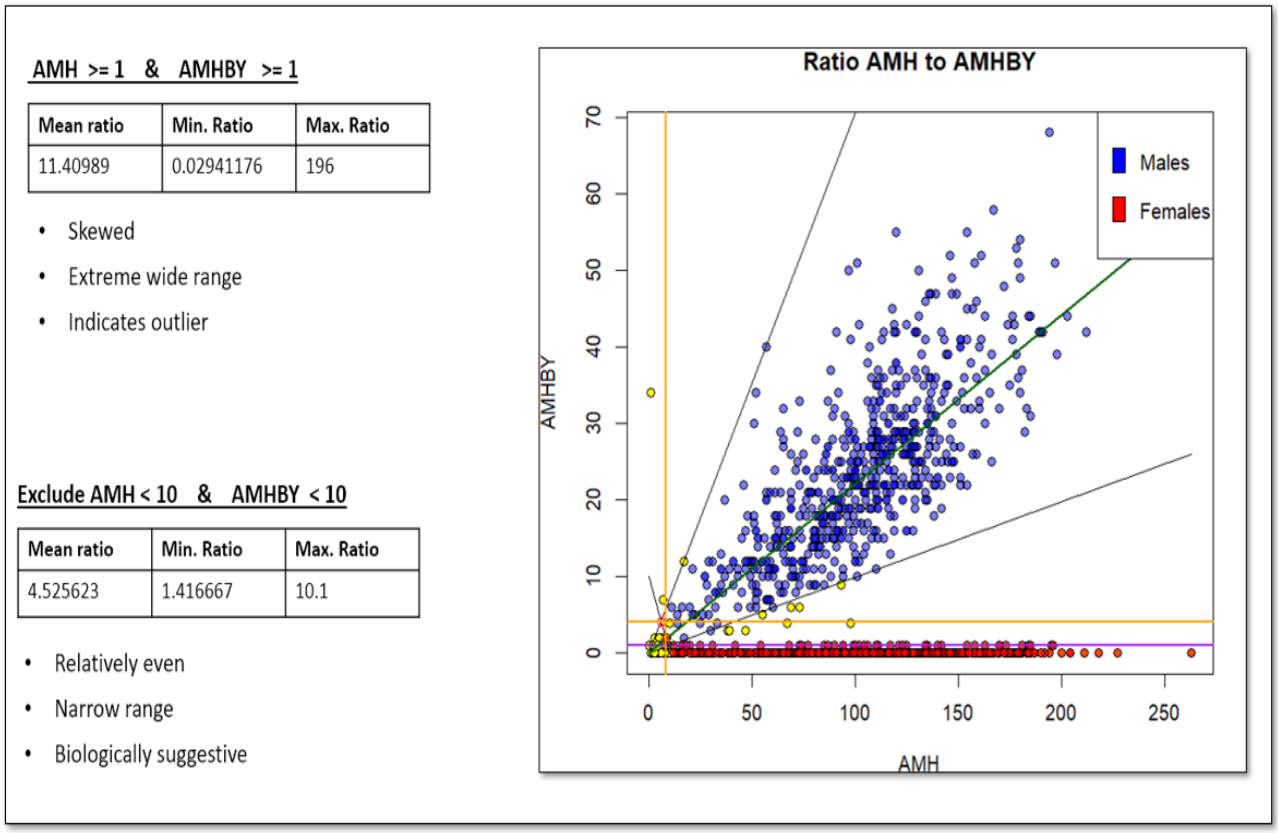
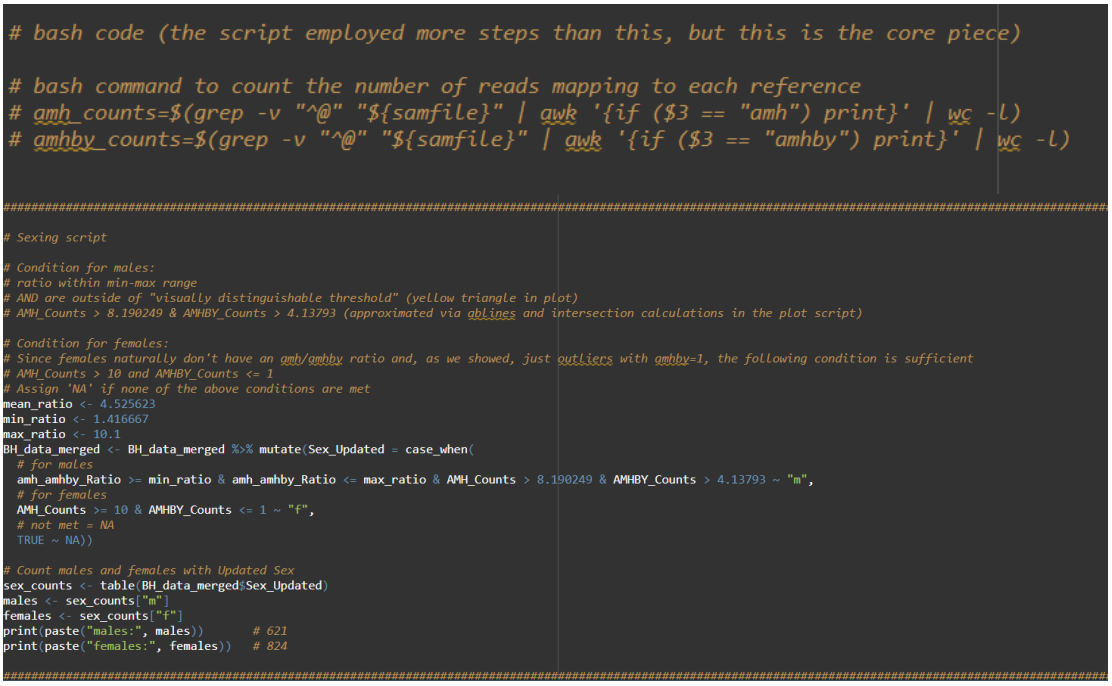


Figure 5a: The explanation of the threshold range and how the individuals in the yellow triangle close to the origin were filtered out. The plot in the right again shows the mean, max and min ratios of amh:amhby in an unfiltered dataset in form of a scatterplot. The table on top left shows the ratios in the unfiltered dataset, the table on the bottom left shows the ratios in the filtered dataset. The plot was used to visualize and eventually define the threshold range. Again, the blue dots represent males, the red dots females, and the yellow dots represent individuals that did not pass the genotyping threshold. The red data cloud parallel to the x-axis represents the females and the individuals above, traversed by the purple line, are the outliers (amhby =1), which were also assigned to the females. The continuous dark-green line indicates the mean ratio, the black lines indicate the ratio range in which individuals are assigned to the male. The ratio range which eventually served as the threshold-range derives from the max and min ratio from a filtered dataset, which was then applied to the unfiltered dataset in form of a ratio-based threshold as shown in Figure 5b, below. This served as final genotyping threshold. The table on top shows the biological unsuggestive ratio range (top), pointing out outliers, the table on the bottom shows the biologically more suggestive range that was then visualized in the plots, and taken as the genotyping threshold. The yellow triangle shown in the Results Fig.2 show the ratio range in which individuals were filtered out as well. This range was calculated based on intersections of the added yellow lines (shown above) with the threshold lines. Based on the intersections with the threshold range lines and the purple line ("female threshold") the area was approximated and converted into an additional threshold, specifically filtering these individuals out. It is notable here that the plot shown in the Results Fig.2 was slightly adjusted for aesthetic reason. The intersection approach is an approximation, and the resulting additional filter threshold would visually not exactly match the yellow triangle shown in the plot in Results Fig.2.



```
# bash code (the script employed more steps than this, but this is the core piece)

# bash command to count the number of reads mapping to each reference
# amh_counts=$(grep -v "^@" "${samfile}" | awk '{if ($3 == "amh") print}' | wc -l)
# amhby_counts=$(grep -v "^@" "${samfile}" | awk '{if ($3 == "amhby") print}' | wc -l)
```

```
################################################################################################################################################

# Sexing script

# Condition for males:
# ratio within min-max range
# AND are outside of "visually distinguishable threshold" (yellow triangle in plot)
# AMH_Counts > 8.190249 & AMHBY_Counts > 4.13793 (approximated via ablines and intersection calculations in the plot script)

# Condition for females:
# Since females naturally don't have an amh/amhby ratio and, as we showed, just outliers with amhby=1, the following condition is sufficient
# AMH_Counts > 10 and AMHBY_Counts <= 1
# Assign 'NA' if none of the above conditions are met
mean_ratio <- 4.525623
min_ratio <- 1.416667
max_ratio <- 10.1
BH_data_merged <- BH_data_merged %>% mutate(Sex_Updated = case_when(
  # for males
  amh_amhby_Ratio >= min_ratio & amh_amhby_Ratio <= max_ratio & AMH_Counts > 8.190249 & AMHBY_Counts > 4.13793 ~ "m",
  # for females
  AMH_Counts >= 10 & AMHBY_Counts <= 1 ~ "f",
  # not met = NA
  TRUE ~ NA))

# Count males and females with Updated Sex
sex_counts <- table(BH_data_merged$Sex_Updated)
males <- sex_counts["m"]
females <- sex_counts["f"]
print(paste("males:", males))       # 621
print(paste("females:", females))   # 824

################################################################################################################################################
```

Figure 5b: Shows the core piece of the bash pipeline to get the amh and amhby counts from the SAM files (top) and the R based sexing pipeline (bottom)

6. Results of the Statistical Analysis of the Sex Ratios

Table 6a: Shows the results of the Pearson's Chi-squared test for all water bodies for each month and for the seasons.

Results of the Chi-squared test and the Nr of fish and the ratios over the year for all woderbodys

In all Month

Month	Fish	Sex Ratio	X-Squared	DF	Probability	P-value
All	1308	1.35	72.25	11	p < 0.005	4.56E-08
January	39	2	4.33	1	p < 0.05	0.03737
February	117	1.34	2.47	1	p > 0.05	0.116
March	487	0.888	1.73	1	p > 0.05	0.1888
April	227	0.802	2.75	1	p > 0.05	0.09705
May	24	0.714	0.67	1	p > 0.05	0.4142
June	64	2.56	12.25	1	p < 0.005	0.0004653
July	61	2.21	8.67	1	p < 0.005	0.003231
August	16	3	4	1	p < 0.05	0.0455
September	42	1.1	0.095	1	p > 0.05	0.7576
October	85	3.47	25.988	1	p < 0.005	3.435e-07
November	79	2.29	12.165	1	p < 0.005	0.0004871
December	67	2.35	10.881	1	p < 0.005	0.0009718

In the Seasons

Season	Fish	Sex Ratio	X-Squared	DF	Probability	P-value
Autumn	206	2.27	31.068	1	< 0,005	2.492e-08
Spring	738	0.854	4.5583	1	> 0,005	0.03276
Summer	141	2.44	24.688	1	< 0,005	6.74e-07
Winter	223	1.69	14.57	1	< 0,005	0.0001351

Table 6b: Shows the results of the Pearson's Chi-squared test for fresh and brackish water catches separately. The months have been combined to seasons.

For Freshwater

Season	Fisch	Sex Ratio	X-Squared	DF	Probability	P-value
All	334	0.885	7.6941	3	p > 0.05	0.05277
Autumn	15	2	1.6667	1	p > 0.05	0.1967
Spring	277	0.787	3.9314	1	p < 0.05	0.04739
Summer	34	1.83	2.9412	1	p > 0.05	0.08635
Winter	8	0.789	0.47059	1	p > 0.05	0.4927

For Brackish Water

Season	Fisch	Sex Ratio	X-Squared	DF	Probalility	P-value
All	948	1.41	47.276	3	< 0.005	3.037e-10
Autumn	191	2.29	29.45	1	< 0.005	5.737e-08
Spring	461	0.897	1.3557	1	> 0.05	0.2443
Summer	107	2.69	22.439	1	< 0.005	2.169e-06
Winter	189	1.95	19.688	1	< 0.005	9.118e-06

Table 6c: Shows the results of the Pearson's Chi-squared test for the different genotype cluster. The months have been combined to seasons.

Sex Ratios within the Genotype Cluster

All Cluster - Total Sex Ratios

Cluster	Total	Female	Male	Sex Ratio	Probability
Blue	358	185	173	1.07	< 0.05
Red	230	128	102	1.25	> 0.05
Red+Blue	588	313	275	1.13	< 0.005
Green	256	160	96	1.67	> 0.05
Purple	144	79	65	1.22	> 0.05
Redblue	322	186	136	1.37	< 0.005
Greenpurple	113	71	42	1.69	> 0.05

Blue Cluster

Season	Fisch	Sex Ratio	X-Squared	DF	Probalility	P-value
All	326	1.0245	25.064	3	< 0.005	1.497e-05
Spring	173	0.617	9.7168	1	< 0.05	0.001826
Summer	43	1.87	3.9302	1	< 0.05	0.04743
Autumn	67	2.35	10.881	1	< 0.05	0.0009718
Winter	44	1.32	0.5814	1	> 0.05	0.4458

Redblue Cluster

Season	Fisch	Sex Ratio	X-Squared	DF	Probalility	P-value
All	305	1.310	24.956	3	< 0.05	1.577e-05
Spring	149	0.753	2.9597	1	> 0.05	0.08536
Summer	29	3.83	9.9655	1	< 0.05	0.001595
Autumn	62	2.65	12.645	1	< 0.05	0.0003765
Winter	65	1.71	4.4462	1	< 0.05	0.03498

Red + Blue (combined)

Season	Fisch	Sex Ratio	X-Squared	DF	Probalility	P-value
All	544	1.09	35.821	3	< 0.005	8.17e-08
Spring	275	0.657	11.815	1	< 0.005	0.0005877
Summer	71	2.23	10.268	1	< 0.005	0.001354
Autumn	107	1.74	7.8598	1	< 0.005	0.005055
Winter	91	1.76	6.8681	1	< 0.05	0.008775

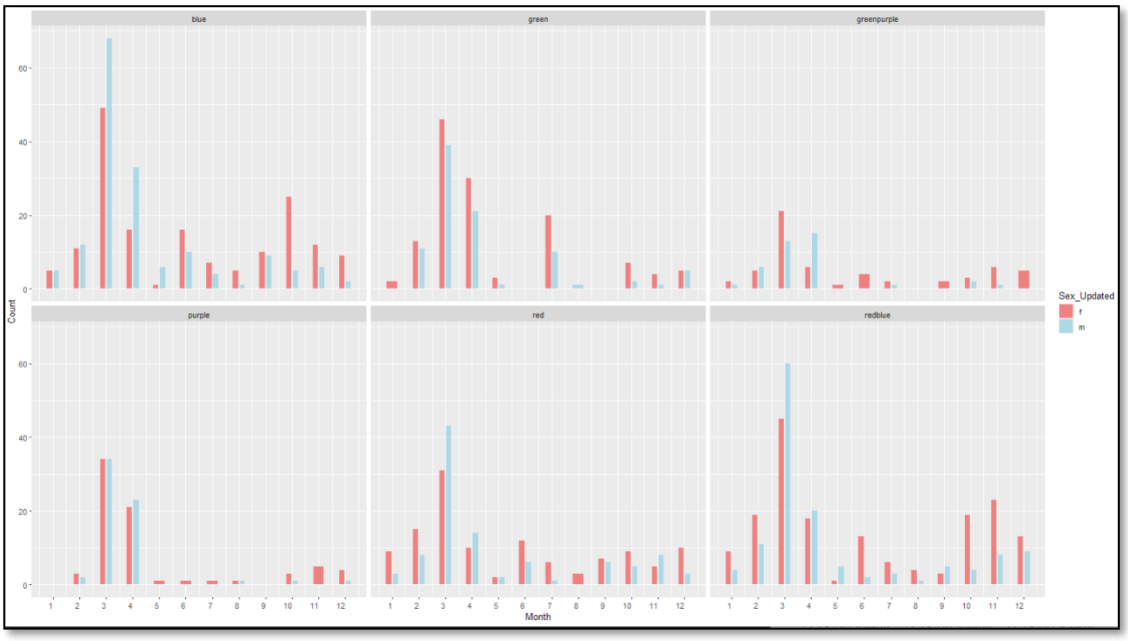


Figure 6: Showing the plots of the counts for males and females over the year for all genotype cluster.

7. Documentation - DNA concentration and dilution

Table 7: Showing all samples of the confirmed female and male pool. Including the Fish ID, the measured DNA concentration, and the absorption parameter as well as the dilution factor and the plate coordinates for the sequencing run

Fish_ID	Con.Mean ng/µL	A260/A280	A260/A230	Dilution	Plate Pos.
BH-91565	547,2	1,87	2,35	1/6	V1/A01
BH-93134	688,2	1,86	2,37	1/7	V1/B01
BH-93164	171,9	1,86	2,37	1/2	V1/C01
BH-93135	570,6	1,87	2,38	1/6	V1/D01
BH-93171	256,8	1,91	2,27	1/3	V1/E01
BH-93185	187,3	1,92	2,25	1/2	V1/F01
BH-93187	166,4	1,86	2,44	1/2	V1/G01
BH-93162	335,1	1,9	2,01	1/4	V1/H01
BH-93183	213	1,93	2,32	1/3	V1/A02
BH-93182	183,3	1,92	2,27	1/2	V1/B02
BH-93154	129,3	1,89	3,54	1/2	V1/C02
BH-93147	117,5	1,88	3,98	1/2	V1/D02
BH-92994	134,6	1,86	2,22	1/2	V1/E02
BH-92995	100,2	1,84	2,08	1/2	V1/F02

BH-93034	104,8	1,88	2,08	1/2	V1/G02
BH-93056	151,8	1,86	2,11	1/2	V1/H02
BH-93144	283,7	1,87	2,77	1/3	V1/A03
BH-93093	183,1	1,85	1,7	1/2	V1/B03
BH-93097	146,9	1,85	1,69	1/2	V1/C03
BH-93059	186,6	1,85	2,16	1/2	V1/D03
BH-93060	86,7	1,85	2,13	0	V1/E03
BH-93088	34,8	1,81	1,37	0	V1/F03
BH-93159	239,6	1,87	2,8	1/3	V1/G03
BH-93156	133,4	189	3,97	1/2	V1/H03
BH-93195	164,6	1,9	2,41	1/2	V1/A04
BH-93196	182,1	1,9	2,33	1/2	V1/B04
BH-93100	145,4	1,86	1,9	1/2	V1/C04
BH-93107	131,1	1,86	3,73	1/2	V1/D04
BH-93136	125,5	1,87	3,69	1/2	V1/E04
BH-93163	125,5	1,87	3,69	1/2	V1/F04
BH-93168	311,3	1,9	2,34	1/4	V1/G04
BH-93193	370	1,92	2,31	1/4	V1/H04
BH-93157	155	1,88	3,33	1/2	V1/A05
BH-93114	209,4	1,87	3	1/3	V1/B05
BH-93122	96	1,85	5	0	V1/C05
BH-93118	184,8	1,87	2,93	1/2	V1/D05
BH-93130	87,2	1,87	6,21	0	V1/E05
BH-93133	151,3	1,81	3,47	1/2	V1/F05
BH-93186	67,2	1,93	2,45	0	V1/G05
BH-93176	277,7	1,87	2,53	1/3	V1/H05
BH-93111	175	1,88	2,26	1/2	V1/A06
BH-93149	95,4	1,89	4	0	V1/B06
BH-93184	149,4	1,91	2,43	1/2	V1/C06
BH-93170	178	1,89	2,41	1/2	V1/D06
BH-93179	198,9	1,9	2,38	1/2	V1/E06
BH-93167	223,9	1,91	2,32	1/3	V1/F06
BH-93177	225,8	1,9	2,29	1/3	V1/G06
BH-93188	131,8	1,91	2,42	1/2	V1/H06
BH-93191	206,2	1,91	2,15	1/3	V1/A07
BH-93166	90,3	1,91	2,59	0	V1/B07
BH-93175	135,1	1,9	2,64	1/2	V1/C07
BH-93061	175,1	1,9	2,29	1/2	V1/D07
BH-93174	98	1,92	2,12	0	V1/E07

BH-93160	122,6	1,89	3,65	1/2	V1/F07
BH-93165	146	1,87	3,11	1/2	V1/G07
BH-93148	133,1	1,9	2,33	1/2	V1/H07
BH-93131	318	1,88	2,7	1/4	V1/A08
BH-93051	576,4	1,88	2,43	1/6	V1/B08
BH-93109	551,7	1,88	2,35	1/6	V1/C08
BH-93116	609	1,88	2,43	1/7	V1/D08
BH-93007	302	1,87	2,23	1/4	V1/E08
BH-93015	289,8	1,89	2,21	1/3	V1/F08
BH-93021	336	1,87	2,29	1/4	V1/G08
BH-93039	304,5	1,86	2,34	1/4	V1/H08
BH-93041	429,5	1,88	2,31	1/5	V1/A09
BH-93044	249,6	1,86	2,29	1/3	V1/B09
BH-93057	262,6	1,89	2,26	1/3	V1/C09
BH-93045	315,2	1,87	2,2	1/4	V1/D09
BH-93074	224,3	1,89	2,29	1/3	V1/E09
BH-93084	239,9	1,87	2,34	1/3	V1/F09
BH-93099	301,1	1,87	2,3	1/4	V1/G09
BH-93108	269,1	1,87	2,21	1/3	V1/H09
BH-93117	229,1	1,87	2,31	1/3	V1/A10
BH-93121	269,1	1,87	2,21	1/3	V1/B10
BH-93127	304,2	1,87	2,21	1/4	V1/C10
BH-93126	376,1	1,91	2,22	1/4	V1/D10
BH-93125	471,4	1,87	2,26	1/5	V1/E10
BH-93132	308,7	1,87	2,26	1/4	V1/F10
BH-93140	196,4	1,88	2,19	1/2	V1/G10
BH-93141	229,5	1,85	2,3	1/3	V1/H10
BH-93143	276,5	1,88	2,21	1/3	V1/A11
BH-93155	534,6	1,88	2,25	1/6	V1/B11
BH-93152	540,6	1,88	2,24	1/6	V1/C11
BH-93169	394,4	1,89	2,26	1/4	V1/D11
BH-93178	227	1,86	2,14	1/3	V1/E11
BH-93172	304	1,87	2,31	1/4	V1/F11
BH-93173	400	1,87	2,3	1/5	V1/G11
BH-93180	198,4	1,87	2,22	1/2	V1/H11
BH-93190	295,4	1,87	2,14	1/3	V1/A12
BH-92990	374,6	1,88	2,25	1/4	V1/B12
BH-91789	396,1	1,87	2,28	1/4	V1/C12
BH-91765	287,9	1,87	2,31	1/3	V1/D12

BH-93013	404,6	1,89	2,27	1/5	V1/E12
BH-93002	413,8	1,89	2,32	1/5	V1/F12
BH-92998	507,6	1,88	2,22	1/6	V1/G12
BH-92991	434	1,89	2,34	1/5	V1/H12
BH-92996	352,1	1,88	2,26	1/4	Mix Plate
BH-92993	402,1	1,88	2,3	1/5	Mix Plate
BH-51246	533,2	1,9	2,37	1/6	Mix Plate
BH-91549	220,6	1,9	2,12	1/3	Mix Plate
BH-91548	430,3	1,89	2,29	1/5	Mix Plate
BH-91534	238,4	1,89	2,3	1/3	Mix Plate
BH-90546	445,6	1,89	2,32	1/5	Mix Plate
BH-51614	566	1,89	2,36	1/6	Mix Plate
BH-51397	575,1	1,9	2,35	1/6	Mix Plate
BH-93005	388,1	1,85	2,39	1/4	Mix Plate
BH-93003	449,4	1,88	2,33	1/5	Mix Plate
BH-93046	442,3	1,88	2,29	1/5	Mix Plate
BH-93006	406,3	1,87	2,26	1/5	Mix Plate
BH-91585	515,8	1,86	2,32	1/6	Mix Plate
BH-93048	644,6	1,88	2,3	1/7	Mix Plate
BH-93047	352,1	1,87	2,32	1/4	Mix Plate

8: Documentation of the amh and amhby counts of the 112 confirmed individuals

Table 8: The initial 112 individuals after genotyping. Showing the ID, the morphological sex, the sex based on the PCR and the sex based on the sequencing, the counts for the two loci and the ratio. The approach for the thresholds as described in Methods 2.5 and Appendix 5 Figure 5a was based on this dataset and was then applied to all individuals for which the sequencing resulted in amh and amhby reads (depending on if it is a male or female obviously). As described before, the min and max ratio after filtering everything <10 was used to define a reasonable threshold ratio, which was then applied to the whole dataset without previously filtering out individuals. The resulting threshold is visualized in Appendix 5 Figure 5a Methods 2.5. Only individuals that do not fall into this range are filtered out. The Outliers (amhby=1) in this table are still assigned to "male" which was then adjusted later in the downstream analysis as described before and were eventually assigned as females

Sample	Morph. Sex	PCR Sex	NGS Sex	Amh	Amhby	Amh/Amhby
BH-51614	F	F	F	137	0	N/A
BH-91585	F	F	F	77	0	N/A
BH-93013	F	F	F	139	0	N/A
BH-93047	F	F	F	128	0	N/A
BH-93048	F	F	F	49	0	N/A
BH-93084	F	F	F	115	0	N/A
BH-93099	F	F	F	140	0	N/A
BH-93108	F	F	F	202	0	N/A

BH-93117	F	F	F	150	0	N/A
BH-93132	F	F	F	133	0	N/A
BH-93140	F	F	F	198	0	N/A
BH-93155	F	F	F	135	0	N/A
BH-51246	F	F	F	111	0	N/A
BH-51397	F	F	F	119	0	N/A
BH-93141	F	F	F	160	0	N/A
BH-91765	F	F	F	120	0	N/A
BH-91789	F	F	F	22	0	N/A
BH-91548	F	F	F	127	0	N/A
BH-93021	F	F	F	119	0	N/A
BH-93039	F	F	F	123	0	N/A
BH-93041	F	F	F	139	0	N/A
BH-93044	F	F	F	108	0	N/A
BH-93045	F	F	F	153	0	N/A
BH-93046	F	F	F	120	0	N/A
BH-93057	F	F	F	141	0	N/A
BH-93125	F	F	F	153	0	N/A
BH-93127	F	F	F	175	0	N/A
BH-90546	F	F	F	131	0	N/A
BH-91534	F	F	F	81	0	N/A
BH-92991	F	F	F	195	0	N/A
BH-92993	F	F	F	89	0	N/A
BH-92996	F	F	F	105	0	N/A
BH-93002	F	F	F	108	0	N/A
BH-93003	F	F	F	32	0	N/A
BH-93005	F	F	F	34	0	N/A
BH-93015	F	F	F	79	0	N/A
BH-93121	F	F	F	85	0	N/A
BH-93143	F	F	F	153	0	N/A
BH-93152	F	F	F	153	0	N/A
BH-93172	F	F	F	103	0	N/A
BH-93173	F	F	F	187	0	N/A
BH-93180	F	F	F	152	0	N/A
BH-93190	F	F	F	150	0	N/A
BH-93116	F	M	M	112	18	6,222222222
BH-91549	F	F	F	1	0	N/A
BH-93169	F	F	F	7	0	N/A
BH-92990	F	F	M	100	1	100
BH-93007	F	F	M	104	1	104
BH-93178	F	F	M	136	1	136
BH-93074	F	F	M	137	1	137
BH-92998	F	F	M	149	1	149
BH-93006	F	F	M	55	1	55
BH-93126	F	F	M	3	1	3
BH-93060	M	M	M	140	36	3,888888889

BH-93093	M	M	M	130	26	5
BH-93097	M	M	M	129	37	3,486486486
BH-93109	M	M	M	85	14	6,071428571
BH-93114	M	M	M	122	26	4,692307692
BH-93122	M	M	M	160	33	4,848484848
BH-93130	M	M	M	119	30	3,966666667
BH-93131	M	M	M	165	39	4,230769231
BH-93133	M	M	M	93	17	5,470588235
BH-93136	M	M	M	133	25	5,32
BH-93147	M	M	M	158	24	6,583333333
BH-93148	M	M	M	148	30	4,933333333
BH-93154	M	M	M	130	28	4,642857143
BH-93100	M	M	M	124	25	4,96
BH-93059	M	M	M	119	28	4,25
BH-93061	M	M	M	120	26	4,615384615
BH-92994	M	M	M	160	26	6,153846154
BH-93034	M	M	M	128	28	4,571428571
BH-93056	M	M	M	134	31	4,322580645
BH-93088	M	M	M	95	20	4,75
BH-93107	M	M	M	131	25	5,24
BH-93135	M	M	M	90	14	6,428571429
BH-93144	M	M	M	125	22	5,681818182
BH-93149	M	M	M	85	16	5,3125
BH-93156	M	M	M	137	34	4,029411765
BH-93157	M	M	M	150	34	4,411764706
BH-93159	M	M	M	134	30	4,466666667
BH-93162	M	M	M	157	38	4,131578947
BH-93163	M	M	M	151	27	5,592592593
BH-93165	M	M	M	136	28	4,857142857
BH-93166	M	M	M	158	32	4,9375
BH-93167	M	M	M	127	22	5,772727273
BH-93168	M	M	M	127	20	6,35
BH-93170	M	M	M	162	38	4,263157895
BH-93171	M	M	M	146	26	5,615384615
BH-93174	M	M	M	143	34	4,205882353
BH-93175	M	M	M	145	25	5,8
BH-93176	M	M	M	117	25	4,68
BH-93177	M	M	M	123	29	4,24137931
BH-93179	M	M	M	95	26	3,653846154
BH-93182	M	M	M	93	11	8,454545455
BH-93183	M	M	M	127	23	5,52173913
BH-93184	M	M	M	101	23	4,391304348
BH-93185	M	M	M	143	28	5,107142857
BH-93186	M	M	M	151	23	6,565217391
BH-93187	M	M	M	142	29	4,896551724

BH-93188	M	M	M	149	38	3,921052632
BH-93191	M	M	M	118	21	5,619047619
BH-93193	M	M	M	141	46	3,065217391
BH-93195	M	M	M	142	34	4,176470588
BH-93196	M	M	M	111	24	4,625
BH-93111	M	M	F	11	0	N/A
BH-93164	M	F	M	111	19	5,842105263
BH-91565	M	F	F	50	0	N/A
BH-93134	M	F	F	101	0	N/A
BH-92995	M	M	M	36	5	7,2
BH-93160	M	M	M	82	9	9,111111111
BH-93118	M	M	M	8	1	8
BH-93051	M	M	M	42	5	8,4