L'INSTITUT agro Rennes Angers

L'Institut Agro Rennes-Angers
☐ Site d'Angers ☒ Site de Rennes

Logo(s) des établissements en co-habilitation, de l'école partenaire accueillant l'étudiant en année de spécialisation, ou de l'école d'origine de l'étudiant

*Logo de la structure d'accueil**

INRAE

Année universitaire : 2024 - 2025 Spécialité : Agronome Spécialisation (et option éventuelle) : SHA (Sciences Halieutiques et Aquacoles) – Option AQUAculture	**Mémoire de fin d'études** ☒ d'ingénieur de l'Institut Agro Rennes-Angers (Institut national d'enseignement supérieur pour l'agriculture, l'alimentation et l'environnement) ☐ de master de l'Institut Agro Rennes-Angers (Institut national d'enseignement supérieur pour l'agriculture, l'alimentation et l'environnement) ☐ de l'Institut Agro Montpellier (étudiant arrivé en M2) ☐ d'un autre établissement (étudiant arrivé en M2)

Diversité génétique de la lignée synthétique et signatures de sélection des lignées grasse et maigre INRAE de truite arc-en-ciel (*Oncorhynchus mykiss*)

Par : Gatien DEBAILLEUL

Soutenu à Rennes ***le 10/09/2025***

Devant le jury composé de :

Président : Bastien SADOUL
Maître de stage : Florence PHOCAS
Enseignant référent : Bastien SADOUL

Autres membres du jury (Nom, Qualité)
Hervé LE BRIS (Enseignant chercheur)
Frédéric HERAULT (Ingénieur de recherche)

Les analyses et les conclusions de ce travail d'étudiant n'engagent que la responsabilité de son auteur et non celle de l'Institut Agro Rennes-Angers

Acknowledgements

I would like to personally thank my two supervisors: Florence PHOCAS and Simon POUIL, for accepting me as an intern on their team. I have learned a lot about rainbow trout genetics during our meetings and discussions. Thanks to their support I developed my knowledge in different research themes, like genetic diversity or the detection of signatures of selection, fields I didn't know much when I arrived. This internship was also an opportunity to acquire a solid foundation in bioinformatics.

I would also like to thank the two of you for the time you dedicated to me, for your corrections and your tips all internship long. By doing this internship, I strengthened my choice of continuing in the research field in aquaculture.

I would also like to thank all the GUPPIE team members for the time spent together and the discussions on a panel of different interesting subjects. Thank you for integrating me so well into this team and for showing me a new side of research.

I would like to thank Sophie for helping me in so many ways. Thank you for your support, your joy, your tips, your jokes and the breaks when we both needed them. I hope your PhD will continue to bring you pride and that it will go well.
Special thanks to Audrey who helped me with the running of the Mongrail code.
Thank you to Michaël, Auguste, Valentin, Delphine, Tristan, Kamenan and Gabriel for making the GUPPIE team such a great team to work with.

Thank you to UMR GABI, that groups passionate and inspiring people that are ready to share any knowledge when asking. Special thanks to Mathilde and Fabienne, our heads of department, for all the work they are doing for us, and for dealing with tricky situations.

Finally, I would like to thank the PEIMA team for responding to our questions, but also for accepting me in such a well-located facility for a week so I could help them with manual work.

Fiche de confidentialité et de diffusion du mémoire

Confidentialité

■ Non □ Oui si oui : □ 1 an □ 5 ans □ 10 ans

Pendant toute la durée de confidentialité, aucune diffusion du mémoire n'est possible [1].

Date et signature du **maître de stage** [2] : 27/08/2025
(ou de l'étudiant-entrepreneur)

A la fin de la période de confidentialité, sa diffusion est soumise aux règles ci-dessous (droits d'auteur et autorisation de diffusion par l'enseignant à renseigner).

Droits d'auteur

L'auteur[3] **DEBAILLEUL Gatien**
autorise la diffusion de son travail (immédiatement ou à la fin de la période de confidentialité)
■ Oui □ Non

Si oui, il autorise

☐ la diffusion papier du mémoire uniquement(4)

☐ la diffusion papier du mémoire et la diffusion électronique du résumé

■ la diffusion papier et électronique du mémoire (joindre dans ce cas la fiche de conformité du mémoire numérique et le contrat de diffusion)

(Facultatif) ☐ accepte de placer son mémoire sous licence Creative commons CC-By-Nc-Nd (voir Guide du mémoire Chap 1.4 page 6)

Date et signature de l'**auteur** :

Le 22/08/2025

Autorisation de diffusion par le responsable de spécialisation ou son représentant

L'enseignant juge le mémoire de qualité suffisante pour être diffusé (immédiatement ou à la fin de la période de confidentialité)

☒ Oui ☐ Non

Si non, seul le titre du mémoire apparaîtra dans les bases de données.

Si oui, il autorise

☐ la diffusion papier du mémoire uniquement(4)

☐ la diffusion papier du mémoire et la diffusion électronique du résumé

☒ la diffusion papier et électronique du mémoire

Date et signature de l'**enseignant** :

(1) L'administration, les enseignants et les différents services de documentation de l'Institut Agro Rennes-Angers s'engagent à respecter cette confidentialité.
(2) Signature et cachet de l'organisme
(3).Auteur = étudiant qui réalise son mémoire de fin d'études
(4) La référence bibliographique (= Nom de l'auteur, titre du mémoire, année de soutenance, diplôme, spécialité et spécialisation/Option)) sera signalée dans les bases de données documentaires sans le résumé

Utilisation de l'Intelligence Artificielle (IA) pour la rédaction du mémoire

L'intelligence artificielle a été utilisée pour améliorer le texte ☒ Oui ☐ Non

Si oui, préciser pour quelle partie de texte et à quelles fins ?

L'intelligence artificielle a été utilisée ponctuellement comme un outil d'assistance. Tout d'abord à des fins de relecture et de correction de la langue (vocabulaire scientifique, clarté de la formulation, correction grammaticale) dans la discussion et la conclusion (parties 4 et 5). L'intelligence artificielle a également été utilisé pour les corrections mineurs de script (R et bash) en cas de message d'erreurs, touchant ainsi aux résultats de la partie 3.

Toute absence de déclaration précise sera considérée comme un manquement aux règles de l'éthique scientifique.

Date et Signature de l'**auteur** :  Le 22/08/2025

List of figures

List of tables

List of annexes

List of abbreviations

DNA – Deoxyribonucleic Acid

EU – European Union

FAO – Food and Agriculture Organization of the United Nations

F_{ROH} – Inbreeding coefficient based on Runs of Homozygosity

F_{ST} – Fixation Index (measure of population differentiation)

GO – Gene Ontology

IBD – Identity By Descent

IBS – Identity By State

INRAE – French National Research Institute for Agriculture, Food and Environment

MAF – Minor Allele Frequency

Ne – Effective population size

PCA – Principal Component Analysis

PC1, PC2 – Principal Components 1 and 2

PEIMA – INRAE Experimental Fish Farming Systems Facility (Sizun, France)

ROH – Run of Homozygosity

SNP – Single Nucleotide Polymorphism

TABLE OF CONTENT

1. Introduction

a. Rainbow trout: Aquaculture and selective breeding

The rainbow trout (*Oncorhynchus mykiss*) is a species native to the Pacific drainages of North America and the Kamchatka Peninsula in Russia. It was introduced in Europe and other parts of the world in the late 19th century to support the expansion of freshwater aquaculture. Thanks to its adaptability to different environments, rapid growth rate, and high consumer acceptance, rainbow trout has become one of the most widely farmed salmonids globally. In France, rainbow trout dominates fish farming, accounting for 79% of national production in 2022, amounting to 32,310 tons (Agreste 2024). The Food and Agriculture Organization (FAO) estimates that French annual production is exceeding 30,000 tons each year (FAO 2022). France maintains a leading role in European production among EU member states, accounting alone for about 20% of the European Union's production in 2019 (Norway, far ahead, not included), with Italy and the emerging Poland also playing an increasingly important role (Directorate General for Maritime Affairs and Fisheries (European Commission), EUMOFA 2021). The rainbow trout (*Oncorhynchus mykiss*) benefits from a long history of domestication and has been the focus of dedicated breeding programs in France since the 1990s (Haffray et al. 2004). Early selection efforts primarily targeted growth performance. Over time, selection objectives have expanded to include additional traits of interest, such as carcass yield, body morphology, and resistance to infectious diseases (Boudry et al. 2021). This evolution has been driven by advancements in animal identification technologies (e.g., PIT-tagging), phenotyping protocols, and genotyping tools, which have significantly improved the precision and efficiency of selection programs. Nevertheless, the success of selective breeding relies heavily on the effective management of genetic diversity. Without maintaining sufficient genetic variability, long-term selection becomes less effective and populations risk losing their adaptive potential in the face of changing environmental and production conditions.

b. Managing diversity and development of genomic tools

Significant advances in rainbow trout breeding have been made possible through continuous scientific and zootechnical efforts, alongside the evolution of selection strategies. Historically, breeding programs moved from phenotypic selection to selection based on pedigree records, and now to genomic selection. While early breeding programs relied primarily on phenotypic selection, later complemented by genealogical information, the availability of a high-quality reference genome has enabled the development of more accurate genomic selection approaches. These methods now make it possible to improve traits of interest that cannot be directly measured on selection candidates, such as disease resistance or fillet quality.

However, the long-term success of selective breeding depends on maintaining sufficient genetic diversity within broodstock populations. Without this, the adaptive capacity of farmed populations to future environmental and climatic changes is compromised. Aquaculture practices often tend to reduce genetic diversity and increase inbreeding (Sekino, Hara, Taniguchi 2002; Lundrigan, Reist, Ferguson 2005; Wang et al. 2012). As inbreeding increases, so does homozygosity, which raises the likelihood of alleles being identical by descent and the expression of deleterious recessive variants (Charlesworth, Charlesworth 1999). Monitoring inbreeding levels is therefore a key component of

genetic management in aquaculture. One approach to limit inbreeding is to incorporate cryopreserved genetic material into breeding programs (Jacques et al. 2023), thereby reintroducing lost alleles and reducing the risk of fixation.

In this context, experimental populations with high genetic diversity are valuable resources. They enable the study of the genetic basis of traits and provide a genetic reservoir that can enhance adaptability in the face of evolving farming conditions, contributing to the long-term sustainability of aquaculture.

Recent decades have also seen a drastic reduction in the costs of genotyping and sequencing, making genomic resources more accessible for aquaculture species, particularly salmonids. The availability of a high-quality rainbow trout genome assembly (Gao et al. 2021), detailed genetic maps (Guyomard et al. 2012), and a medium-density SNP array containing 57,000 markers (Palti et al. 2015) has opened new perspectives for genetic management. These tools allow for precise measurement of genetic diversity (Paul 2022), detection and monitoring of inbreeding (D'Ambrosio et al. 2019), and the identification of selection signatures resulting from both artificial and natural pressures (Cádiz et al. 2021; Paul 2022).

c. Application of genomic tools to experimental rainbow trout lines

This study used genomic resources to address two major research questions concerning experimental rainbow trout lines developed by INRAE geneticists in the 1980s and 1990s. These lines are maintained at the PEIMA experimental facility in Sizun (Finistère), a key site for research in genetics, nutrition, immunology, and reproduction. Another experimental site is located in the south of France, dedicated to rearing the same rainbow trout lines. This site, called Lees-Athas, serves as a mirror facility. In practice, the lines maintained at PEIMA are also kept there, though in smaller groups. Unlike at PEIMA, these fish are not used for experiments; instead, the site provides a safeguard, ensuring the preservation of the lines in case an incident at PEIMA.

Objective 1 - preserving genetic diversity in the INRAE synthetic line

The INRAE synthetic line was established in the early 1980s by crossing breeders from French, European, and American rainbow trout farms to maximize genetic variability. Since its creation, it has been maintained without any external genetic input or deliberate selection pressure to preserve its genetic diversity (D'Ambrosio et al. 2019). To safeguard this resource, sperm from the 2010 and 2012 cohorts was cryopreserved for potential use in future experiments and genetic analyses.

Over the past decade, this line has faced two major mortality events at PEIMA. In 2018, a chemical release in the flow system caused the loss of most of the fish. In 2023, an outbreak of lactococcosis, a bacterial disease causing acute infection (Shahin et al. 2022), led to the culling of all fish to prevent its spread and to sanitize the facility. Recovery was possible thanks to the mirror INRAE unit at Lees-Athas (Pyrénées). Fish were transferred back to PEIMA in 2019 and again in 2024, but each time only a subset of the stock could be moved, raising concerns over potential losses of genetic diversity in the reconstituted cohorts.

In 2019, cryopreserved sperm from earlier cohorts was incorporated into the breeding plan to produce the 2020 progeny. Although this strategy was not based on targeted genetic selection, subsequent

analyses confirmed its usefulness for maintaining diversity (Figure 1). A preliminary study in 2022 (Colon 2022) suggested that cryopreserved material could again be used to boost diversity in the 2021 and 2022 cohorts, but the resulting fish could not be genotyped before the 2023 outbreak. In the present work, we assessed the genetic diversity of the 2023 synthetic cohort transferred from Lees-Athas and evaluated whether reintroducing cryopreserved sperm from the 2010 and 2012 cohorts could help restore diversity for the next generation.

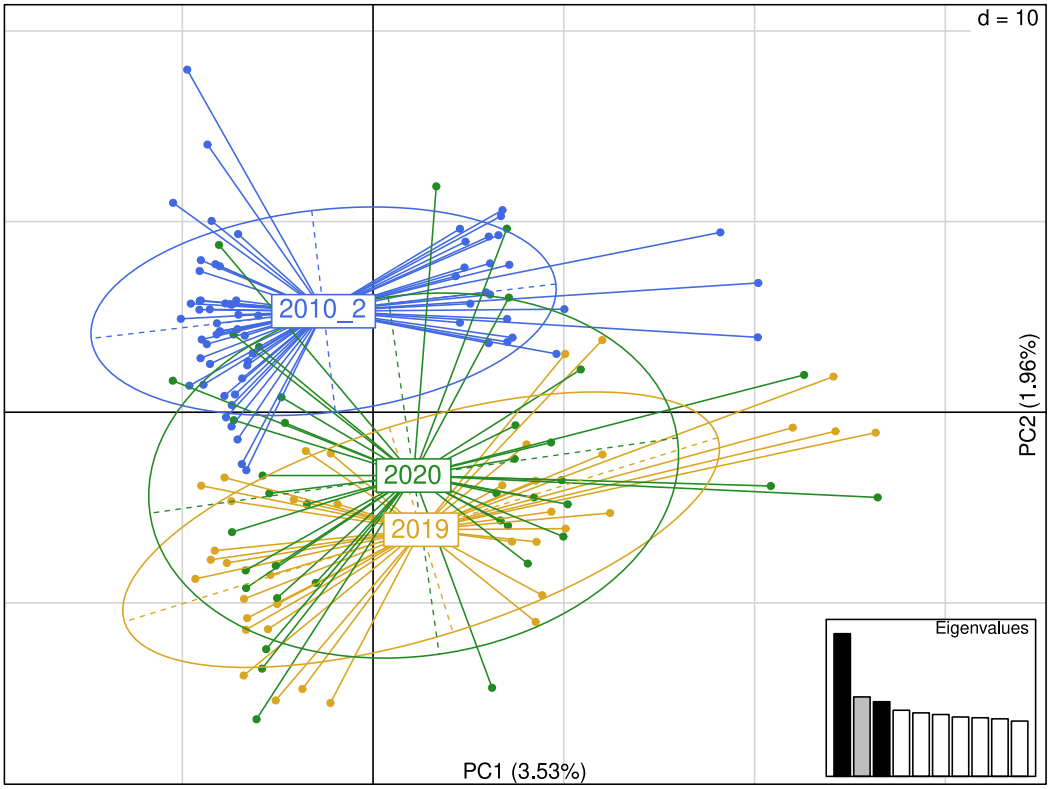

Figure 1: Principal Component Analysis (PCA) illustrates the distribution of individuals from three cohorts of the synthetic line ("2010_2" are cryoconserved straws, "2019" are the genotyped individuals from the cohort imported from the mirror site and "2020" is the progeny cohort resulting from the breeding of the two previous stocks presented)

Objective 2 – Detecting signatures of selection in the INRAE divergent fat and lean lines

The second objective focused on two experimental lines derived from the same INRAE spring-spawning population, which have undergone divergent selection for muscle fat content over eight generations (Quillet et al. 2005). By the seventh generation, muscle fat content was approximately three times higher in the fat line than in the lean line (Lefèvre et al. 2024).

Muscle fat content is measured at each selection cycle using a Fish Torry Fat-meter (Douirin et al. 1998), a non-destructive device that estimates lipid content by measuring the dielectric properties of muscle tissue, which are influenced by water content. Calibration equations are then used to convert water content estimates into lipid content values. For selection, individuals are ranked according to a weight-to-log(fat value) ratio: the bottom 10% in the lean line and the top 10% in the fat line are chosen to produce the next generation.

These two lines share the same genetic origin but have been subjected to opposite selection pressures, providing an ideal model to detect genomic signatures of selection - molecular footprints left when selection drives the frequency of certain alleles upward (Nielsen 2005). Such processes can reduce local genetic diversity, increase homozygosity, and cause marked allele frequency differences between populations.

We used two complementary approaches to detect these signatures. The first identified regions of extended homozygosity within each line, which may indicate the fixation of advantageous alleles. The second measured allele frequency divergence (F_{ST}) between the lines to pinpoint regions that may have been differentially targeted by selection. Combining these methods allowed us to identify candidate genomic regions potentially associated with traits related to lipid metabolism, energy storage, or muscle physiology. These findings are of practical interest for aquaculture, as adequate fat content is essential for processing products such as smoked fillets, which are currently lacking in some commercial strains.

In this context, the present study aims to assess genetic diversity and detect signatures of selection within experimental rainbow trout lines by making use of available genomic resources. The following section details the biological material used and the methodological framework adopted.

2. Materials and methods

a. Ethical statement

All the fish used for the analyses are currently reared or their sperm is cryoconserved in straws at the INRAE experimental facilities (PEIMA, INRAE, 2021, Fish Farming systems Experimental Facility, DOI: 10.15454/1.5572329612068406E12, Sizun, France) authorized for animal experimentation under the French regulation D29–277-02. Fish were reared under conditions similar to those of commercial aquaculture facilities, and the experimental procedures did not require prior approval from an ethical committee.

b. Genotyping

A small fin clip was collected from each individual at the PEIMA experimental facilities. The samples were then sent for DNA extraction and genotyping to the INRAE genotyping platform GETYANE (Clermont-Ferrand).
In total, there were:

- 135 individuals genotyped for the 2023 cohort of the synthetic line
- 30 cryoconserved sperm straws genotyped for the 2010 cohort of the synthetic line
- 38 cryoconserved sperm straws genotyped for the 2012 cohort of the synthetic line
- 94 individuals genotyped for the 2023 cohort of the fat line
- 98 individuals genotyped for the 2023 cohort of the lean line

Genotyping was processed for 57 501 SNPs with the DNA chip Axiom™ Trout Genotyping array (Palti et al. 2015). All SNPs with unknown or multiple locations on the Arlee reference genome assembly (USDA_Omyka_1.1) (Gao et al. 2021) were removed from the analyses.
The remaining SNPs had to pass quality control filters using PLINK v1.9 (Chang et al. 2015). The quality control criteria employed in this study were as follows. First, SNPs were filtered on the Hardy-Weinberg Equilibrium at the threshold of 10^{-6} to eliminate the last genotyping errors. A filter on the minor allele frequency (MAF) was also implemented to keep SNPs with MAF equal or above 1% (Kanaka et al. 2023). Another filter was the call rate for each SNP, fixed at 0.03, meaning that to pass, each SNP needed to be genotyped for 97% of the individuals. Finally, animals with over 10% missing genotypes were excluded, in our case, all the individual genotypes passed this final filter.
At the end, to proceed with our analyses, we had a total of:

- 22 185 SNPs and 203 individuals for the evaluation of genetic diversity of the synthetic line
- 25 649 SNPs and 192 individuals for the detection of signatures of selection in the fat and lean lines

c. Hypothesis of genetic transmission within closed populations

Because no external genetic material has been introduced since the beginning of the experimental lines, any genomic similarities observed between individuals reflect inheritance through direct descent. Thanks to this closed population structure, we can interpret the identity by state (IBS) across individuals as identity by descent (IBD), which reflects a shared ancestry.

This closed population framework also supports the interpretation of our proxy of inbreeding (F_{ROH}) as a true coefficient of inbreeding, since homozygous genomic segments are expected to result from inheritance of identity by descent haplotypes.

d. Characterization of genetic diversity in the synthetic line

i. Principal Component Analysis

Principal Component Analysis (PCA) was conducted to visualize population structure and diversity. In other words, we wanted to see if and how different lines of rainbow trout were clustered based on their genotypes. PCAs were performed by using the glPca function from the adegenet package (Jombart 2008) in R version 4.2.1 (R Core Team 2022) and run on RStudio software (RStudio Team 2022).

We then used the ggplot2 package (Wickham 2016) to plot the PCA results and observe a potential clustering between the lines through graphs representing two-by-two the first three principal components.

A first PCA was carried out between the current and the past cohorts of the INRAE synthetic line, to see if the current living stock is different enough from the frozen sperm of past cohorts so that the use of the cryoconserved sperm straws could help reintroduce genetic diversity in the future cohort.

A second PCA was carried out to compare the fat line and the lean line of rainbow trout. The objective was to see if they were really forming two very distinct clusters because of the 8 generations of divergent selection those lines have undergone on their muscle fat content.

ii. Identity by State and Identity by Descent

Because the objective was to select useful cryopreserved sperm straws to reintroduce genetic diversity into the synthetic line, we first needed to evaluate how different each sperm genotype is from genotypes of the current stock. In addition, as we want to use as little as possible the cryopreserved sperm straws, we also evaluated how close the frozen material was to each other.

To do that, we need to calculate the Identity by State (IBS) probabilities. Identity by State refers to alleles or DNA segments that are identical in sequence between individuals, regardless of whether or not those similarities originated from a common ancestor. When the similarities are inherited from a common ancestor, an IBS probability is in fact an Identity by Descent (IBD) probability (Wright 1922). IBS and IBD are estimated pairwise between individuals. Based on the values of the IBD probability between two individuals, we can know the expected relationship between different individuals (Table 1).

Table 1: Parental link based on IBD values between two individuals (Wright 1922)

Parental link	Parent ~ Offspring or Full Siblings	Half Sibs	First Cousins
IBD values	0.5	0.25	0.125

The concept of IBD is based on pedigree, meaning that we need to know the information of the parents of each individual, information that we don't have in our case. Instead, we must rely on IBS values, which don't need ancestry data. In our specific study, because the population is closed, with all

individuals originating from the same cohort, we can assume that shared segments between individuals are likely inherited, and so IBS segments can be interpreted as IBD segments (Figure 2). Accessing the IBS values, and so the IBD values and the coefficient of relatedness, is possible thanks to PLINK v1.9 (Chang et al. 2015) by using the --genome option.

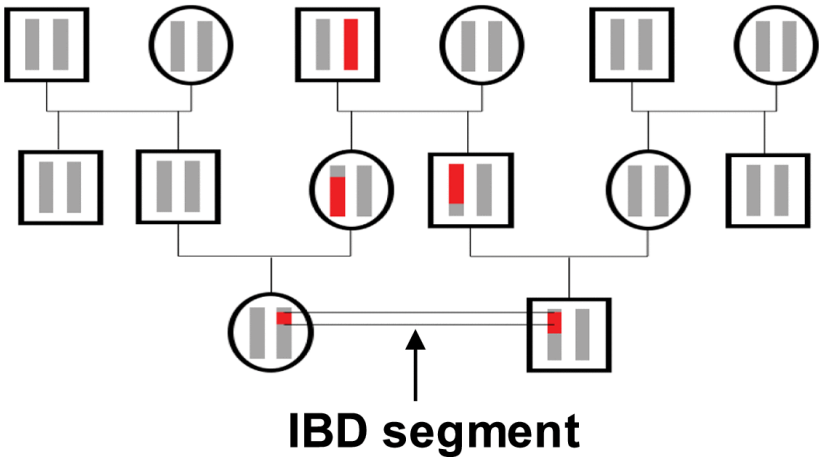

Figure 2: Representation of the appearance of an IBD segment between two individuals from a common ancestor (Zhang et al. 2023)

iii. Run of Homozygosity

A Run of Homozygosity (ROH) is a long continuous region of the genome where an individual is homozygous across all markers, while there is polymorphism observed at the marker positions in the corresponding population (Figure 3). This double copy of the same haplotype is assumed to be inherited from a common ancestor and thus considered as IBD segments (McQuillan et al. 2008). The haplotype is then autozygous, meaning homozygous by descent.

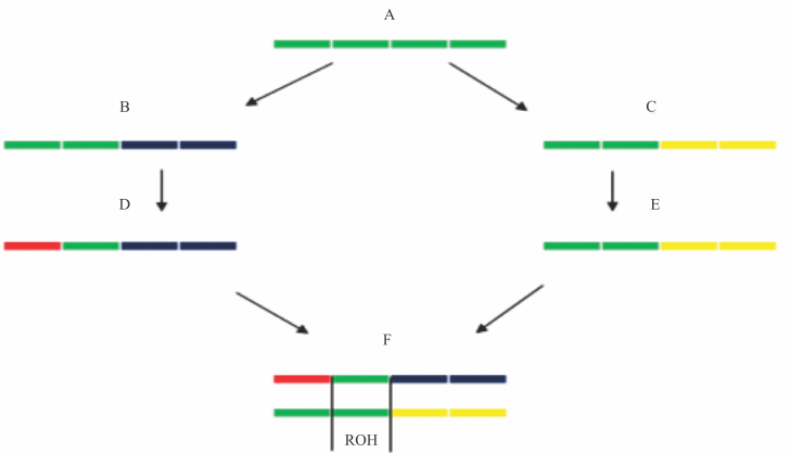

Figure 3: Representation of the formation of a ROH. Individual F presents a ROH (in green) formed by the pairing of stretches in homozygosis of the homologous chromosome of a common ancestor A (Rebelato, Caetano 2018)

ROH were calculated thanks to PLINK 1.9 (Chang et al. 2015) by using the "--homozyg" option. All the options we used are described in the following table (Table 2).

Table 2: Description of the options used to obtain the Runs of Homozygosity (ROH) with PLINK v1.9 (D'Ambrosio et al. 2019)

Option	Value	Description
homozyg		Enables the detection of ROH in the entry file
homozyg-density	100	Definition of the minimal SNP density in a ROH (1 SNP per 100 kb)
homozyg-kb	1000	Definition of the minimal length of your ROH (in kb)
homozyg-het	1	Definition of the maxima number of heterozygous SNP in a ROH (in SNP)
homozyg-snp	30	Definition of the minimum number of SNP in a ROH (in SNP)
homozyg-window-snp	30	Definition of the number of SNP per sliding window (in SNP)
homozyg-window-missing	2	Definition of the maximum missing SNP allowed in a window (in SNP)
homozyg-gap	100	Definition of the maximum distance allowed between two SNPs in a ROH (in kb)

ROH can be used to estimate the coefficient of inbreeding of individuals.

iv. Coefficient of inbreeding

A genomic measure of the level of autozygosity and so the level of inbreeding is defined as the proportion of genome that is included in Runs of Homozygosity:

$$F_{ROH} = \frac{\sum L_{ROH}}{L_{genome}}$$

Where F_{ROH} is the coefficient of inbreeding, $\sum L_{ROH}$ is the total length of the genome of an individual that is in ROHs and L_{genome} is the length of the genome of the individual that is covered by SNPs.

e. Description of the optimization algorithm used to select cryoconserved sperm straws to increase genetic diversity of the future cohort of the synthetic line

To optimize the genetic diversity of the future cohort while respecting operational constraints about relatedness across individuals, we used an algorithm based on Simulated Annealing implemented in the stats package in R. The function used is constrOptim(). The simulated annealing framework was chosen because our optimization problem involves multiple continuous parameters and nonlinear constraints on relatedness among individuals (Kirkpatrick, Gelatt, Vecchi 1983). Unlike gradient-based methods that may get trapped in local optimum, simulated annealing probabilistically explores the solution space by accepting worse solutions early on, thus increasing the chance of finding a global

optimum. This makes it particularly suited for complex genetic optimization tasks, such as maximizing diversity while controlling relatedness. In other words, this algorithm will try to converge to a minimum (or a maximum) while accepting some worst solutions occasionally to avoid local minimum. This method was previously explored and described by a previous student (Colon 2022).

This algorithm allows us to minimize, or maximize, a score for a target function while respecting some constraints. In our case, it was used to maximize the genetic diversity of a future cohort by identifying the optimal proportion as well as the specific cryopreserved genetic material to use relative to fresh sperm from the current cohort 2023.

The goal of the sets of constraints is to eliminate any candidate sperm that would not respect either one of the constraints, meaning, any cryoconserved sperm straw that would be too related to the current cohort as well as to avoid selecting cryoconserved sperm straws from related males.

The following three constraints were considered:

- Constraint 1 (C1): Cryoconserved straws were prioritized based on the median of the coefficient of relatedness between the selected straws and the individuals of the current cohort.
- Constraint 2 (C2): Cryoconserved straws were prioritized based on the third quartile of the coefficient of relatedness between the selected straws and individuals of the current cohort.
- Constraint 3 (C3): Cryoconserved straws were prioritized based on the median of the coefficient of relatedness among the straws themselves.

At first, we tested the scenario with no constraints (meaning the maximum values observed for all three coefficients of relatedness), meaning that all the individuals were allowed to be selected by the algorithm; this will be referred to as scenario C0. We then tested all the constraints one by one; we tested them by reducing the value of the constraint while keeping the maximum value for the other coefficients of relatedness and assessed the effect on the number of individuals selected the mean of heterozygosity of the future cohort and the gain in genetic diversity.

With this approach, we created different scenarios:

- CX_a: we constrain CX by limiting the value of the coefficient of relatedness to the 80% of the individuals with the lowest values.
- CX_b: we constrain CX by limiting the value of the coefficient of relatedness to the 50% of the individuals with the lowest values.

Where CX is either the constraint C1, C2 or C3.

f. Optimized value: Mean of heterozygosity and the proportion of cryoconserved sperm straws to use

To increase the genetic diversity of the future cohort, we considered two potential optimization criteria: the mean and the variance of the expected heterozygosity across individuals of the future cohort. To determine which metric was more relevant, we focused on the one that would most be reliable to measure high genetic diversity across all individuals.

While the maximization of the variance of the expected heterozygosity could create offspring with extreme genetic profiles (a part of the cohort largely homozygous and another part highly heterozygous) it would then be necessary to assess whether the homozygous regions contain rare

alleles or simply reflect redundancy of already common variants. This approach introduces uncertainty or more complexity to verify the information. The formulas for the variance are presented in annex I.
In contrast, the optimization of the mean heterozygosity ensures that most individuals maintain a high level of heterozygosity. This strategy promotes a more balanced genetic distribution across the cohort and reduces the probability of producing highly inbred individuals and so supporting long term population resilience.
The mean heterozygosity of the future cohort is calculated with this formula:

$$E_{div}(Next\_Cohort) = (1-p) * E_{div}(Current\_Cohort) + p * E_{div}(Cryoconserved\_Straws)$$

With:
p = proportion on cryo-conserved genetic material used for the reproduction of the next cohort in a mating plan for 100 males used with equal amount of sperm for each of them to contribute to the next generation.

$$E_{div}(Y) = \frac{\sum_{i=1}^{n}(1 - F_{ROH_i})}{n}$$

Where:

- Y is the genotyped individuals of the current cohort or the cryoconserved straws selected
- n is the number of individuals considered for the cohort
- F_{ROH_i} is the coefficient of inbreeding for the individual i and $1 - F_{ROH_i}$ is an indicator of heterozygosity

The model considers next-generation heterozygosity as a linear combination of parental contributions, providing a simplified yet practical framework.
It should be noted that if there is no use of cryoconserved sperm straws (the scenario where $p = 0$), the mean of the expected heterozygosity of the future cohort remains unchanged, with a value of $E_{div}(2025_{p=0}) = E_{div}(Current\_Cohort) = 0.8918$.
Based on this result, we obtain a reference. It is now possible to compute the gain of each optimization scenario we studied based on our set of constraints. The gain will represent the percentage increase in expected heterozygosity that would result from using the specific cryopreserved straws selected by the algorithm to produce the future cohort.

$$Gain\ (\%) = \frac{E_{div}(2025_{p_1}) - E_{div}(2025_{p=0})}{E_{div}(2025_{p=0})} * 100$$

Based on this approach, we evaluated several scenarios involving different sets of cryoconserved straws and mating constraints and calculated the expected gain in heterozygosity for each.

g. Genetic structure in fat and lean lines

A PCA was used as an exploratory tool to visualize the main axes of genetic variation within and between the two lines. This multivariate approach allows the projection of individuals onto a reduced

set of dimensions that capture most of the genotypic variance, enabling the detection of genetic clustering patterns. A clear separation between the fat and lean lines in the PCA space would indicate a strong genetic differentiation consistent with the applied divergent selection. To better interpret the observed clustering patterns, we complemented PCA with a systematic inspection for outliers that could represent admixed individuals. Previous analyses of the genetic structure in earlier cohorts of the fat and lean lines using PCA revealed that a small number of individuals did not cluster clearly with either line, and these were considered potential hybrids between the two populations (Figure 4). With this precedent in mind, the PCA of the current dataset was also examined for intermediate individuals potentially resulting from admixture. When such individuals were detected, we further investigated their genetic background using the Mongrail program (Chakraborty, Rannala 2023), which assigns individuals to genealogical classes based on phased haplotype data. This model distinguishes pure parental types, F_1 hybrids, F_2 hybrids, and backcrosses. For this analysis, two reference populations were defined using individuals previously classified as belonging to the “fat” or “lean” lines. The genealogical classes considered were: pure fat (offspring of two fat-line parents), pure lean (offspring of two lean-line parents), F_1 hybrids (offspring of one fat-line and one lean-line parent), F_2 hybrids (offspring of two F_1 individuals), fat backcrosses (offspring of a fat-line individual and an F_1), and lean backcrosses (offspring of a lean-line individual and an F_1). For each individual, Mongrail calculates posterior probabilities of membership in each genealogical class. This approach enabled the detection and classification of hybrids with greater confidence than PCA alone, ensuring that only genetically well-assigned individuals were retained for downstream analyses of genetic structure and signatures of selection.

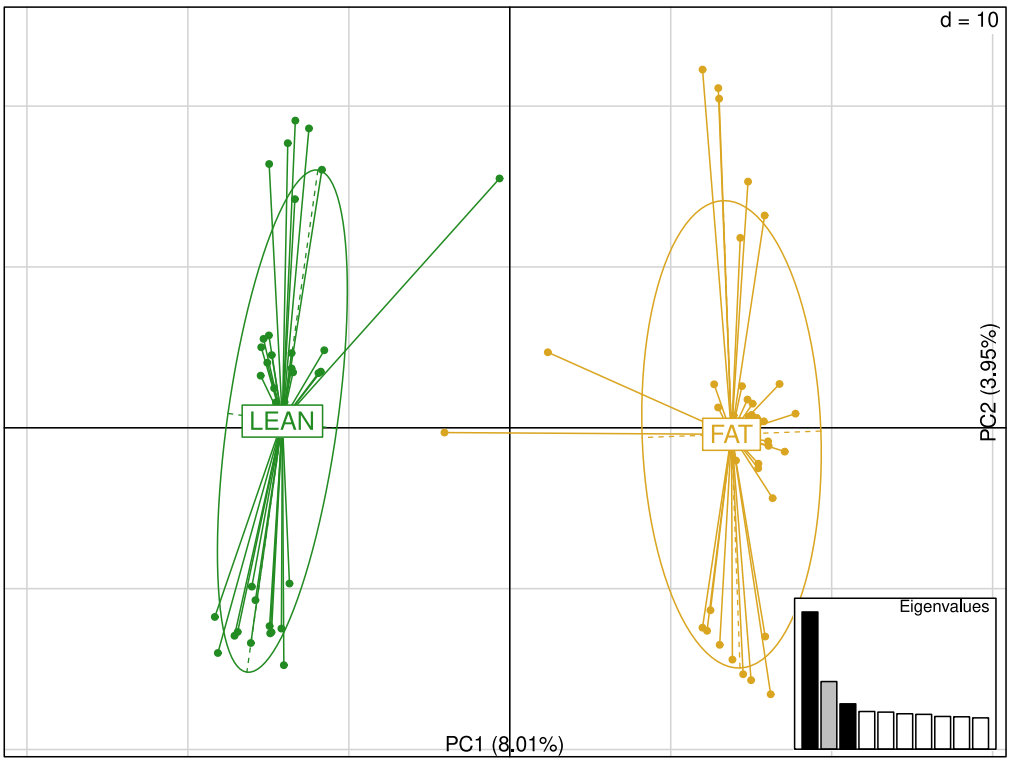

Figure 4: PCA of the previous cohorts of the lean and fat lines, already showing the existence of 3 hybrids between the two lines (2 from the fat cohort and 1 from the lean cohort)

h. Detection of signatures of selection

To detect potential signatures of selection in the fat and lean lines, we applied complementary approaches based on the detection of Runs of Homozygosity (ROH), and estimation of the fixation index (F_{ST}).

ROH analysis was performed to identify potential genomic regions under selection within each line (Pemberton et al. 2012). We calculated the proportion of the occurrences of a SNP in a ROH by counting the number of times the SNP was detected in those ROH across individuals. This recurrence was plotted against the physical position of SNPs along the chromosomes. The genomic regions most frequently encompassed by ROH were defined as those containing SNPs within the top 0.5% of recurrence values, following the threshold used by Xu *et al.* (2019).

The fixation index F_{ST} was used to quantify genetic differentiation between the fat and lean lines at each SNP. F_{ST} measures the proportion of genetic variance attributable to allele frequency differences between populations, ranging from 0 (identical allele frequencies) to 1 (complete fixation of alternative alleles). High F_{ST} values are indicative of genomic regions subjected to divergent selection, where different alleles have been preferentially fixed in each population (Qanbari et al. 2011; Zhao et al. 2015). We calculated per-site F_{ST} values using VCFtools (Danecek et al. 2011) with the Weir and Cockerham (1984) estimator, applied to SNP data from both lines. Negative F_{ST} values, which are non-biological artefacts, were set to zero following previous recommendations (Akey et al. 2002; Maiorano et al. 2018). SNPs were plotted according to their genomic position, and the top 0.1% of F_{ST} values were retained as candidate signals of selection, in line with thresholds used in earlier studies on aquaculture species (Kijas et al. 2012; Zhao et al. 2015). Combining ROH and F_{ST} results could allow a cross-validation of candidate regions and increase confidence in the identification of the regions we considered as signatures of selection.

i. Identification of candidate genes

A genomic region was considered to be under selection pressure if it contained at least one significant SNP based on the ROH or F_{ST} threshold values. For both approaches, genomic regions were defined using a sliding window procedure. Starting from the most significant SNP among the outliers (top 0.1% for F_{ST} and top 0.5% for ROH), we extended the window 500 kb upstream and downstream. If another significant SNP was located within this 1 Mb interval, it was incorporated into the same window. The process was then repeated from each newly incorporated SNP, extending the search by 500 kb on either side until no additional significant SNPs were found within this distance. The final region was then extended by an additional 500 kb upstream and downstream to ensure coverage of potentially linked loci. Genes located within the windows of interest were identified from the NCBI *O. mykiss* genome assembly release 101 (USDA_OmykA_1.1) and are presented in annexes V to VII.

To explore the biological relevance of these regions, we performed functional enrichment analysis using the FishEnrichr online tool (Chen, Tan, Kou 2013; Kuleshov, Jones, Rouillard 2016), which is specifically designed for enrichment studies in fish species. The analysis was carried out using zebrafish (Danio rerio) as the reference organism. Gene lists derived from the ROH- and F_{ST}-defined regions were mapped to Gene Ontology (GO) terms. FishEnrichr provides several metrics for ranking functions: the p-value, which assesses the probability of observing a given function by chance; the adjusted p-value,

which corrects for multiple testing to reduce false positives; the z-score, which quantifies the deviation of observed functional ranks from the expected distribution; and the combined score, which integrates both statistical significance and effect size by combining the log-transformed p-value with the z-score. Given the exploratory nature of our study and the aim of generating biological hypotheses, we chose not to exclude functions solely based on the adjusted p-value to avoid discarding potentially relevant signals. The classification of enriched functions was based on the top ten GO terms ranked by combined score, with particular attention given to functions related to lipid metabolism, energy balance, and muscle biology, as these were expected to be among the main targets of divergent selection between the fat and lean lines.

3. Results

a. Characterization of genetic diversity in the synthetic line

i. PCA of the current genetic diversity in our lines

Based on the genotyping data of the current and past cohorts of the synthetic line, a PCA was generated (Figure 5). This analysis allows the exploration of the genetic structure and highlights the variation between the living cohort and the deceased cohorts.

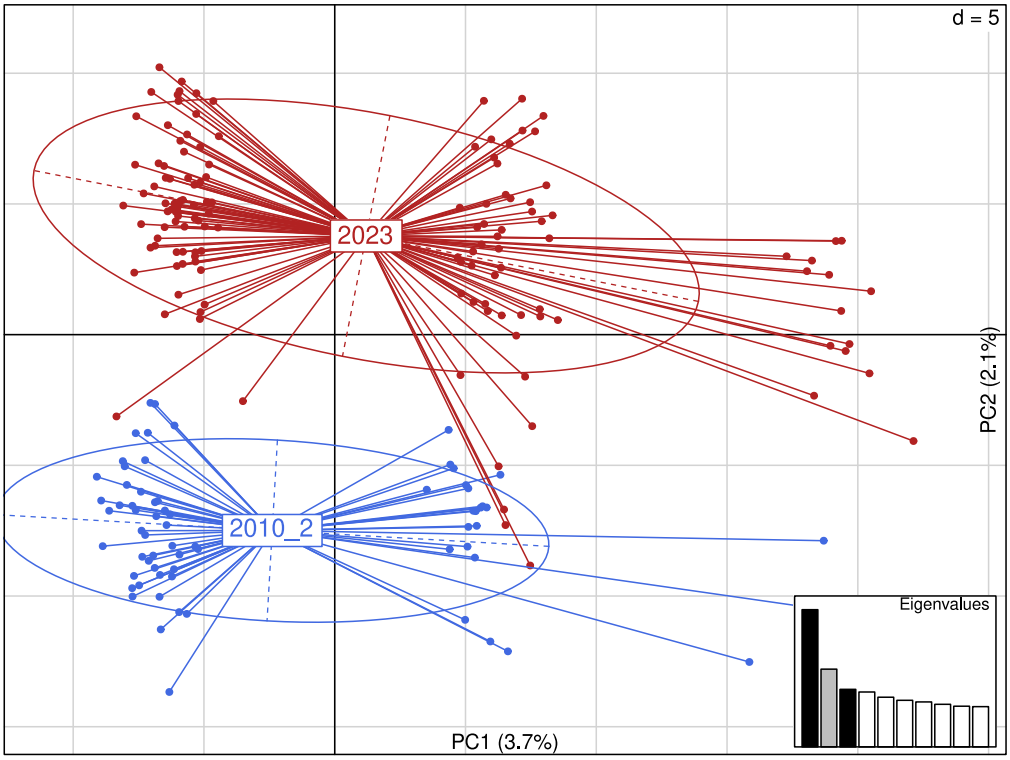

Figure 5: Principal Components Analysis representing the genetic diversity of the synthetic line's cohorts – current synthetic cohort of 2023 (in red) and deceased cohorts of 2010 and 2012 (in blue)

The first two axes explain respectively 3.7% and 2.1% of the global variance. The result indicates that the current and past cohorts are moderately differentiated based on SNPs. Thus, using cryoconserved sperm straws from past cohorts could help reintroduce genetic diversity in the synthetic line.

ii. Coefficient of inbreeding

The inbreeding coefficients are presented in the following table (Table 3).

Table 3: Distribution of F_{ROH} for the different cohorts studied

Cohorts	Mean	Min	Q1	Med	Q3	Max
Current cohort	0.1082	0.0511	0.0847	0.1007	0.1081	0.2576
Cryoconserved sperm straws	0.0955	0.0412	0.0776	0.0875	0.1080	0.2646

As shown in table 3, coefficients of inbreeding are lower than those reported in the commercial lines of rainbow trout (D'Ambrosio et al. 2019). This result is very much expected because the synthetic line has not been selected in any way before. However, as inbreeding is not managed a few individuals have high inbreeding coefficients with values between 0.20 and 0.26, both in the recent and past cohorts.

iii. Coefficient of relatedness

Since the main goal is to see if we need to reintroduce genetic diversity into the current synthetic line using cryopreserved sperm straws, it was essential to evaluate both the coefficients of relatedness between each cryoconserved material and the individuals of the current cohort, and the coefficient of relatedness among the individuals with cryoconserved sperm straws. That information is crucial for identifying sperm samples that are genetically distant from the current population, thus maximizing the potential gain in diversity, while also avoiding redundancy among selected cryoconserved sperm.
The following table shows the distribution of IBS probabilities calculated between individuals within the current cohort (2023), within the past cohorts (2010_2) and between the current and past cohorts (Table 4).

Table 4: Distribution of the IBS coefficients across pairs

Pairs of cohorts	Mean	Min	Q1	Med	Q3	Max
2023 x 2023	0.0812	0	0.0317	0.0860	0.1181	0.5509
2010_2 x 2010_2	0.0357	0	0	0.0303	0.0624	0.5447
2010_2 x 2023	0.0431	0	0	0.0336	0.0691	0.4237

The current synthetic cohort has the highest levels of relatedness among its members compared to the past cohorts (within themselves and when confronted to the current individuals) This suggests that the cryoconserved sperm straws are more original, genetically speaking, and they could possess alleles that have been lost in the current cohort.
However, in any cohort, IBS values can range from very low to quite high. Indicating that we have some closely related individuals (with a full sibling's relation or a parent offspring relation) and unrelated individuals.

The figure 6 shows the distributions of median and third-quartile coefficients of relatedness between the cryoconserved sperm straws and the individuals of the current cohort (top and middle panels), as well as the median coefficient of relatedness within the cryoconserved sperm straws themselves (bottom panel).
The distribution of estimated pairwise coefficients of relatedness are presented below (Figure 6).

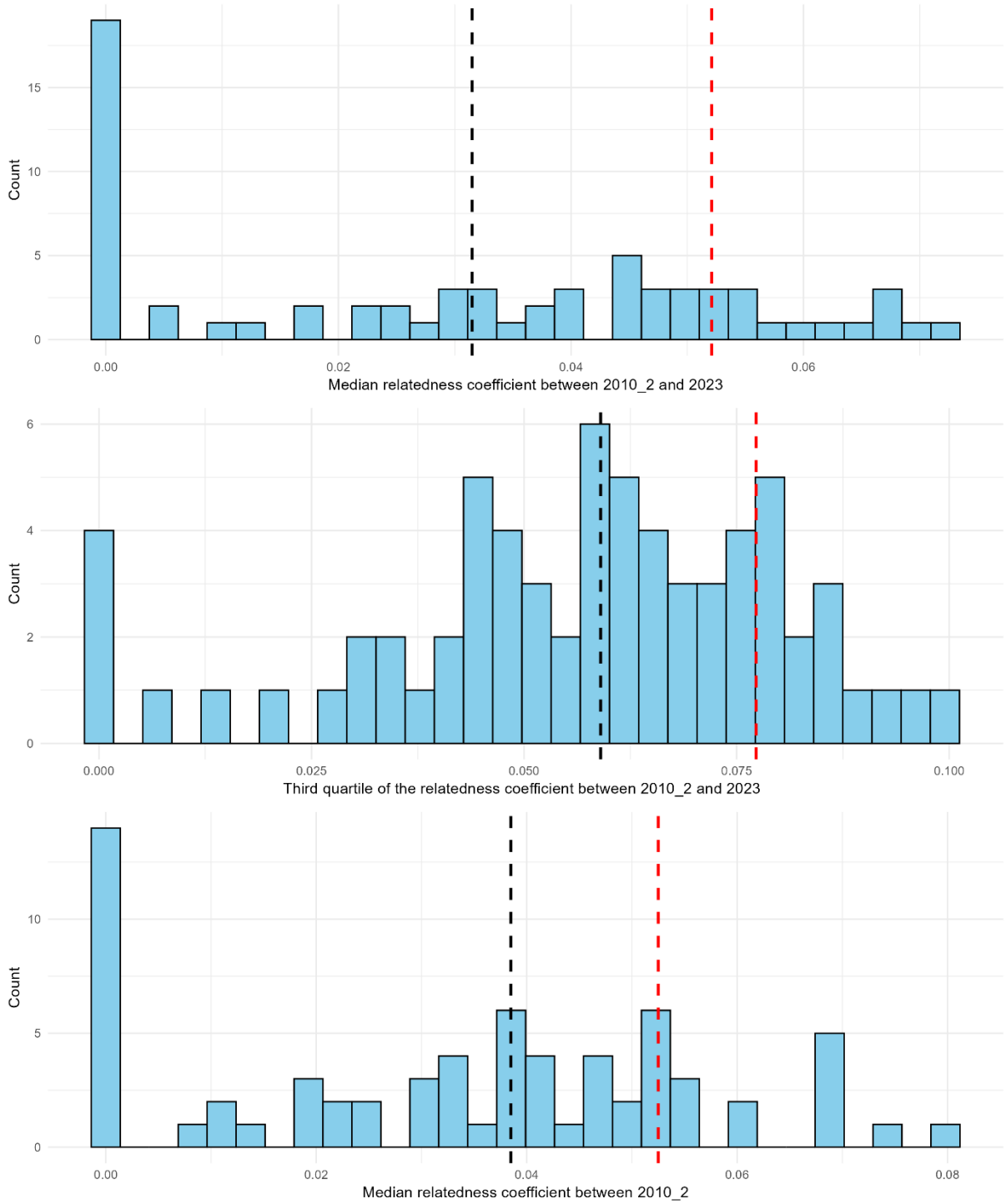

Figure 6: Distributions of relatedness coefficients between cryoconserved sperm straws "2010_2" and the individuals of the current cohort "2023" (top and middle panel) and within the past cohorts themselves (bottom panel). The dashed lines correspond to different percentile values (the red dashed lines correspond to the 80th and the black dashed lines to the 50th, i.e. the median)

These distributions inform us about the structure of coefficient of relatedness. They are highlighting a wide range of coefficient of relatedness, confirming that some individuals from the past cohorts are not related to the current cohort or even to other past individuals, which could make their

cryopreserved sperm straws interesting for our goal. Those distributions were important to set thresholds for our constraints in the optimization problem.

b. Optimization of the mean heterozygosity of the next cohort

To assess the potential gain of genetic diversity by introducing cryoconserved sperm straws, different sets of constraints were tested. As indicated in the Material & Methods section, if we do not introduce any cryoconserved straw in the next reproduction, the mean heterozygosity will be $E_{div}(2025) = 0.8918$. This value served as a baseline for the evaluation of the gain of heterozygosity.
We applied three types of constraints, each with two levels of discrimination to see if it was possible to improve the genetic diversity while limiting cryoconserved sperm straws usage. The following table summarizes the interesting scenarios tested (Table 5).

Table 5: The effects of the different tested constraints of the genetic diversity of the future cohort of the synthetic line

Constraint tested	**None**	**Constraint 1**		**Constraint 2**		**Constraint 3**	
Name	**C0**	**C1_a**	**C1_b**	**C2_a**	**C2_b**	**C3_a**	**C3_b**
Constraint 1	0.0722	0.0521	0.0315	0.0722	0.0722	0.0722	0.0722
Constraint 2	0.0995	0.0995	0.0995	0.0773	0.0590	0.0995	0.0995
Constraint 3	0.0798	0.0798	0.0798	0.0798	0.0798	0.0525	0.0385
Selected fish	66	54	34	54	34	51	33
Mean diversity	0.8994	0.8959	0.8906	0.8959	0.8909	0.8945	0.8899
Gain (%)	0.86	0.46	-0.13	0.47	-0.10	0.30	-0.21

First, the optimization algorithm was run without applying any constraint (C0) to see what the maximum of the mean of heterozygosity could be achievable in the future cohort. The scenario C0 shows that we could be at 0.8994, which means that we could only gain 0.86% of genetic diversity. In that case, we must use sperm straws of 66 out of the 68 individuals from previous cohorts. However, this scenario is not optimal in our opinion because it would nearly exhaust the cryobank.

Then we tested constraint 1 with threshold values C1_a or C1_b on the maximum median relatedness between the selected cryopreserved sperm and the recent cohort; meaning that we try to take out of the optimization straws that were too close to the current synthetic cohort. Respectively 54 and 34 out of the 68 cryopreserved individuals had to be selected to maximize the mean heterozygosity of the future cohort. For the scenario C1_a, only a gain of 0.46% of expected heterozygosity can be expected compared to the mean heterozygosity of the current cohort. For scenario C1_b, the situation is even worse as no gain can be expected (-0.13%) while using sperm straws from half of the cryopreserved fish.

Considering constraint 2 with scenarios dedicated to limit relatedness between the current cohort and the third quartile of cryopreserved fish, we got very similar results to those observed with constraint 1. There are no scenarios respecting the constraint and our choice of the maximum straws used that would increase the genetic diversity of the future cohort.

Finally, scenarios introducing constraint 3 on the maximum coefficient of relatedness among selected sperm of the past cohorts indicated that to use only about half of the cryopreserved sperm no gain in heterozygosity can be expected for the future cohort (-0.21% in scenario C3_b).

c. Signatures of selection in divergent lines

i. The clustering of the lines

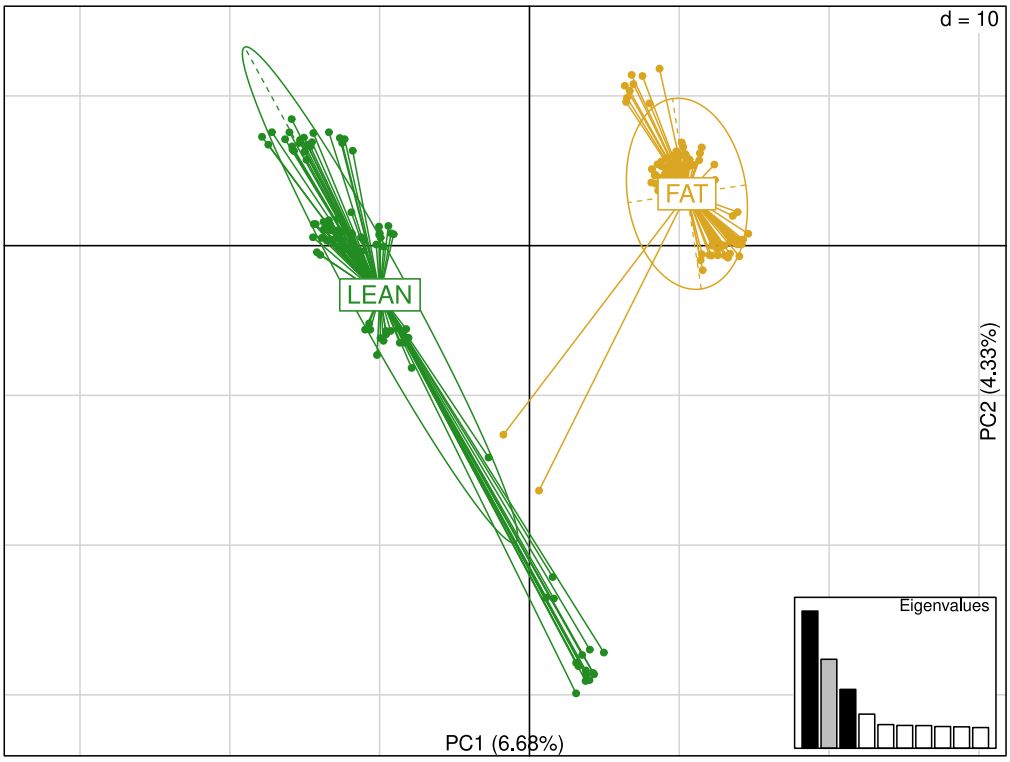

Figure 7: PCA plot of the two first principal components (PC) showing the segregation between the fat line (yellow) and the lean line (green) and individuals considered as hybrids of those lines

As shown in the PCA (Figure 7), PC1, explaining 6.68% of the variance, clearly separates the two lines of rainbow trout into two distinct clusters. PC2 also contributes to this separation, further structuring each line. Each cluster appears to consist of three subgroups of individuals, suggesting genetic differentiation and reduced relatedness among subgroups within each line. The first two components explain 6.68% and 4.33% of the variance, respectively.

Some individuals are not fully separated by PC1 while strongly segregated by PC2. A total of 18 individuals (2 from the fat line and 16 from the lean line) formed a third cluster, referred to as the hybrid cluster.

Hybrids identified by the first two principal components of the PCA were analyzed using the Mongrail framework. None of the 18 individuals has reached a sufficient value that would allow a confident classification (95%). However, for the two hybrids originated from the fat line and 14 out of 16 hybrids from the lean line, the largest proportion of their genome is estimated to be derived from the “pure lean” class, with posterior probabilities ranging from 35% to 93%. The last two individuals are showing a genetic pattern most similar to the “pure fat” class and “lean backcross” class with percentage of 53% and 41% respectively. As a result, none of the hybrids were reassigned to any class and they were excluded from subsequent analyses.

After the hybrids were taken away, the genetic structure of the data set from the two new lines was captured again by the first two principal components of a PCA (Figure 8).

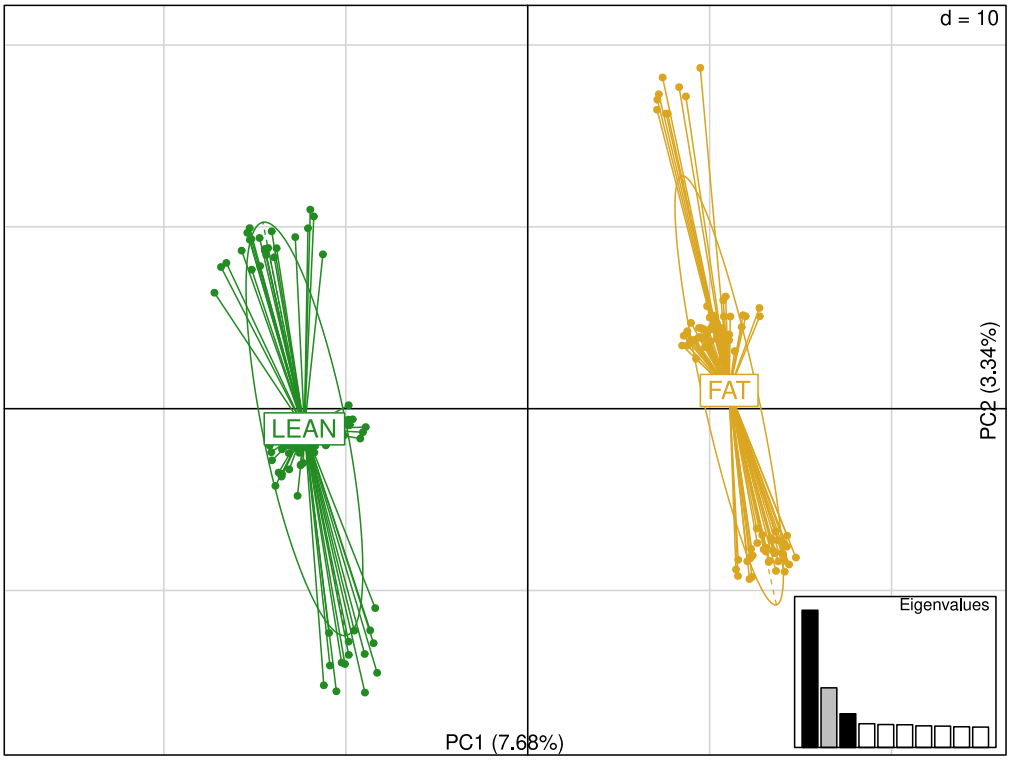

Figure 8: PCA plot of the two first principal components (PC) showing the segregation between the fat line (yellow) and the lean line (green) without any hybrid individual

Once again, PC1 allows a clear segregation between the two lines of rainbow trout into two separated clusters (Figure 8). PC2 also reveals the same repartition inside each line. Each cluster seems to be formed by 3 groups, which shows diversity in each line. This time, PC1 and PC2 are explaining 7.68% and 3.34% variance respectively.

After exploring the segregation and the clustering between the fat and lean lines of rainbow trout, we excluded suspect individuals for the subsequent analysis aiming to identify signatures of selection using ROH and F_{ST} approaches.

ii. Signature of selection

To identify the genomic regions that were most commonly associated with ROH in all individuals, the top 0.5% of SNPs with the highest occurrences in ROH were considered as candidates, as outliers. This represents 127 SNPs selected as outliers for this study and it means we selected frequencies above 48.91% and 45.78% for the fat line and the lean line respectively. The SNP candidates were located in chromosomes 7, 14, 23, 30 and 32 for the lean line with percentage ranging from 45.78% to 53.01% (Figure 9, top panel). The set of candidate SNPs for the fat line was located in chromosomes 6, 8 and 16, ranging from 48.91% to 58.69% (Figure 9, bottom panel).

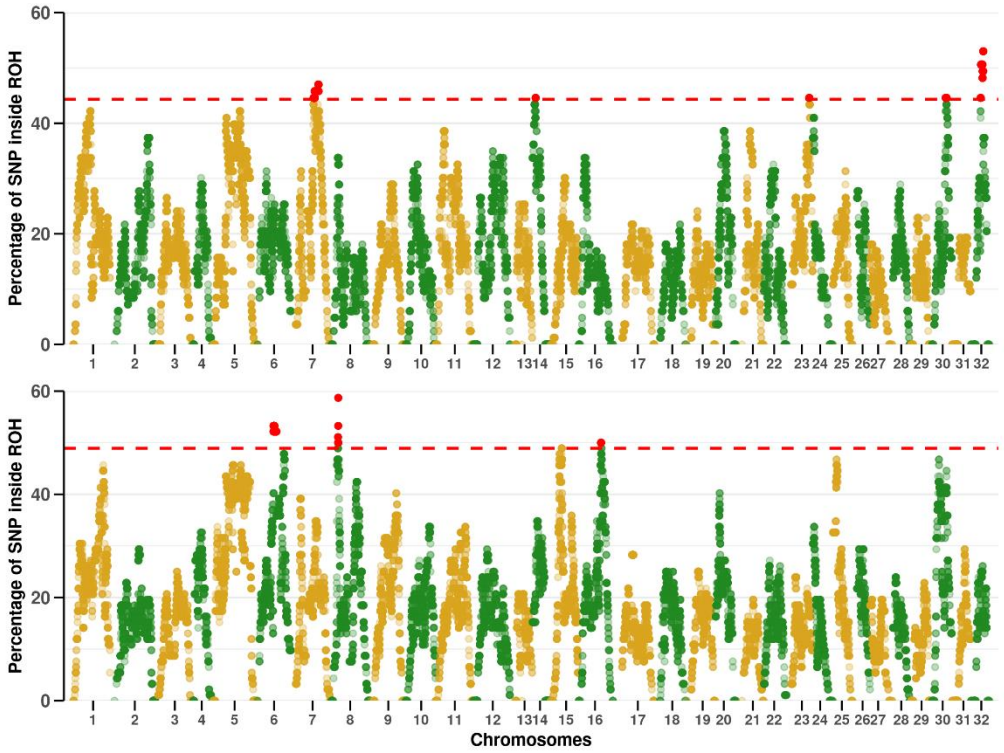

Figure 9: Distribution of percentage at which each SNP appears within ROH segments across individual from the lean line (top panel) and the fat line (bottom panel). The threshold is represented by the red dashed lines and the outliers by the red dots

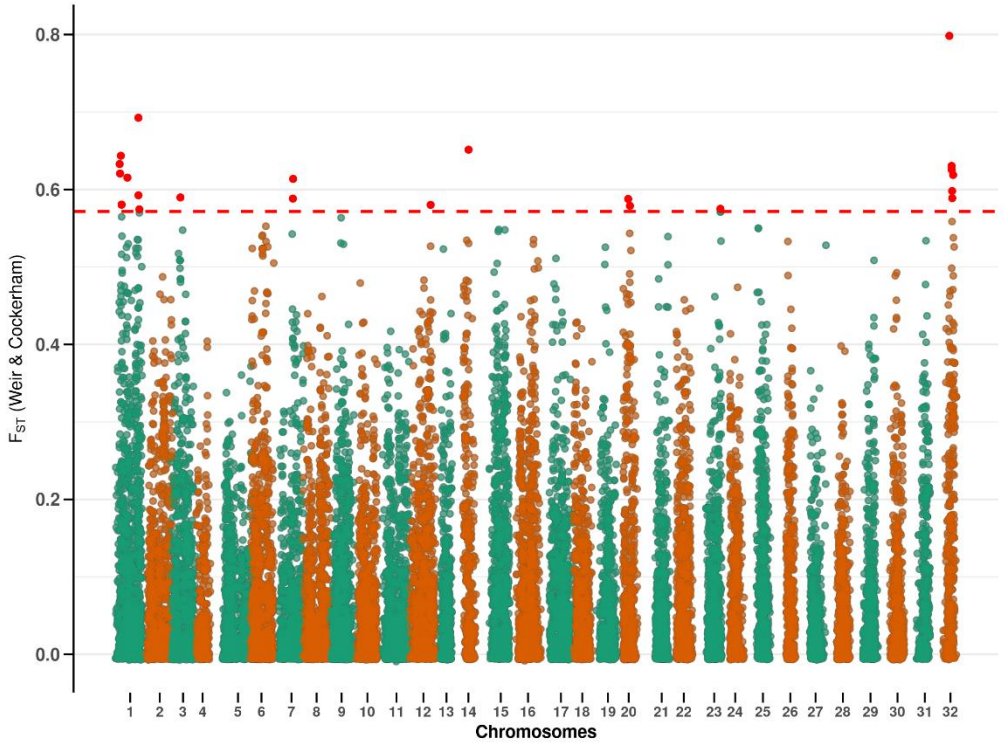

Figure 10: Genomic distribution of F_{ST} values. SNPs are plotted relative to their physical position within each chromosome. The cutoff to call SNP outliers is represented by the red dashed line and the outliers are represented as red dots

Another way to identify regions under selection pressure was to evaluate genetic differentiation. This data varied throughout the genome (Figure 10) with an average value of 0.080. By considering only the top 0.1% of the F_{ST} values, 32 SNPs were selected, ranging from 0.574 to 0.798, and located on chromosomes 1, 3, 7, 12, 14, 20, 23 and 32.

Thanks to the ROH and F_{ST} approaches we were able to define multiple regions of interest. In total, we obtained:

- 11 regions from the ROH approach of the lean line
- 4 regions from the ROH approach for the fat line
- 19 regions from the F_{ST} approach comparing the lean and the fat line

These regions are covering 18,735,691bp, 10,489,807bp and 19,579,891bp respectively based on the 500 kb-sliding window method used to select regions around the top SNPs.
Based on those regions, we collected the known genes in them. Genes were grouped by approach and analyzed for functional enrichment using FishEnrichr. A total of 3 enrichments were then performed. For each of them, we only took the GO terms that were ranked in the top ten by combined score of the tables from "GO Molecular Function 2018", "GO Biological Process 2018" and "GO Cellular Component 2018" (Annexes II to IV). From the ten GO terms of each table, we only kept GO terms that could be related to lipid metabolism, energy storage or muscle biology.
The selected GO terms included functions related to nuclear hormone receptor activity, microRNA-mediated gene silencing, and stress granule formation in the lean line; hormone receptor binding, steroid metabolism, lipid signaling, MAPK pathway activation, and cytoskeletal organization in the fat line; and developmental regulation, steroid hormone response, lipid metabolism enzymes, opioid receptors, and intracellular trafficking in the FST comparison.
The complete list of enriched GO terms that could be related to lipid metabolism, energy storage or muscle biology is presented in Table 6.

Table 6: Compilation of the enriched GO terms the most related to lipid metabolism, energy storage or muscle biology in regions identified by ROH or F_{ST} approaches in lean and fat lines

Approach - Line	GO Category	Function (Name + Description)
ROH – lean line	Molecular Function	**ligand-dependent nuclear receptor transcription coactivator activity (GO:0030374)** – ability of a protein factor to bind to a ligand-activated nuclear receptor (steroid, thyroid hormone, retinoid) to stimulate transcription of target genes.
	Biological Process	**production of miRNAs involved in gene silencing by miRNA (GO:0035196)** – process of generating microRNAs that participate in repression of gene expression by binding to target mRNAs.
	Cellular Component	**cytoplasmic stress granule (GO:0010494)** – cytoplasmic aggregates of proteins and mRNAs formed in response to cellular stress.

Approach - Line	GO Category	Function (Name + Description)
ROH – fat line	Molecular Function	**estrogen receptor binding (GO:0030331)** – binding to the estrogen receptor, involved in hormonal regulation. **steroid dehydrogenase activity (GO:0033764)** – oxidation-reduction of steroids via NAD(P). **prostaglandin E receptor activity (GO:0004957)** – cellular response to prostaglandins.
	Biological Process	**activation of MAPKK activity (GO:0000186)** – activation of MAPK pathways involved in cell growth and stress. **regulation of focal adhesion assembly (GO:0051893)** – control of the assembly of cell adhesion complexes to the extracellular matrix.
	Cellular Component	**anchored component of external side of plasma membrane (GO:0031362)** – proteins anchored to the external plasma membrane via lipid linkage. **intrinsic component of external side of plasma membrane (GO:0031233)** – transmembrane proteins exposing their active part to the cell exterior. **contractile actin filament bundle (GO:0097517)** – organized bundles of actin involved in cell contraction or shape. **actomyosin (GO:0042641)** – actin–myosin complex responsible for cellular contraction.
FST – lean vs fat	Molecular Function	**opioid receptor activity (GO:0004985)** – opioid receptors modulating cellular signaling. **palmitoyl-CoA hydrolase activity (GO:0016290)** – enzyme hydrolyzing palmitoyl-CoA, involved in lipid metabolism.
	Biological Process	**regulation of mesodermal cell fate specification (GO:0042661)** – control of mesodermal cell fate during embryonic development. **response to steroid hormone (GO:0048545)** – cellular response to steroid hormones.
	Cellular Component	**BLOC-1 complex (GO:0031083)** – protein complex involved in sorting and maturation of endosomal organelles.

These enrichment results provide a first overview of the potential biological processes, molecular functions and cellular components underlying the divergence between the lean and fat lines. While they highlight pathways consistent with differences in lipid metabolism, energy storage, and muscle biology, they also raise questions about their broader physiological implications. In the following discussion, we explore these findings considering previous studies on rainbow trout lines, considering both their biological relevance and their possible links with observed phenotypic traits.

4. Discussion

a. Management of the genetic diversity in the synthetic line

i. The genetic erosion

The results that we obtained in this study have shown a slight decrease in genetic diversity in the synthetic line of rainbow trout in the INRAE's PEIMA stock following the two major mortality episodes, causing bottlenecks in the last years. Indeed, coefficient of inbreeding tends to be higher in the current cohort than in past cohorts of the synthetic line, which means heterozygosity decreased a bit.
The PCA has shown a genetic differentiation between the current cohort and the past cohorts, suggesting a genetic drift and hinting that reintroducing previous genetic material could be a solution to gain genetic diversity.
Inbreeding coefficients were higher in the current cohort than in the past cohorts of the synthetic line (see Table 3). This heterozygosity decrease may have been caused by the bottlenecks induced by the recent slaughter episodes. Nevertheless, the average level of inbreeding in the current cohort is much lower than observed in the selected commercial lines used for French production (D'Ambrosio et al. 2019), indicating that the INRAE synthetic line is still an interesting population as a reservoir of genetic diversity.

ii. Advantages and limitations of cryopreserved sperm straws

Although the cryopreserved material represents a valuable reservoir of genetic diversity, our calculations indicated that its use would result in only a marginal effective gain of 0.86% in genetic diversity when employed to produce the next cohort of the synthetic line. Moreover, achieving this modest gain would require using cryopreserved samples from 66 out of the 68 individuals, thereby depleting a massive part of the stock. Such a strategy would severely limit the resources available for future reintroductions of genetic diversity. In the event of another bottleneck in the synthetic line, having consumed most of the cryopreserved stock for such a limited benefit could compromise future opportunities to restore genetic variability when it might be most urgently needed.
A further limitation lies in the choice of the metric used to optimize genetic diversity. Relying on the mean heterozygosity may not have been optimal; the variance of heterozygosity could represent a more informative criterion. Indeed, selecting breeders with homozygosity at multiple loci but carrying rare alleles would favor the retention and propagation of these alleles in subsequent generations. This approach could ultimately enhance the frequency of rare variants and lead to a more substantial increase in genetic diversity over time.

iii. Choice of the constraints used

Two types of constraints were tested, two constraints (C1 and C2) aimed to limit the relationships between the cryopreserved sperm and the current cohort and one (C3) was used to limit the relatedness of the selected sperm.

Results showed that there was no scenario helpful to increase the genetic diversity of the future cohort of the synthetic line while limiting the number of past breeders used to half the list of the 68 cryopreserved sperm.
Results can also be a hint that the current cryobank diversity may be insufficient for impactful short-term gains but might be useful for possible coming events.
Because of this point, and the fact that IBS values between cryopreserved sperm straws can be quite low, we could identify the individuals of the current cohort that are very different from the current cryoconserved stock. Those candidates could be sampled in order to create a new stock of cryopreserved sperm straws.

iv. Alternatives to increase genetic diversity

We may suggest alternative strategies to maintain genetic diversity.
First, it can be proposed to constitute an additional stock of cryopreserved sperm using the new breeders selected in cohort 2023. As we have demonstrated that more than half the cryopreserved sperm from cohorts 2010 and 2012 are almost unrelated (IBS probabilities < 0.04) to the fish from cohort 2023, a second cryopreserved stock from the new breeders could be useful to reintroduce genetic diversity that may be lost in a future crisis. We could recreate a new stock of cryopreserved sperm straws periodically.
Other studies have shown that we could also reintroduce external genetic resources to gain genetic diversity (Johnson et al. 2025). Unfortunately, our goals are different. INRAE's work is not to develop a line with all the diversity of the rainbow trout genome, but to keep the genetic diversity of the synthetic as close as it originally was.
We could also increase the number of breeders used for each generation to potentially increase effective population size (Ne). The higher the Ne, the lower is the heterozygosity decrease across generations (Wang 2005). Higher Ne counters genetic drift reduces the rate of inbreeding and maintains genetic diversity. However, managing a high Ne isn't an easy solution, especially for the PEIMA team that has to manage numerous rainbow trout lines (Lallias et al. 2021; Lagarde et al. 2023; Pouil et al. 2023) and in the context of newly launched experiments and logistical constraints of the facility.
Another limit that we could consider is the narrow sample size. We only had 134 individuals from the 2023 synthetic cohort that were genotyped, and the stock of cryopreserved sperm straws only counted 68 different individuals and some of them have already been used previously. While this restricted sampling might not fully represent the genetic landscape, it has been estimated in population genetics that 30–40 individuals are sufficient to estimate allele frequencies and Ne. Thus, sample size is unlikely to bias our main conclusions.

b. Signatures of selection in the fat and lean lines

i. Candidate genes and functional enrichment

To better understand the biological processes underlying the genetic regions identified, we examined the functional enrichment of genes located within the genomic regions detected using ROH and F_{ST}.

This analysis aimed to highlight pathways that could explain phenotypic differences between the lean and fat lines.

In the lean line, several enriched GO terms converged on mechanisms related to transcriptional regulation, post-transcriptional control, and cellular stress responses. Among them, the ligand-dependent nuclear receptor transcription coactivator activity (GO:0030374) stands out, reflecting the role of nuclear hormone receptors in metabolic pathways such as thyroid and steroid signaling. This is particularly relevant since nuclear receptor coactivators such as PGC-1α have been shown to reduce oxytocin expression and thereby regulate appetite and energy intake in zebrafish (Danio rerio) (Blechman et al. 2011). Such a mechanism provides a plausible molecular basis for the more efficient feed utilization and lower feed conversion ratio reported in the lean line (Blanc et al. 2025). Another important enrichment was the production of miRNAs involved in gene silencing by miRNA (GO:0035196), which points to tighter post-transcriptional regulation of gene expression. This is consistent with the role of miRNAs in fine-tuning developmental processes, metabolism, and cellular plasticity (O'Brien et al. 2018). It also resonates with physiological studies showing that the lean line displays improved resource allocation and more efficient growth despite lower lipid deposition (Blanc et al. 2025). Finally, enrichment for cytoplasmic stress granule formation (GO:0010494) highlights the cellular strategies used to cope with stress, such as oxidative stress, heat shock, or hypoxia, by transiently halting translation and storing mRNAs for later use or degradation. This may underpin the greater hypoxia tolerance observed in the lean line (Pouil et al. 2024), reinforcing the idea that this line has developed mechanisms of stress resilience and metabolic flexibility. Together, these enrichments point to a functional profile where the lean line maintains homeostasis under environmental challenges through regulatory plasticity and efficient use of available nutrients.

In the fat line, enriched GO terms revealed three major functional themes: hormonal regulation, intracellular signaling, and structural adaptation of adipose tissue. At the molecular function level, estrogen receptor binding (GO:0030331), steroid dehydrogenase activity (GO:0033764), and prostaglandin E receptor activity (GO:0004957) all indicate strong hormonal control of lipid metabolism and storage. In fish, estrogen receptors such as esr1, esr2a, and esr2b are known to modulate lipid homeostasis (Lu et al. 2017). Similarly, steroid dehydrogenases locally activate or inactivate steroid hormones within adipose and muscle tissues, thereby influencing energy partitioning and adipocyte differentiation (Aizawa et al. 2007). Prostaglandin signaling, via its receptors, has been shown to promote triglyceride accumulation in mice (Cai, Ying, Song 2015), suggesting a potential role in energy storage and metabolic inflammation. These hormonal pathways are consistent with experimental findings in rainbow trout, where the fat line exhibits higher hepatic expression of lipogenic genes, increased fatty acid bioconversion, and greater glycogen storage compared to the lean line, without improved glucose utilization (Kamalam et al. 2012).

At the biological process level, enrichment for activation of MAPKK activity (GO:0000186) suggests MAPK pathway modulation, which is known in mice to promote lipid accumulation and adaptation to metabolic stress (Ding et al. 2023). This result is consistent with observations that the fat line tends to accumulate more lipid throughout development (Quillet et al. 2007; Blanc et al. 2025). Another biological process, regulation of focal adhesion assembly (GO:0051893), points to extracellular matrix remodeling, a process required for the structural adaptation of adipose tissue during expansion. At the cellular component level, enrichments for intrinsic plasma membrane components, contractile actin

filament bundles, and actomyosin highlight cytoskeletal and membrane remodeling processes. These can modulate adipogenic signaling and provide mechanical adaptation to lipid overload, as shown in mice (Ramadori et al. 1990; Jeong, Ko, Park 2007; Chen et al. 2020). Such cytoskeletal remodeling could explain experimental findings that selection for muscle lipid content alters flesh quality traits such as texture and oxidative stability in trout (Lefevre et al. 2015), since muscle mechanical properties are closely linked to actomyosin structure and extracellular matrix composition. Collectively, the enrichment profile in the fat line depicts a system strongly oriented towards lipid accumulation and storage, with both metabolic and structural adaptations to support this phenotype.

The F_{ST}-based comparison between lean and fat lines further highlighted functional differences that directly relate to experimentally observed traits. For example, enrichment for opioid receptor activity (GO:0004985) is particularly relevant given that differences in feeding behavior have been reported between the lines. Blanc et al. (2025) showed that the fat line has a higher feed conversion ratio, requiring more feed to produce the same biomass as the lean line. The identification of opioid receptor activity, known to regulate appetite and energy balance in mammals (Tabarin et al. 2005), suggests a molecular basis for this behavioral difference. Similarly, variation in palmitoyl-CoA hydrolase activity (GO:0016290) reflects differences in lipid turnover, consistent with studies in humans showing its role in fat tissue composition (Watt et al. 2003). Developmental terms such as regulation of mesodermal cell fate specification (GO:0042661) suggest divergence in adipocyte precursor differentiation during early ontogeny (Sebo et al. 2018), which matches the experimental observation that the fat line shows long-term differences in flesh quality (Lefevre et al. 2015) and body composition (Kamalam et al. 2012). Likewise, enrichment for response to steroid hormone (GO:0048545) supports the idea that differential hormonal sensitivity contributes to distinct metabolic profiles, as recently suggested in mammals (Liang et al. 2023). Finally, enrichment of the BLOC-1 complex (GO:0031083), involved in endosomal trafficking of receptors and lipid-related proteins (Di Pietro et al. 2006), may contribute to differences in receptor recycling and lipid handling between the two lines.

Altogether, the integration of enrichment analyses with experimental data reinforces the view that the two lines represent distinct metabolic strategies shaped by divergent selection on muscle lipid content. The lean line relies on stress adaptation, miRNA-mediated post-transcriptional regulation, and efficient nutrient utilization, explaining its improved feed efficiency, growth trajectory, and tolerance to environmental stress. By contrast, the fat line is characterized by strong hormonal regulation, MAPK pathway modulation, adipogenesis, and cytoskeletal remodeling, which sustain higher lipid deposition, altered feeding behavior, and modified flesh quality.

In this study, we chose to focus only on GO terms most relevant to the detection of selection signatures. However, additional GO terms within the identified regions may also provide valuable insights into lipid metabolism, muscle biology, and energy storage, and these may have been overlooked.

ii. Thresholds and statistical considerations

For the detection of candidate regions, thresholds were applied to filter extreme signals: the top 0.5% of SNPs for ROH and the top 0.1% for F_{ST}. These stringent criteria reduced the likelihood of false positives but may also have excluded regions under moderate selection that remain biologically relevant. Relaxing these thresholds could reveal additional candidate regions but would come at the cost of specificity. Future analyses should explore how varying thresholds influence the enrichment results.

A further limitation lies in the statistical framework used for GO term selection. When applying adjusted p-values, no enriched GO terms reached significance (all > 0.1), reflecting the limited number of genes in candidate regions and reduced statistical power. To address this, we relied on the combined score, which integrates both statistical significance (p-value) and deviation from expected rank (z-score). This approach allowed us to highlight potentially meaningful biological signals, even in the absence of strict significance. Nonetheless, future studies could systematically compare thresholds based on adjusted p-values, z-scores, or combined scores, rather than selecting only the top-ranked terms.

iii. Shared regions between approaches

Hopefully, we had three regions that were shared by the two different approaches (F_{ST} and ROH) and only for the lean line. Those regions were placed on chromosome 7 (from 53,945,579bp to 54,945,579bp) chromosome 32 (from 25,952,315bp to 27,007,193bp and from 27,656,469bp to 28,464,650bp). It means that the markers in those regions are in the way of getting fixed for the lean line and that it appears that they are different from the one from the fat line. Unfortunately, after collecting the known genes from those regions and processing enrichment, no GO terms were relevant from our analyses based on lipid metabolism, muscle biology, and energy storage.

Taken together, these results highlight both the management challenges of maintaining genetic diversity in the synthetic line and the biological insights gained from exploring genomic signatures in the fat and lean lines. While the analyses provide consistent evidence for divergent metabolic strategies and point to practical recommendations for conservation and the use of cryopreserved sperm in breeding programs, they also underline important limitations related to statistical power, thresholds, and the feasibility of applying cryoconservation strategies.

5. Conclusion

This study investigated the impact of successive mortality events on the genetic diversity of the INRAE synthetic rainbow trout line, as well as the genomic basis of divergence between experimental lean and fat lines.

Analysis of the 2023 cohort of the synthetic line revealed higher homozygosity compared with the 2010 and 2012 cohorts, indicating a decline in genetic diversity likely caused by both mortality events and genetic drift in this relatively small experimental population. The use of cryopreserved sperm from older cohorts was considered as a possible way to restore diversity. However, our results showed that this strategy would yield only a very limited gain, insufficient to justify its use given both the cost and the scarcity of the available material. It therefore appears more practical to rely on breeders from the current cohort to produce the next generation.
Careful breeding management, as currently implemented, remains essential. Regular monitoring of inbreeding and genetic diversity allows informed decisions on whether cryopreserved sperm should be reintroduced in response to future bottlenecks.
Since the present cohort still retains relatively high diversity, this represents an opportunity to establish a new cryoconserved stock as insurance against future mortality events. Ideally, such a stock should be built from carefully selected individuals to maximize the preservation of rare alleles and genetic originality. At present, however, sperm cryopreservation at the facility is performed opportunistically during reproduction events, generating a largely random cryobank. Genotyping occurs only afterwards, and because individuals in the 2023 cohort are not identified or tagged, it is impossible to retrospectively target promising candidates identified through PCA, F_{ROH}, or IBS analyses. Integrating systematic cryopreservation into the breeding program would therefore help maintain long-term genetic diversity and allow rapid recovery following potential bottlenecks.

In parallel with these management considerations, we also explored the genomic divergence between the experimental lean and fat lines of rainbow trout. These lines, originally derived from the same genetic background but subjected to contrasting selection pressures, provide a unique framework to study the genetic basis of lipid metabolism, energy storage, and muscle physiology.
Our analysis reveals clear differences in genetic diversity between the two lines, reflecting historical selection pressures. Interestingly, the analysis also detected 18 hybrids that were excluded from the analysis showing limitations in the maintenance of these lines only based on phenotypic information.
Using ROH and F_{ST} approaches, we identified candidate regions potentially under selection, although only small regions were shared between the two methods. Several regions overlapped with genes and pathways previously implicated in metabolic regulation, adipocyte differentiation, and intracellular trafficking, supporting their functional relevance. Others may instead reflect linked selection or population-specific drift, and their interpretation must therefore remain cautious.

In conclusion, this study demonstrates the value of combining complementary approaches to detect selection signatures while emphasizing the importance of preserving genetic diversity in experimental lines. These findings contribute to the understanding of the genetic basis of lean versus fat phenotypes

and provide a framework for monitoring and managing genetic variation in breeding programs, ensuring both the persistence of desirable traits and the long-term adaptability of the lines.

Nonetheless, some limitations should be acknowledged. The analyses were conducted on relatively small experimental populations, and some results may therefore be context-dependent rather than universally applicable to aquaculture breeding. In addition, the imperfect separation between the lean and fat lines highlights the practical challenges of maintaining experimental lines over the long term. Future work integrating larger datasets and focusing on functional validation could help refine these findings and assess their relevance for broader breeding applications to produce lines based on their fat content to use them for better smoked filets.

6. Bibliography

AGRESTE, 2024. Enquête aquaculture 2022 | Agreste, la statistique agricole. [en ligne]. 2024. [Consulté le 21 juillet 2025]. Disponible à l'adresse : https://agreste.agriculture.gouv.fr/agreste-web/disaron/Chd2403/detail/

AIZAWA, Katsuji, IEMITSU, Motoyuki, MAEDA, Seiji, JESMIN, Subrina, OTSUKI, Takeshi, MOWA, Chishimba N., MIYAUCHI, Takashi et MESAKI, Noboru, 2007. Expression of steroidogenic enzymes and synthesis of sex steroid hormones from DHEA in skeletal muscle of rats. *American Journal of Physiology-Endocrinology and Metabolism*. février 2007. Vol. 292, n° 2, pp. E577-E584. DOI 10.1152/ajpendo.00367.2006.

AKEY, Joshua M., ZHANG, Ge, ZHANG, Kun, JIN, Li et SHRIVER, Mark D., 2002. Interrogating a high-density SNP map for signatures of natural selection. *Genome Research*. décembre 2002. Vol. 12, n° 12, pp. 1805-1814. DOI 10.1101/gr.631202.

BLANC, Fanny, KERNÉIS, Thierry, COLON, Floriane, BIDEAU, Marjorie, RIGAUDEAU, Dimitri, BRUNEAU, Nicolas, MARCHAND, Michaël, LEGUEN, Isabelle et POUIL, Simon, 2025. *Growth, metabolic, and immune responses of lean and fat rainbow trout lines during a starvation-refeeding cycle*. Manuscrit non publié (soumis). Université Paris-Saclay, INRAE, AgroParisTech, GABI.

BLECHMAN, Janna, AMIR-ZILBERSTEIN, Liat, GUTNICK, Amos, BEN-DOR, Shifra et LEVKOWITZ, Gil, 2011. The metabolic regulator PGC-1α directly controls the expression of the hypothalamic neuropeptide oxytocin. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*. 19 octobre 2011. Vol. 31, n° 42, pp. 14835-14840. DOI 10.1523/JNEUROSCI.1798-11.2011.

BOUDRY, Pierre, ALLAL, François, ASLAM, Muhammad L., BARGELLONI, Luca, BEAN, Tim P., BRARD-FUDULEA, Sophie, BRIEUC, Marine S. O., CALBOLI, Federico C. F., GILBEY, John, HAFFRAY, Pierrick, LAMY, Jean-Baptiste, MORVEZEN, Romain, PURCELL, Catherine, PRODÖHL, Paulo A., VANDEPUTTE, Marc, WALDBIESER, Geoffrey C., SONESSON, Anna K. et HOUSTON, Ross D., 2021. Current status and potential of genomic selection to improve selective breeding in the main aquaculture species of International Council for the Exploration of the Sea (ICES) member countries. *Aquaculture Reports*. 1 juillet 2021. Vol. 20, pp. 100700. DOI 10.1016/j.aqrep.2021.100700.

CÁDIZ, María I., LÓPEZ, María E., DÍAZ-DOMÍNGUEZ, Diego, CÁCERES, Giovanna, MARIN-NAHUELPI, Rodrigo, GOMEZ-UCHIDA, Daniel, CANALES-AGUIRRE, Cristian B., OROZCO-TERWENGEL, Pablo et YÁÑEZ, José M., 2021. Detection of selection signatures in the genome of a farmed population of anadromous rainbow trout (*Oncorhynchus mykiss*). *Genomics*. 1 septembre 2021. Vol. 113, n° 5, pp. 3395-3404. DOI 10.1016/j.ygeno.2021.07.027.

CAI, Yin, YING, Fan et SONG, Erfei, 2015. Mice lacking prostaglandin E receptor subtype 4 manifest disrupted lipid metabolism attributable to impaired triglyceride clearance. *The FASEB Journal* [en ligne]. 2015. [Consulté le 12 août 2025]. DOI 10.1096/fj.15-274597. Disponible à l'adresse : https://faseb.onlinelibrary.wiley.com/doi/abs/10.1096/fj.15-274597

CHAKRABORTY, Sneha et RANNALA, Bruce, 2023. An efficient exact algorithm for identifying hybrids using population genomic sequences. *Genetics*. 6 avril 2023. Vol. 223, n° 4, pp. iyad011. DOI 10.1093/genetics/iyad011.

CHANG, Christopher C, CHOW, Carson C, TELLIER, Laurent CAM, VATTIKUTI, Shashaank, PURCELL, Shaun M et LEE, James J, 2015. Second-generation PLINK: rising to the challenge

of larger and richer datasets. *GigaScience*. 1 décembre 2015. Vol. 4, n° 1, pp. s13742- 015-0047-8. DOI 10.1186/s13742-015-0047-8.

CHARLESWORTH, B. et CHARLESWORTH, D., 1999. The genetic basis of inbreeding depression. *Genetical Research*. décembre 1999. Vol. 74, n° 3, pp. 329-340. DOI 10.1017/s0016672399004152.

CHEN, E, TAN, C et KOU, Y, 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013. Vol. 14, pp. 128. DOI 10.1186/1471-2105-14-128.

CHEN, Xihang, DENG, Zilong, HE, Yunfan, LU, Feng et YUAN, Yi, 2020. Mechanical Strain Promotes Proliferation of Adipose-Derived Stem Cells Through the Integrin β1-Mediated RhoA/Myosin Light Chain Pathway. *Tissue Engineering Part A*. septembre 2020. Vol. 26, n° 17-18, pp. 939-952. DOI 10.1089/ten.tea.2019.0266.

COLON, Floriane, 2022. *La préservation de la diversité génétique et l'amélioration de l'efficacité alimentaire : Deux grands enjeux pour une truitticulture plus durable*. Rapport de stage de fin d'études. AgroParisTech.

D'AMBROSIO, Jonathan, PHOCAS, Florence, HAFFRAY, Pierrick, BESTIN, Anastasia, BRARD-FUDULEA, Sophie, PONCET, Charles, QUILLET, Edwige, DECHAMP, Nicolas, FRASLIN, Clémence, CHARLES, Mathieu et DUPONT-NIVET, Mathilde, 2019. Genome-wide estimates of genetic diversity, inbreeding and effective size of experimental and commercial rainbow trout lines undergoing selective breeding. *Genetics Selection Evolution*. 6 juin 2019. Vol. 51, n° 1, pp. 26. DOI 10.1186/s12711-019-0468-4.

DANECEK, Petr, AUTON, Adam, ABECASIS, Goncalo, ALBERS, Cornelis A., BANKS, Eric, DEPRISTO, Mark A., HANDSAKER, Robert E., LUNTER, Gerton, MARTH, Gabor T., SHERRY, Stephen T., MCVEAN, Gilean et DURBIN, Richard, 2011. The variant call format and VCFtools. *Bioinformatics*. 1 août 2011. Vol. 27, n° 15, pp. 2156-2158. DOI 10.1093/bioinformatics/btr330.

DI PIETRO, Santiago M., FALCÓN-PÉREZ, Juan M., TENZA, Danièle, SETTY, Subba R.G., MARKS, Michael S., RAPOSO, Graça et DELL'ANGELICA, Esteban C., 2006. BLOC-1 Interacts with BLOC-2 and the AP-3 Complex to Facilitate Protein Trafficking on Endosomes. *Molecular Biology of the Cell*. septembre 2006. Vol. 17, n° 9, pp. 4027-4038. DOI 10.1091/mbc.e06-05-0379.

DING, Jiexia, WU, Lili, ZHU, Guoxian, ZHU, Jing, LUO, Pingping et LI, Youming, 2023. HADHA alleviates hepatic steatosis and oxidative stress in NAFLD via inactivation of the MKK3/MAPK pathway. *Molecular Biology Reports*. 1 février 2023. Vol. 50, n° 2, pp. 961-970. DOI 10.1007/s11033-022-07965-2.

DIRECTORATE GENERAL FOR MARITIME AFFAIRS AND FISHERIES (EUROPEAN COMMISSION) et EUMOFA, 2021. *Portion trout in the EU: price structure in the supply chain focus on Germany, Italy and Poland* [en ligne]. Publications Office of the European Union. [Consulté le 21 juillet 2025]. ISBN 978-92-76-28833-6. Disponible à l'adresse : https://data.europa.eu/doi/10.2771/98441KL-01-21-015-EN-N

DOUIRIN, C., HAFFRAY, P., VALLET, J. L. et FAUCONNEAU, B., 1998. Determination of the lipid content of rainbow trout (Oncorhynchus mykiss) fillets with the Torry Fish Fat Meter®. *Sciences des Aliments*. 1998. Vol. 18, n° 5, pp. 527-535.

FAO, 2022. *The State of World Fisheries and Aquaculture 2022* [en ligne]. FAO ; [Consulté le 21 juillet 2025]. ISBN 978-92-5-136364-5. Disponible à l'adresse : https://openknowledge.fao.org/handle/20.500.14283/cc0461en

GAO, Guangtu, MAGADAN, Susana, WALDBIESER, Geoffrey C, YOUNGBLOOD, Ramey C, WHEELER, Paul A, SCHEFFLER, Brian E, THORGAARD, Gary H et PALTI, Yniv, 2021. A long reads-based de-novo assembly of the genome of the Arlee homozygous line reveals chromosomal rearrangements in rainbow trout. *G3 Genes|Genomes|Genetics*. 1 avril 2021. Vol. 11, n° 4, pp. jkab052. DOI 10.1093/g3journal/jkab052.

GUYOMARD, René, BOUSSAHA, Mekki, KRIEG, Francine, HERVET, Caroline et QUILLET, Edwige, 2012. A synthetic rainbow trout linkage map provides new insights into the salmonid whole genome duplication and the conservation of synteny among teleosts. *BMC Genetics*. 16 mars 2012. Vol. 13, n° 1, pp. 15. DOI 10.1186/1471-2156-13-15.

HAFFRAY, P., PINCENT, C., RAULT, P. et COUDURIER, B., 2004. Domestication et amélioration génétique des cheptels piscicoles français dans le cadre du SYSAAF. *INRAE Productions Animales*. 29 juillet 2004. Vol. 17, n° 3, pp. 243-252. DOI 10.20870/productions-animales.2004.17.3.3598.

JACQUES, Alicia, LEROY, Grégoire, ROGNON, Xavier, VERRIER, Etienne, TIXIER-BOICHARD, Michèle et RESTOUX, Gwendal, 2023. Reintroducing genetic diversity in populations from cryopreserved material: the case of Abondance, a French local dairy cattle breed. *Genetics, Selection, Evolution : GSE*. 19 avril 2023. Vol. 55, pp. 28. DOI 10.1186/s12711-023-00801-6.

JEONG, Ju Ah, KO, Kyung-Min et PARK, Hyung Soon, 2007. Membrane proteomic analysis of human mesenchymal stromal cells during adipogenesis. *PROTEOMICS* [en ligne]. 2007. [Consulté le 12 août 2025]. Disponible à l'adresse : https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/pmic.200700502

JOHNSON, Peter, CROSSMAN, James, MILLER, Andy, NICHOLS, Brent, MCLELLAN, Jason, HOWELL, Matthew et SCHREIER, Andrea, 2025. Conservation aquaculture of wild-origin offspring preserves genetic diversity in an endangered population of white sturgeon. *Conservation Genetics*. 1 avril 2025. Vol. 26, n° 2, pp. 335-346. DOI 10.1007/s10592-024-01670-2.

JOMBART, Thibaut, 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics (Oxford, England)*. 1 juin 2008. Vol. 24, n° 11, pp. 1403-1405. DOI 10.1093/bioinformatics/btn129.

KAMALAM, Biju Sam, MEDALE, Françoise, KAUSHIK, Sadasivam, POLAKOF, Sergio, SKIBA-CASSY, Sandrine et PANSERAT, Stephane, 2012. Regulation of metabolism by dietary carbohydrates in two lines of rainbow trout divergently selected for muscle fat content. *The Journal of Experimental Biology*. 1 août 2012. Vol. 215, n° Pt 15, pp. 2567-2578. DOI 10.1242/jeb.070581.

KANAKA, K. K., SUKHIJA, Nidhi, GOLI, Rangasai Chandra, SINGH, Sanjeev, GANGULY, Indrajit, DIXIT, S. P., DASH, Aishwarya et MALIK, Anoop Anand, 2023. On the concepts and measures of diversity in the genomics era. *Current Plant Biology*. 1 janvier 2023. Vol. 33, pp. 100278. DOI 10.1016/j.cpb.2023.100278.

KIJAS, James W., LENSTRA, Johannes A., HAYES, Ben, BOITARD, Simon, NETO, Laercio R. Porto, CRISTOBAL, Magali San, SERVIN, Bertrand, MCCULLOCH, Russell, WHAN, Vicki, GIETZEN, Kimberly, PAIVA, Samuel, BARENDSE, William, CIANI, Elena, RAADSMA,

Herman, MCEWAN, John, DALRYMPLE, Brian et CONSORTIUM, other members of the International Sheep Genomics, 2012. Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PLOS Biology*. 7 février 2012. Vol. 10, n° 2, pp. e1001258. DOI 10.1371/journal.pbio.1001258.

KIRKPATRICK, S., GELATT, C. D. et VECCHI, M. P., 1983. Optimization by Simulated Annealing. *Science*. 13 mai 1983. Vol. 220, n° 4598, pp. 671-680. DOI 10.1126/science.220.4598.671.

KULESHOV, M, JONES, M et ROUILLARD, A, 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*. 2016. Vol. 44. DOI 10.1093/nar/gkw377.

LAGARDE, H., PHOCAS, F., POUIL, S., GOARDON, L., BIDEAU, M., GUYVARC'H, F., LABBÉ, L., DECHAMP, N., PRCHAL, M., DUPONT-NIVET, M. et LALLIAS, D., 2023. Are resistances to acute hyperthermia or hypoxia stress similar and consistent between early and late ages in rainbow trout using isogenic lines? *Aquaculture*. 15 janvier 2023. Vol. 562, pp. 738800. DOI 10.1016/j.aquaculture.2022.738800.

LALLIAS, Delphine, BERNARD, Maria, CIOBOTARU, Céline, DECHAMP, Nicolas, LABBÉ, Laurent, GOARDON, Lionel, LE CALVEZ, Jean-Michel, BIDEAU, Marjorie, FRICOT, Alexandre, PRÉZELIN, Audrey, CHARLES, Mathieu, MOROLDO, Marco, COUSIN, Xavier, BOUCHEZ, Olivier, ROULET, Alain, QUILLET, Edwige et DUPONT-NIVET, Mathilde, 2021. Sources of variation of DNA methylation in rainbow trout: combined effects of temperature and genetic background. *Epigenetics*. septembre 2021. Vol. 16, n° 9, pp. 1031-1052. DOI 10.1080/15592294.2020.1834924.

LEFÈVRE, Florence, BUGEON, Jérôme, GOARDON, Lionel, KERNEIS, Thierry, LABBÉ, Laurent, PANSERAT, Stéphane, MÉDALE, Françoise et QUILLET, Edwige, 2024. Seven generations of selection for muscle fat content greatly affect rainbow trout flesh quality and muscle fiber size. *Aquaculture Reports*. octobre 2024. Vol. 38, pp. 102343. DOI 10.1016/j.aqrep.2024.102343.

LEFEVRE, Florence, CARDINAL, Mireille, BUGEON, Jérôme, LABBE, Laurent, MEDALE, Françoise et QUILLET, Edwige, 2015. Selection for muscle fat content and triploidy affect flesh quality in pan-size rainbow trout, *Oncorhynchus mykiss*. *Aquaculture*. 1 novembre 2015. Vol. 448, pp. 569-577. DOI 10.1016/j.aquaculture.2015.06.029.

LIANG, Junzhi, ZHANG, Bowen, HU, Yannan, NA, Zhijing et LI, Da, 2023. Effects of steroid hormones on lipid metabolism in sexual dimorphism: A Mendelian randomization study. *Frontiers in Endocrinology* [en ligne]. 16 janvier 2023. Vol. 13. [Consulté le 12 août 2025]. DOI 10.3389/fendo.2022.1119154. Disponible à l'adresse : https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2022.1119154/full

LU, Huijie, CUI, Yong, JIANG, Liwen et GE, Wei, 2017. Functional Analysis of Nuclear Estrogen Receptors in Zebrafish Reproduction by Genome Editing Approach. *Endocrinology*. 1 juillet 2017. Vol. 158, n° 7, pp. 2292-2308. DOI 10.1210/en.2017-00215.

LUNDRIGAN, Tricia, REIST, Jim et FERGUSON, Moira, 2005. Microsatellite variation within and among Arctic charr (Salvelinus alpinus) from aquaculture and natural populations in North America. *Aquaculture*. 1 février 2005. Vol. 244, pp. 63-75. DOI 10.1016/j.aquaculture.2004.11.027.

MAIORANO, AM, LOURENCO, DL, TSURUTA, S, OSPINA, AMT, STAFUZZA, NB et MASUDA, Y, 2018. Assessing genetic architecture and signatures of selection of dual purpose

Gir cattle populations using genomic information | PLOS One. [en ligne]. 2018. Disponible à l'adresse : https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0200694

MCQUILLAN, Ruth, LEUTENEGGER, Anne-Louise, ABDEL-RAHMAN, Rehab, FRANKLIN, Christopher S., PERICIC, Marijana, BARAC-LAUC, Lovorka, SMOLEJ-NARANCIC, Nina, JANICIJEVIC, Branka, POLASEK, Ozren, TENESA, Albert, MACLEOD, Andrew K., FARRINGTON, Susan M., RUDAN, Pavao, HAYWARD, Caroline, VITART, Veronique, RUDAN, Igor, WILD, Sarah H., DUNLOP, Malcolm G., WRIGHT, Alan F., CAMPBELL, Harry et WILSON, James F., 2008. Runs of homozygosity in European populations. *American Journal of Human Genetics*. septembre 2008. Vol. 83, n° 3, pp. 359-372. DOI 10.1016/j.ajhg.2008.08.007.

NIELSEN, Rasmus, 2005. Molecular signatures of natural selection. *Annual Review of Genetics*. 2005. Vol. 39, pp. 197-218. DOI 10.1146/annurev.genet.39.073003.112420.

O'BRIEN, Jacob, HAYDER, Heyam, ZAYED, Yara et PENG, Chun, 2018. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Frontiers in Endocrinology*. 3 août 2018. Vol. 9, pp. 402. DOI 10.3389/fendo.2018.00402.

PALTI, Y., GAO, G., LIU, S., KENT, M. P., LIEN, S., MILLER, M. R., REXROAD, C. E. et MOEN, T., 2015. The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. *Molecular Ecology Resources*. mai 2015. Vol. 15, n° 3, pp. 662-672. DOI 10.1111/1755-0998.12337.

PAUL, Katy, 2022. *Evolution de la diversité génétique dans des populations domestiquées de truites arc-en-ciel* [en ligne]. These de doctorat. université Paris-Saclay. [Consulté le 24 avril 2025]. Disponible à l'adresse : https://theses.fr/2022UPASB067

PEMBERTON, Trevor J., ABSHER, Devin, FELDMAN, Marcus W., MYERS, Richard M., ROSENBERG, Noah A. et LI, Jun Z., 2012. Genomic patterns of homozygosity in worldwide human populations. *American Journal of Human Genetics*. 10 août 2012. Vol. 91, n° 2, pp. 275-292. DOI 10.1016/j.ajhg.2012.06.014.

POUIL, Simon, HARTÉ, Elodie, MARCHAND, Michaël, GOARDON, Lionel, TERRENNE, Tina, LABBÉ, Laurent, DUPONT-NIVET, Mathilde, LEGUEN, Isabelle et LALLIAS, Delphine, 2024. Cold temperature during embryonic development and its influence on responses to acute confinement and hyperthermia challenges in juvenile rainbow trout. *Aquaculture*. 15 décembre 2024. Vol. 593, pp. 741318. DOI 10.1016/j.aquaculture.2024.741318.

POUIL, Simon, KERNÉIS, Thierry, QUILLET, Edwige, LABBÉ, Laurent, LALLIAS, Delphine, PHOCAS, Florence et DUPONT-NIVET, Mathilde, 2023. Isogenic lines of rainbow trout (*Oncorhynchus mykiss)* as a tool to assess how growth and feeding behaviour are correlated to feed efficiency in fish. *Aquaculture*. 15 décembre 2023. Vol. 577, pp. 739904. DOI 10.1016/j.aquaculture.2023.739904.

QANBARI, Saber, GIANOLA, Daniel, HAYES, Ben, SCHENKEL, Flavio, MILLER, Steve, MOORE, Stephen, THALLER, Georg et SIMIANER, Henner, 2011. Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. *BMC Genomics*. 16 juin 2011. Vol. 12, n° 1, pp. 318. DOI 10.1186/1471-2164-12-318.

QUILLET, Edwige, LE GUILLOU, Sandrine, AUBIN, Joël et FAUCONNEAU, Benoît, 2005. Two-way selection for muscle lipid content in pan-size rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*. 4 mars 2005. Vol. 245, n° 1, pp. 49-61. DOI 10.1016/j.aquaculture.2004.12.014.

QUILLET, Edwige, LE GUILLOU, Sandrine, AUBIN, Joël, LABBÉ, Laurent, FAUCONNEAU, Benoît et MÉDALE, Françoise, 2007. Response of a lean muscle and a fat muscle rainbow trout (*Oncorhynchus mykiss*) line on growth, nutrient utilization, body composition and carcass traits when fed two different diets. *Aquaculture*. 14 septembre 2007. Vol. 269, n° 1, pp. 220-231. DOI 10.1016/j.aquaculture.2007.02.047.

R CORE TEAM, 2022. *R: A language and environment for statistical computing* [en ligne]. 2022. Vienna, Austria. Disponible à l'adresse : https://www.R-project.org/

RAMADORI, G., VEIT, T., SCHWÖGLER, S., DIENES, H. P., KNITTEL, T., RIEDER, H. et MEYER ZUM BÜSCHENFELDE, K. -H., 1990. Expression of the gene of the α-smooth muscle-actin isoform in rat liver and in rat fat-storing (ITO) cells. *Virchows Archiv B*. 1 décembre 1990. Vol. 59, n° 1, pp. 349-357. DOI 10.1007/BF02899424.

REBELATO, Arnaldo Basso et CAETANO, Alexandre Rodrigues, 2018. Runs of homozygosity for autozygosity estimation and genomic analysis in production animals. *Pesquisa Agropecuária Brasileira*. septembre 2018. Vol. 53, pp. 975-984. DOI https://doi.org/10.1590/S0100-204X2018000900001.

RSTUDIO TEAM, 2022. *RStudio: Integrated Development Environment for R* [en ligne]. 2022. Boston, MA. Disponible à l'adresse : http://www.rstudio.com/

SEBO, Zachary L., JEFFERY, Elise, HOLTRUP, Brandon et RODEHEFFER, Matthew S., 2018. A mesodermal fate map for adipose tissue. *Development*. 17 août 2018. Vol. 145, n° 17, pp. dev166801. DOI 10.1242/dev.166801.

SEKINO, Masashi, HARA, Motoyuki et TANIGUCHI, Nobuhiko, 2002. Loss of microsatellite and mitochondrial DNA variation in hatchery strains of Japanese flounder *Paralichthys olivaceus*. *Aquaculture*. 18 octobre 2002. Vol. 213, n° 1, pp. 101-122. DOI 10.1016/S0044-8486(01)00885-7.

SHAHIN, Khalid, VEEK, Tresa, HECKMAN, Taylor I., LITTMAN, Eric, MUKKATIRA, Kaveramma, ADKISON, Mark, WELCH, Timothy J., IMAI, Denise M., PASTENKOS, Gabrielle, CAMUS, Alvin et SOTO, Esteban, 2022. Isolation and characterization of Lactococcus garvieae from rainbow trout, Onchorhyncus mykiss, from California, USA. *Transboundary and Emerging Diseases*. juillet 2022. Vol. 69, n° 4, pp. 2326-2343. DOI 10.1111/tbed.14250.

TABARIN, Antoine, DIZ-CHAVES, Yolanda, CARMONA, Maria del Carmen, CATARGI, Bogdan, ZORRILLA, Eric P., ROBERTS, Amanda J., COSCINA, Donald V., ROUSSET, Sophie, REDONNET, Anabelle, PARKER, Graham C., INOUE, Koki, RICQUIER, Daniel, PÉNICAUD, Luc, KIEFFER, Brigitte L. et KOOB, Georges F., 2005. Resistance to Diet-Induced Obesity in μ-Opioid Receptor–Deficient Mice : Evidence for a "Thrifty Gene". *Diabetes*. 1 décembre 2005. Vol. 54, n° 12, pp. 3510-3516. DOI 10.2337/diabetes.54.12.3510.

WANG, Jinliang, 2005. Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 13 juillet 2005. Vol. 360, n° 1459, pp. 1395-1409. DOI 10.1098/rstb.2005.1682.

WANG, Le, SHI, Xiaofeng, SU, Yongquan, MENG, Zining et LIN, Haoran, 2012. Loss of genetic diversity in the cultured stocks of the large yellow croaker, Larimichthys crocea, revealed by microsatellites. *International Journal of Molecular Sciences*. 2012. Vol. 13, n° 5, pp. 5584-5597. DOI 10.3390/ijms13055584.

WATT, Matthew J., HEIGENHAUSER, George J. F., O'NEILL, Marcus et SPRIET, Lawrence L., 2003. Hormone-sensitive lipase activity and fatty acyl-CoA content in human skeletal

muscle during prolonged exercise. *Journal of Applied Physiology*. juillet 2003. Vol. 95, n° 1, pp. 314-321. DOI 10.1152/japplphysiol.01181.2002.

WICKHAM, Hadley, 2016. *ggplot2: Elegant Graphics for Data Analysis* [en ligne]. Springer-Verlag New York. ISBN 978-3-319-24277-4. Disponible à l'adresse : https://ggplot2.tidyverse.org/

WRIGHT, S, 1922. Coefficients of inbreeding and relationship. *The American Naturalist*. 1922. Vol. 56, n° 645, pp. 330-338. DOI https://doi.org/10.1086/279872.

XU, Zhong, SUN, Hao, ZHANG, Zhe, ZHAO, Qingbo, OLASEGE, Babatunde Shittu, LI, Qiumeng, YUE, Yang, MA, Peipei, ZHANG, Xiangzhe, WANG, Qishan et PAN, Yuchun, 2019. Assessment of Autozygosity Derived From Runs of Homozygosity in Jinhua Pigs Disclosed by Sequencing Data. *Frontiers in Genetics*. 2019. Vol. 10, pp. 274. DOI 10.3389/fgene.2019.00274.

ZHANG, Yanjun, BAI, Guangdong, LI, Xue, NEPAL, Surya, GROBLER, Marthie, CHEN, Chen et KO, Ryan K. L., 2023. Preserving Privacy for Distributed Genome-Wide Analysis Against Identity Tracing Attacks. *IEEE Transactions on Dependable and Secure Computing*. 1 juillet 2023. Vol. 20, n° 04, pp. 3341-3357. DOI 10.1109/TDSC.2022.3186672.

ZHAO, Fuping, MCPARLAND, Sinead, KEARNEY, Francis, DU, Lixin et BERRY, Donagh P., 2015. Detection of selection signatures in dairy and beef cattle using high-density genomic information. *Genetics Selection Evolution*. 19 juin 2015. Vol. 47, n° 1, pp. 49. DOI 10.1186/s12711-015-0127-3.

7. Annexes

Annex I: Formula for the variance of genetic diversity

The variance of heterozygosity of the future cohort could be calculated by this formula:

$$Var_{div}(Next\_Cohort) = (1-p)^2 * Var_{div}(Current\_Cohort) + p^2 * Var_{div}(Cryoconserved\_Straws)$$

With:
p = proportion on cryo-conserved genetic material used for the reproduction of the next cohort in a mating plan for 100 males used with equal amount of sperm for each of them to contribute to the next generation.

$$Var_{div(Y)} = Var(1 - F_{ROH_i})$$

Where:

- Y is either the genotyped individuals of the current cohort or the cryoconserved straws selected
- n is the number of individuals considered for the cohort
- F_{ROH_i} is the coefficient of inbreeding for the individual i and $1 - F_{ROH_i}$ is an indicator of heterozygosity

Annex II: Gene Ontology (GO) terms significantly enriched in genomic regions identified through Runs of Homozygosity (ROH) analysis in the Lean line of rainbow trout

GO Molecular Function 2018					
Index	Name	P-value	Adjusted p-value	Z-score	Combined score
1	RNA polymerase II transcription factor activity, sequence-specific transcription regulatory region DNA binding (GO:0001133)	0.001203	0.1432	-2.59	17.40
2	opioid receptor activity (GO:0004985)	0.05424	0.3411	-4.84	14.11
3	polynucleotide adenylyltransferase activity (GO:0004652)	0.05424	0.3411	-4.55	13.25
4	translation release factor activity (GO:0003747)	0.05424	0.3411	-4.43	12.90
5	histone demethylase activity (H3-trimethyl-K4 specific) (GO:0034647)	0.05424	0.3411	-4.15	12.08
6	carbon-nitrogen ligase activity, with glutamine as amido-N-donor (GO:0016884)	0.07166	0.3411	-4.19	11.03
7	ligand-dependent nuclear receptor transcription coactivator activity (GO:0030374)	0.00410	0.1626	-2.00	10.99
8	lysine-acetylated histone binding (GO:0070577)	0.06299	0.3411	-3.92	10.83
9	RNA polymerase II transcription coactivator activity (GO:0001105)	0.07166	0.3411	-3.90	10.28
10	retinol dehydrogenase activity (GO:0004745)	0.05424	0.3411	-3.38	9.86

GO Biological Process 2018					
Index	Name	P-value	Adjusted p-value	Z-score	Combined score
1	production of miRNAs involved in gene silencing by miRNA (GO:0035196)	0.003647	0.3832	-3.03	17.01
2	anterior lateral line development (GO:0048899)	0.07166	0.3832	-5.67	14.93
3	negative regulation of histone methylation (GO:0031061)	0.05424	0.3832	-4.88	14.24
4	spinal cord development (GO:0021510)	0.00719	0.3832	-2.87	14.17
5	primary neural tube formation (GO:0014020)	0.05424	0.3832	-4.73	13.78
6	tube closure (GO:0060606)	0.06299	0.3832	-4.93	13.62
7	protein K11-linked deubiquitination (GO:0035871)	0.05424	0.3832	-4.61	13.42
8	cerebellar granule cell differentiation (GO:0021707)	0.05424	0.3832	-4.60	13.42
9	pigment cell development (GO:0070285)	0.06299	0.3832	-4.77	13.20
10	mRNA polyadenylation (GO:0006378)	0.008256	0.3832	-2.55	12.23

GO Cellular Component 2018					
Index	Name	P-value	Adjusted p-value	Z-score	Combined score
1	mRNA cleavage and polyadenylation specificity factor complex (GO:0005847)	0.08025	0.4372	-3.55	8.95
2	cytoplasmic stress granule (GO:0010494)	0.08025	0.4372	-3.24	8.18
3	core mediator complex (GO:0070847)	0.07166	0.4372	-3.05	8.05
4	THO complex part of transcription export complex (GO:0000445)	0.07166	0.4372	-2.82	7.43
5	BLOC-1 complex (GO:0031083)	0.09720	0.4372	-3.12	7.27
6	RNA polymerase II transcription factor complex (GO:0090575)	0.005937	0.3206	-1.25	6.42
7	ESC/E(Z) complex (GO:0035098)	0.1056	0.4372	-2.67	6.00
8	mitochondrial intermembrane space (GO:0005758)	0.1382	0.4372	-2.66	5.27
9	chromosomal region (GO:0098687)	0.1302	0.4372	-2.54	5.18
10	site of double-strand break (GO:0035861)	0.1462	0.4372	-2.62	5.04

Annex III: Gene Ontology (GO) terms significantly enriched in genomic regions identified through Runs of Homozygosity (ROH) analysis in the Fat line of rainbow trout

GO Molecular Function 2018					
Index	Name	P-value	Adjusted p-value	Z-score	Combined score
1	estrogen receptor binding (GO:0030331)	0.04893	0.2224	-5.80	17.49
2	NADPH-hemoprotein reductase activity (GO:0003958)	0.04893	0.2224	-4.72	14.25
3	mannose binding (GO:0005537)	0.04294	0.2224	-4.33	13.64
4	GABA receptor activity (GO:0016917)	0.008997	0.2224	-2.41	11.36
5	GABA-A receptor activity (GO:0004890)	0.008997	0.2224	-2.39	11.24
6	steroid dehydrogenase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor (GO:0033764)	0.04893	0.2224	-3.42	10.32
7	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen (GO:0016709)	0.01142	0.2224	-2.26	10.13
8	prostaglandin E receptor activity (GO:0004957)	0.05487	0.2224	-3.44	9.98
9	oxidoreductase activity, acting on NAD(P)H, heme protein as acceptor (GO:0016653)	0.07249	0.2224	-3.76	9.86
10	deoxyribonuclease activity (GO:0004536)	0.06078	0.2224	-3.51	9.82

GO Biological Process 2018					
Index	Name	P-value	Adjusted p-value	Z-score	Combined score
1	activation of MAPKK activity (GO:0000186)	0.002054	0.1089	-4.32	26.73
2	positive regulation of developmental growth (GO:0048639)	0.002889	0.1089	-3.50	20.47
3	negative regulation of blood pressure (GO:0045776)	0.03692	0.2346	-5.70	18.79
4	positive regulation of peptidyl-tyrosine phosphorylation (GO:0050731)	0.003858	0.1089	-3.25	18.06
5	neuron projection fasciculation (GO:0106030)	0.002889	0.1089	-3.05	17.82
6	regulation of intracellular estrogen receptor signaling pathway (GO:0033146)	0.03692	0.2346	-5.24	17.28
7	regulation of neuron migration (GO:2001222)	0.03692	0.2346	-5.18	17.09
8	positive regulation of synapse assembly (GO:0051965)	0.04294	0.2346	-5.30	16.70
9	regulation of focal adhesion assembly (GO:0051893)	0.04294	0.2346	-5.29	16.64
10	regulation of peptidyl-tyrosine phosphorylation (GO:0050730)	0.003357	0.1089	-2.81	15.99

GO Cellular Component 2018					
Index	Name	P-value	Adjusted p-value	Z-score	Combined score
1	stereocilia ankle link (GO:0002141)	0.03692	0.2780	-4.39	14.49
2	GABA-A receptor complex (GO:1902711)	0.008997	0.2780	-2.95	13.90
3	microtubule minus-end (GO:0036449)	0.04294	0.2780	-3.69	11.60
4	anchored component of external side of plasma membrane (GO:0031362)	0.04294	0.2780	-3.45	10.85
5	intrinsic component of external side of plasma membrane (GO:0031233)	0.04294	0.2780	-3.25	10.22
6	contractile actin filament bundle (GO:0097517)	0.04893	0.2780	-3.30	9.95
7	actomyosin (GO:0042641)	0.05487	0.2780	-3.36	9.75
8	stress fiber (GO:0001725)	0.04893	0.2780	-2.25	6.80
9	microtubule end (GO:1990752)	0.07829	0.2780	-2.62	6.67
10	preribosome, large subunit precursor (GO:0030687)	0.1289	0.3536	-2.52	5.16

Annex IV: Gene Ontology (GO) terms significantly enriched in genomic regions identified through FST-based genomic differentiation between Lean and Fat lines of rainbow trout

GO Molecular Function 2018					
Index	Name	P-value	Adjusted p-value	Z-score	Combined score
1	opioid receptor activity (GO:0004985)	0.002050	0.1079	-4.97	30.78
2	enhancer sequence-specific DNA binding (GO:0001158)	0.001633	0.1079	-2.98	19.12
3	RNA polymerase II distal enhancer sequence-specific DNA binding (GO:0000980)	0.0008310	0.1079	-2.67	18.92
4	phosphatase activator activity (GO:0019211)	0.005959	0.1883	-2.94	15.06
5	protein phosphatase activator activity (GO:0072542)	0.007226	0.1903	-2.89	14.25
6	endodeoxyribonuclease activity (GO:0004520)	0.01337	0.3018	-2.90	12.53
7	kinase activator activity (GO:0019209)	0.06932	0.3901	-4.54	12.11
8	protein phosphatase binding (GO:0019903)	0.002800	0.1106	-2.06	12.10
9	U6 snRNA binding (GO:0017070)	0.08040	0.3901	-4.46	11.25
10	palmitoyl-CoA hydrolase activity (GO:0016290)	0.06932	0.3901	-4.21	11.24

GO Biological Process 2018					
Index	Name	P-value	Adjusted p-value	Z-score	Combined score
1	tube closure (GO:0060606)	0.002847	0.3343	-5.11	29.94
2	primary neural tube formation (GO:0014020)	0.002050	0.3327	-4.83	29.89
3	neural tube closure (GO:0001843)	0.005959	0.3627	-3.72	19.08
4	regulation of mesodermal cell fate specification (GO:0042661)	0.004804	0.3343	-3.51	18.74
5	rhombomere 4 morphogenesis (GO:0021661)	0.06932	0.4269	-6.83	18.22
6	enteric nervous system development (GO:0048484)	0.0003198	0.1164	-2.06	16.60
7	response to steroid hormone (GO:0048545)	0.004804	0.3343	-3.02	16.14
8	L-amino acid transport (GO:0015807)	0.06932	0.4269	-5.53	14.75
9	microtubule depolymerization (GO:0007019)	0.008604	0.4269	-3.04	14.44
10	positive regulation of transporter activity (GO:0032411)	0.06932	0.4269	-5.24	13.98

GO Cellular Component 2018					
Index	Name	P-value	Adjusted p-value	Z-score	Combined score
1	Rad51B-Rad51C-Rad51D-XRCC2 complex (GO:0033063)	0.08040	0.4868	-5.42	13.66
2	mitochondrial proton-transporting ATP synthase complex, coupling factor F(o) (GO:0000276)	0.08040	0.4868	-3.88	9.79
3	spliceosomal tri-snRNP complex (GO:0097526)	0.04072	0.4868	-2.59	8.30
4	actin filament (GO:0005884)	0.007586	0.4868	-1.65	8.04
5	microvillus (GO:0005902)	0.1022	0.4868	-3.52	8.04
6	mRNA cleavage and polyadenylation specificity factor complex (GO:0005847)	0.1022	0.4868	-3.52	8.02
7	mitochondrial proton-transporting ATP synthase complex (GO:0005753)	0.09134	0.4868	-3.16	7.56
8	condensed chromosome (GO:0000793)	0.1234	0.4868	-3.59	7.50
9	U6 snRNP (GO:0005688)	0.09134	0.4868	-3.05	7.30
10	BLOC-1 complex (GO:0031083)	0.1234	0.4868	-3.09	6.47

Annex V: Genes identified in genomic regions highlighted by Runs of Homozygosity (ROH) analysis in the lean line of rainbow trout

aadacl4	abcf1	adck5	agr2	alpl
ANLN	APBB2	aqp4	arid1ab	arl4ab
artna	asns	atf1	atf7b	atxn7
barhl2	bicc2	bloc1s4	brdt	bzw2
C5orf49	cadpsa	camk1b	ccdc3b	cdc42ep4b
cdh2	chrna9	cited4a	cpsf1	ctdp1
ctps1a	dhrs3b	dip2bb	dipk1ab	dtnbp1b
ece1	edn1	edn2	EEPD1	eif4g3a
ephb2b	ephx4	etf1a	eva1bb	EVI5
eya3	faim2b	FAM135B	fbxw11a	FEZF2
fgf18a	foxo6b	gadd45ab	gfi1ab	glcci1a
glt8d1	glut1a	gmeb1	gnl3	gs01
hivep1	hivep3a	hp1bp3	hs3st3b1b	HSBP1
igsf21a	irx1b	irx4b	jarid2b	kcng2
KCNK9	kctd1	kdf1b	kdm4aa	KHDRBS3
kif17	klhdc7a	lactbl1b	larp4ab	lin28a
LPCAT1	lrig2	lrrc38b	LRRC3B	magi3b
map7d1b	MATCAP2	med10	med8	meox2b
mrps18b	mtf	myo1g	ndufs6	nedd9
nek4	NGLY1	nkiras1	nmur3	nr0b2b
nr1d2b	nr1d4b	NSUN7	ntf2	ntl
nudc	oprd1a	oscp1a	pad	paqr7a
pard6gb	pbx1b	pcdh2ac	pcdhb	pdik1l
phactr4a	phc2b	phf14	ppp1r10	ppp1r8a
prickle2b	psd6	psma8	ptprfa	ptprga
ptprua	rab42a	RARB	rbms2a	rpa3
rpl15	RPL5	scin	sdhaf3	sdhaf3
sdk2b	SEPTIN7	serinc2	sf3a3	SH2D5
sh3bgrl3	si:ch1073-296d18.1	si:ch1073-342h5.2	si:ch211-195b13.1	si:ch211-215a10.4
si:ch211-286o17.1	si:ch73-233f7.1	si:ch73-379j16.2	si:ch73-379j16.2	si:ch73-389b16.2
si:dkey-202e17.1	si:dkey-202e17.1	si:dkeyp-92c9.2	SLC66A2	smim13
SNORD69	sostdc1b	spcs1	spock1	srgap3
ST3GAL1	st3gal3a	stmn1a	SYNPR	tac1
tent4a	tent4a	tgfbr3	thoc7	thrap3b
thrb	thsd7aa	tlr1	tmem170b	tmem222b
top2b	tpbga	tspan13b	txnl4a	U4
U7	U7	ube2ka	ube2ql1	ube2ql1
umad1	Vault	Vault	yod1	zdhhc18b
zgc:158258	znf644b			

Annex VI: Genes identified in genomic regions highlighted by Runs of Homozygosity (ROH) analysis in the fat line of rainbow trout

abca2	adamts12	agpat9l	ajap1	AKIRIN2
ANKRD6	AOPEP	araf	avp	b3gnt7l
bach2b	bag4	c7a	camsap1b	casp8ap2
cd53	cdk11b	CFAP206	cnr1	col27a1b
coq5	creb3l3l	cx52.6	dffb	dock8
dph7	efna5b	elmod3	elovl7a	epha7
ercc6l2	fabp1b.1	fancc	fbxl17	fbxo21
fbxo4	fbxw8	fgf10b	fgf10b	foxd5
gabrr1	gabrr2b	gadd45ga	ghra	gk
grk5l	hsd17b3	hspb8	HTR1B	HTR1E
htra4	imp4	ipo11	ksr2	lman2lb
loxhd1b	lrrc47	lrrc74b	LSM1	lzts3b
map3k7	MIB2	MIR23B	MIR455	MMP23B
MRPL41	mrps27	NADK	NDUFAF2	nf2a
niban2b	nim1k	nos1	noxa1	npdc1b
nsmfb	nt5e	oxct1a	paip1	pgm5
plcxd3	pnrc1	PTCH1	ptger4a	rabl6b
RFC5	rgs7bpa	RIMOC1	rnf165b	rnft2
rngtt	SELENOP	si:ch1073-459j12.1	si:ch211-241j12.3	si:dkey-1k23.3
SLC35A1	slc35d2	SLC35E2B	smim15	SMIM8
snx14	SPRING1	srrm4	star	stxbp1b
suds3	SYNCRIP	taok3a	tbx3b	tesca
tm2d2	tnc	traf2a	ube2j1	ubiad1
ubox5	VSIG10	whrnb	WSB2	xbp1
zgc:162952	znf131	znf292b	znf366	znf367
ZSWIM6				

Annex VII: Genes identified in genomic regions showing FST-based genomic differentiation between fat and lean lines of rainbow trout

aadacl4	ABRAXAS2	acot11b	adck5	aldh18a1
aldh4a1	anxa5a	AP1M1	aqp8a.1	arf2b
arhgap17a	arhgef10la	arhgef1a	arid1aa	arid1ab
arl4d	atcaya	atp5mc1	atpaf1	axdnd1
b3gat2	BBS7	BFSP1	bloc1s4	bricd5
btf3l4	calr3a	casc3	ccl25b	CCNA2
CEP170	cers1	cfap20	chd5	cherp
chst15	chtopa	clptm1l	comp	cpped1
cpsf1	crygmx	crygmxl2	CTBP2	ctdp1
ctns	ddx49	dennd3a	dgat1a	dhrs3b
dmc1	dtnbp1b	edn2	EHF	eif1
eif1	elovl1b	EPS15L1	ercc4	eve1
FAM163A	FAM53B	fgfbp1b	fgfbp2a	foxj1b
FSTL5	gpn2	gpr26	grb7	grid2ipb
grnb	her2	hgh1	hivep3a	hivep3b
hmx2	hmx3a	hoxb13a	hoxb1a	hoxb2a
HOXB3	hoxb4a	hoxb5a	hoxb6a	hoxb7a
hoxb8a	hoxb9a	HSBP1	igf2bp1	igsf21a
ikzf5	ints3	jarid2b	jtb	JUND
jupb	kcnab2a	kcng2	kdf1b	klf2a
klhdc7a	klhl26	lhx8b	loxl5a	lpl
lrrc38b	lrriq3	lsm4	matk	megf6b
MIR9-1	mknk1	mlst8	mob3c	mrpl54
mrtfba	mrto4	msl1a	msrb1b	mydgf
nbr1a	ncln	neurl2	NKX1-2	noxo1b
npr1a	nr0b2b	nr1d1	ntf2	ntl
nudc	ogfrl1	oprd1a	oprd1b	OTOR
pamr1	paqr7a	paqr7b	pard6gb	park7
pcdh15a	pcsk2	pdhx	pdia8	pex11b
pgp	PGPEP1	phactr4a	phactr4b	pigv
plek	pole4	prf1.2	prf1.3	prokr1b
prom1a	psme3	ptgfr	ptk2aa	QDPR
rab11fip3	rab3ab	rab3b	rad23ab	rapgefl1
rcc1	rhbdf1b	rit1	RNF207	rpl22
rplp1	RS27	rx1	s100v2	s100v2
safb	scrt1a	scxa	sema4ab	serinc2
serinc2l	sh3bgrl3	shisa9a	si:ch1073-296d18.1	si:ch1073-396h14.1
si:ch211-1a19.3	si:ch211-248l17.3	si:ch73-109i22.2	si:dkey-222b8.1	si:dkey-225f5.4
si:dkey-240h12.4	si:dkey-27c15.3	si:dkey-89b17.4	si:dkeyp-113d7.1	slc1a2b
slc1a6	slc44a5b	SLC66A2	smap1	snapin
snx29	sost	ssbp3b	ssrb	ST3GAL1
st6galnac5b	stmn1a	stmn1b	stx1b	syt11a
tal1	tapt1a	tax1bp3	tent5ba	tex2l
tmem106a	tmem204	tmem222b	TMEM33	tnfaip8l1
tnni3k	tpbga	tpm3	tpm4a	txnl4a
U7	ube2z	wapla	wipf2b	xcr1a.1
ZBTB18	zdhhc18a	zdhhc18b	zdhhc4	zfr2
zfyve9b	zgc:163057	zgc:6386table3		

L'INSTITUT agro Rennes Angers	Diplôme : Ingénieur Spécialité : Agronome Spécialisation / option : SHA (AQUA) Enseignant référent : Bastien SADOUL

Auteur(s) : Gatien DEBAILLEUL Date de naissance* : 17/07/2000	Organisme d'accueil : INRAE Ile de France Jouy-en-Josas - Anthony Adresse : Domaine du Vilvert | 78352 Jouy-en-Josas Maître de stage : Florence PHOCAS
Nb pages : 29 Annexe(s) : 7	
Année de soutenance : 2025	

Titre français : Diversité génétique de la lignée synthétique et signatures de sélection des lignées grasse et maigre INRAE de truite arc-en-ciel (*Oncorhynchus mykiss*)

Titre anglais: Genetic diversity of the synthetic line and signatures of selection of fat and lean lines of INRAE rainbow trout (*Oncorhynchus mykiss*)

Résumé (1600 caractères maximum) :

Ce mémoire analyse la diversité génétique et les signatures de sélection dans des lignées expérimentales de truite arc-en-ciel (*Oncorhynchus mykiss*) développées par l'INRAE. La première partie évalue la diversité de la lignée synthétique, créée pour conserver un réservoir génétique. Les résultats montrent une légère érosion liée à deux crises sanitaires majeures, mais une diversité encore intéressante. L'utilisation de semences cryoconservées de cohortes anciennes ne permettrait qu'un faible gain (0,86 %) et épuiserait presque le stock disponible, rendant cette stratégie peu pertinente à court terme. Il est recommandé d'établir un nouveau stock cryoconservé à partir de la cohorte actuelle, afin d'assurer une résilience face à de futurs goulots d'étranglement.

La seconde partie porte sur les lignées divergentes « grasse » et « maigre », sélectionnées pendant huit générations sur la teneur en lipides musculaires. L'analyse génomique (ROH, F_{ST}) met en évidence des régions sous sélection liées au métabolisme lipidique, à la régulation hormonale, à la plasticité métabolique et à la structure musculaire. La lignée maigre se caractérise par une meilleure efficacité alimentaire et une résilience au stress, tandis que la lignée grasse présente des mécanismes orientés vers le stockage énergétique et des adaptations structurales du muscle.

Ce travail souligne l'importance de combiner la conservation de la diversité génétique et l'identification des signatures de sélection pour sécuriser la durabilité des lignées aquacoles expérimentales et améliorer les programmes de sélection.

Abstract (1600 caractères maximum):

This work investigates genetic diversity and selection signatures in experimental rainbow trout (*Oncorhynchus mykiss*) lines developed by INRAE. The first objective was to assess the synthetic line, initially created as a genetic reservoir. Results indicate a moderate erosion of diversity following two major mortality events, yet overall diversity remains higher than in commercial lines. The use of cryopreserved sperm from past cohorts would only yield a minor gain (0.86%) while depleting nearly all available samples, making it an unsustainable strategy. Instead, establishing a new cryobank from the 2023 breeders is recommended to safeguard against future bottlenecks.

The second objective focused on fat and lean lines subjected to eight generations of divergent selection for muscle lipid content. Genomic analyses (ROH, F_{ST}) revealed candidate regions under selection associated with lipid metabolism, hormonal regulation, metabolic plasticity, and muscle structure. The lean line appears characterized by efficient feed conversion and greater stress resilience, while the fat line shows stronger hormonal and structural mechanisms promoting energy storage and muscle lipid deposition.

Overall, this work highlights the dual importance of maintaining genetic diversity in experimental lines and identifying genomic regions linked to relevant traits. These findings provide insights for the sustainable management of aquaculture breeding programs and the long-term adaptability of farmed fish populations.

Mots-clés : Diversité génétique, Cryoconservation, Sélection divergente, Métabolisme lipidique, Truite arc-en-ciel

Key Words: Genetic diversity, Cryoconservation, Divergent selection, Lipid metabolism, Rainbow trout

* *Elément qui permet d'enregistrer les notices auteurs dans le catalogue des bibliothèques universitaires*