

Article

Genotyping Strategies Using ddRAD Sequencing in Farmed Arctic Charr (*Salvelinus alpinus*)

Fotis Pappas and Christos Palaiokostas *

Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, P.O. Box 7090, 750 07 Uppsala, Sweden; foas0001@stud.slu.se

* Correspondence: christos.palaiokostas@slu.se

Simple Summary: Animal breeding in recent years has benefited greatly from the availability of large-scale genetic information. The most widely applied genomic tools in selective breeding are specialized arrays that use DNA hybridization. However, the high financial investments accompanying this practice impair the profitability of emerging aquaculture species, including Arctic charr. The aim of the current study was to assess and compare the potential of two cost-efficient genotyping strategies applicable in a variety of breeding-related tasks, such as pedigree verification, genetic diversity screening and detection of genomic regions that are associated with phenotypes of economic importance. Both strategies are based on reduced representation sequencing but differ in sequencing coverage (low and high). The low coverage strategy offers a higher density of DNA markers, but also presents a greater proportion of missing data in the marker set and is characterized by more uncertainty in determining heterozygosity compared to high coverage. Our results show that while high coverage genotyping performs better in genetic diversity and kinship analyses, a low coverage strategy is more successful in identifying genomic regions associated with phenotypic traits, leading to the conclusion that both strategies could be of value into selection schemes.



Citation: Pappas, F.; Palaiokostas, C. Genotyping Strategies Using ddRAD Sequencing in Farmed Arctic Charr (*Salvelinus alpinus*). *Animals* **2021**, *11*, 899. <https://doi.org/10.3390/ani11030899>

Academic Editor: Montse Pérez

Received: 12 February 2021

Accepted: 16 March 2021

Published: 21 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Incorporation of genomic technologies into fish breeding programs is a modern reality, promising substantial advances regarding the accuracy of selection, monitoring the genetic diversity and pedigree record verification. Single nucleotide polymorphism (SNP) arrays are the most commonly used genomic tool, but the investments required make them unsustainable for emerging species, such as Arctic charr (*Salvelinus alpinus*), where production volume is low. The requirement to genotype a large number of animals for breeding practices necessitates cost effective genotyping approaches. In the current study, we used double digest restriction site-associated DNA (ddRAD) sequencing of either high or low coverage to genotype Arctic charr from the Swedish national breeding program and performed analytical procedures to assess their utility in a range of tasks. SNPs were identified and used for deciphering the genetic structure of the studied population, estimating genomic relationships and implementing an association study for growth-related traits. Missing information and underestimation of heterozygosity in the low coverage set were limiting factors in genetic diversity and genomic relationship analyses, where high coverage performed notably better. On the other hand, the high coverage dataset proved to be valuable when it comes to identifying loci that are associated with phenotypic traits of interest. In general, both genotyping strategies offer sustainable alternatives to hybridization-based genotyping platforms and show potential for applications in aquaculture selective breeding.

Keywords: reduced representation sequencing; *Salvelinus alpinus*; selective breeding; GBS; salmonid

1. Introduction

Over the last decade, genomic technologies have transformed the field of aquaculture, opening up exciting new avenues for in-depth studies of practically any trait of interest [1]. Implementations of genomic technologies in aquaculture breeding schemes relying on

single nucleotide polymorphism (SNP) arrays have been most beneficial for the major aquaculture species [2]. Nevertheless, SNP arrays require considerable financial investments and running costs that are particularly difficult to sustain in the case of rare aquaculture species. Reduced-representation genotyping platforms offer a cost-effective solution for detecting genome-wide genetic polymorphisms irrespective of the availability of prior genomic information [3]. An abundance of reduced-representation based methodologies have become available over the last years following the introduction of restriction site-associated DNA (RAD) sequencing [4].

The RAD-derived platforms supported by a plethora of available restriction enzymes have proven to be particularly flexible in their applications [5]. Most commonly used type II restriction enzymes recognize between 4 and 8 bases of sequence motifs. Since the sequencing output per lane and the associated cost are in general fixed for each platform (minimal variation between sequencing runs), a balance needs to be obtained between genotyping density and sequencing cost per sample. Naturally, more frequent enzymatic cutters will result in higher genotyping densities. On the other hand, the required sequencing effort, and as such, the associated cost per sample for identifying genomic polymorphisms increase in correspondence with the number of interrogated sites [6]. In comparison, the choice of less frequent enzymatic cutters would allow multiplexing a higher number of animals in a sequencing lane, reducing the genotyping cost per sample at the expense though of reducing the obtained genotyping density.

A number of common enzymatic cutters used for genotyping by sequencing (GBS) [7] have been successfully applied in the form of low read coverage per interrogated site for deciphering genomic relationships at a high resolution [8]. Low coverage genotyping by sequencing has been successfully applied for estimating genome-wide linkage disequilibrium [9], genetic map construction [10], gender prediction [11], parentage assignment [12] and genomic selection purposes [13,14]. Nevertheless, the aforementioned strategy usually results in an increase of missing genotypic data and in a reduction of confidently identifying heterozygotes due to the low number of supporting reads as opposed to high coverage genotyping approaches.

Arctic charr (*Salvelinus alpinus*) is an attractive candidate for diversifying the Nordic aquaculture industry, with an ongoing breeding program having been operational in Sweden for approximately 40 years [15,16]. Nevertheless, limited implementation of modern genomic technologies has taken place to guide selection decisions. Double digest RAD (ddRAD) relying on the simultaneous usage of two enzymes is one of the most popular reduced-representation sequencing platforms. The combination of a relatively easy library construction workflow and cost-efficiency [6] makes ddRAD particularly useful for studying the genetic diversity of populations [17–20], constructing genetic maps [21–24] and quantitative trait loci mapping [25–29].

In the current study, we compared two genotyping strategies based on ddRAD that were applied on 277 (ten full-sibling families) and 188 (eight full-sibling families; subset of the 277 samples) samples, respectively, of Arctic charr originating from the national Swedish breeding program. During the first genotyping scenario, we applied less frequent cutting restriction enzymes (SbfI-SphI) while aiming to identify SNPs supported by high read coverage. In the second scenario, we constructed shallow coverage ddRAD libraries through the usage of more frequent enzymatic cutters (PstI-NlaIII). SNPs were identified under both scenarios and were used for estimating genetic diversity metrics, genomic relationships, and identifying growth-related quantitative trait loci (QTLs).

2. Materials and Methods

2.1. Sample Background and Phenotypes

Arctic charr from the national Swedish breeding program was used for the needs of our study. The animals were located at the facilities of Aquaculture Center North (ACN) in Kälarne, Sweden. Growth measurements (body weight and total length) were recorded from 277 animals (10 months old) of the 2017 class, representing 10 full-sibling families

(with two families having the same sire). In addition, the condition factor (K) was calculated for all the genotyped animals using the formula $K = 10^5 \times \text{weight}/\text{length}^3$. Finally, fin-clips were collected for DNA extraction and ddRAD library preparation.

2.2. DNA Extraction and Quantification

Fin clips of Arctic charr of about 3 mm² were collected and preserved in 100% ethanol. Genomic DNA from individual samples was extracted using a salt precipitation method [30] and eluted in 30 µL of 5 mM Tris. The quality of the obtained DNA was assessed through gel electrophoresis (1% agarose gel). DNA samples were quantified using a Qubit fluorimeter (Thermo Fisher Scientific, Waltham, MA, USA) and diluted with Tris buffer (5 mM) to 15 ng/µL.

2.3. ddRAD Library Preparation and Sequencing

ddRAD libraries were simulated in silico using RADinitio v1.1.1 [31] and were prepared according to a standard protocol [6], with minor modifications that have been previously described [32]. Two different types of ddRAD libraries were prepared using either less frequent (SbfI-SphI; $n = 277$; File S1) or more frequent enzymatic cutters (PstI-NlaIII; $n = 188$; File S2). Briefly, each sample (15 ng/µL DNA) was digested at 37 °C for 60 min in the same reaction with either SbfI (recognizing the CCTGCA|GG motif) and SphI (recognizing the GCATG|C motif) or PstI (recognizing the CTGCA|G motif) and NlaIII (recognizing the CATG|N motif) high fidelity restriction enzymes (New England Biolabs, UK; NEB), by using 6 U of each enzyme per microgram of genomic DNA in 1 × Reaction Buffer 4 (NEB). The reactions (6 µL final volume) were then heat inactivated at 65 °C for 20 min. Individual-specific combinations of P1 and P2 adapters, each with a unique 5 or 7 base pair (bp) barcode, were ligated to the digested DNA at 22 °C for 120 min by adding 1 µL SbfI-PstI compatible P1 adapter (25 nM), 0.7 µL SphI-NlaIII compatible P2 adapter (100 nM), 0.06 µL 100 mmol/L rATP (Promega, Southampton, UK), 0.95 µL 1 × Reaction Buffer 2 (NEB), 0.05 µL T4 ligase (NEB, 2 × 10⁶ U/mL) and reaction volumes made up to 12 µL with nuclease-free water for each sample. Following heat inactivation at 65 °C for 20 min, the ligation reactions were slowly cooled to room temperature (over 1 h) then combined in a single pool (for one sequencing lane) and purified. Size selection (300–600 bp) was performed by agarose gel separation and followed by gel purification and PCR amplification. A total of 100 µL each of the amplified libraries (12–14 PCR cycles) was purified using an equal volume of AMPure beads. The libraries were eluted into 20 µL EB buffer (MinElute Gel Purification Kit, Qiagen, Chadstone, Australia). The libraries were sequenced on two SP flow cells of an Illumina NovaSeq6000 using 300 cycles and v1.5 chemistry kits (150 bp paired end-reads) at the National Genomics Infrastructure in Uppsala, Sweden.

2.4. Sequence Data Analysis and SNP Genotyping

Reads of low quality ($Q < 20$) and missing the expected restriction sites were discarded. The retained reads were aligned to the *Salvelinus* sp. reference genome assembly GCA_002910315.2 using bowtie2 [33]. Stacks v2.5 [34] was used to identify and extract single nucleotide polymorphisms (SNPs) using the *gstacks* module. An alpha threshold of 0.01 and 0.05 for calling SNPs and genotypes respectively was applied using the *marukilow* model of *gstacks*. From each putative ddRAD locus, only a single SNP was used for downstream analysis. SNPs with minor allele frequency (MAF) < 0.05 and maximum heterozygosity > 0.7 across the tested samples were discarded using the *populations* module of Stacks. Only SNPs found in at least 50% and 70% of the samples in the low and high coverage datasets respectively were retained for downstream analysis.

2.5. Genetic Diversity and Kinship

For the direct comparison of the two genotyping strategies, the intersecting population (individuals present in both SbfI-SphI and PstI-NlaIII filtered datasets; $n = 175$) was used.

Expected (H_e) and observed (H_o) heterozygosity metrics were calculated separately for the two datasets using the R package *adegenet* 2.1.3 [35]. Principal component analysis (PCA) was performed for both genotyping scenarios with the aforementioned software to gain information regarding the population structure. Additionally, discriminant analysis of principal components (DAPC) was conducted to further investigate the presence of genetic clusters. The optimal number of clusters (K) was selected for each dataset using the Bayesian Information Criterion (BIC) [36]. Furthermore, a cross-validation scheme was performed to test the ability of the two marker sets to discriminate between individuals from different full-sibling families. More specifically, data from ~75% of the individuals were used as training sets and the rest were utilized as testing sets (one family was excluded since after quality control it consisted of only 2 members). Information derived from DAPC (*predict.dapc*) performed on the training data sets was used to predict the family of origin for the members of the testing sets and assignment accuracies were calculated. Finally, in order to determine the effect of MAF on the reassignment performance of the low coverage set, we applied a series of MAF filters (0.05, 0.1, 0.15, 0.2 and 0.25) and subjected the yielded sets to the same cross-validation method.

The intersecting population was also used to compute genomic relationship matrices (GRMs) and compare the two datasets in regard to kinship estimations among full siblings. For this purpose, we employed R/*rrblup* 4.6.1 [37,38] and calculated the GRMs for each dataset setting a threshold of 0.3 for maximum missing data and using the Expectation-Maximization (EM) algorithm [39]. Genomic relatedness for full-sibling and non-sibling pairs was then expressed using the following formula:

$$GR_{ij} = \frac{g_{ij}}{\sqrt{g_{ii} \cdot g_{jj}}} \quad (1)$$

where GR_{ij} is the genomic relatedness between individuals i and j , and g refers to the genomic relationships as calculated in the GRMs mentioned above.

2.6. Association with Phenotypic Traits

Association studies for the recorded growth traits were performed separately for the two genotyping strategies and their respective populations. Analyses were performed using the R package *gaston* 1.5.5 [40]. The tests were carried out aiming to identify genomic regions associated with body length and \log_2 transformed K factor. A linear mixed model was applied:

$$Y = X\alpha + Z\beta + S\omega + \varepsilon \quad (2)$$

where α is the vector of fixed effects other than marker effects (intercept, sex, family-tank), β is the fixed effect of each SNP marker, ω is the vector of animal random effect $\sim N(0, G\sigma_\alpha^2)$ and ε is the vector of residuals $\sim N(0, I\sigma_\varepsilon^2)$. X , Z and S are incidence matrices relating Y with α and β respectively. G is a genomic relationship matrix computed with the *GRM* function, I is an identity matrix, σ_ε^2 is the residual variance and σ_α^2 represents the additive genetic variance. Multi-test adjustment of the genome-wide significance thresholds was achieved using Bonferroni corrections (0.05α level/ n), where n is the number of tested SNPs. Furthermore, Benjamini–Hochberg (BH) adjustment of p -values was performed to control for false discovery rate.

3. Results

3.1. Genotypic Information

Due to high rate of missing data, 24 animals from the high coverage dataset (HC) and nine animals from the low coverage dataset (LC) were removed. Therefore, the HC dataset was comprised of 253 animals, and 179 were finally represented in the LC dataset. The number of individuals successfully genotyped with both *SbfI-SphI* and *PstI-NlaII* was 175 (intersecting population) and was used for analyses where direct comparisons were necessary. The expected number of RAD loci as estimated with the in-silico library

stimulations was 1524 for SbfI-SphI (HC) and 305,077 for PstI-NlaIII (LC). In contrast, the filtered variant sets in our study consisted of 1034 SNP genotypes for SbfI-SphI, and 38,224 SNPs for PstI-NlaIII. The mean coverage for the former was 428X, while for the latter the mean coverage was 3X. The distributions of MAF and SNP call rates after filtering are shown in Figure 1. The marker set derived from the shallow coverage approach was characterized by a greater proportion of low call-rate genotypes and MAF compared to high coverage.

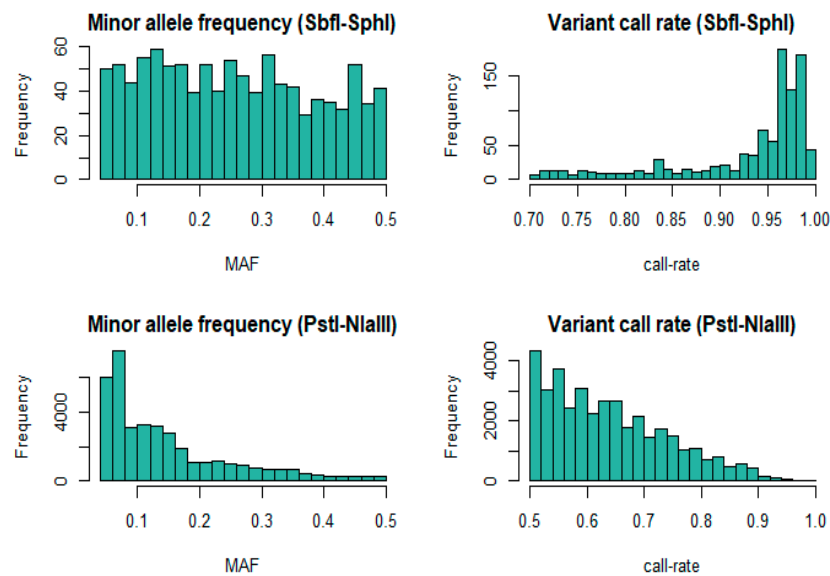


Figure 1. Distributions of post-filtering minor allele frequency (MAF) and single nucleotide polymorphism (SNP) call rate for SbfI-SphI ($n = 253$) and the intersecting ($n = 175$) animals that were also genotyped with PstI-NlaIII.

3.2. Descriptive Statistics of Phenotypic Traits

The distributions of phenotypic data are graphically presented in Figure 2. The overall mean body length of the studied sample was 169.03 mm (sd = 19.01), and the mean wet body weight was 55.88 gr (sd = 19.57). For condition factor K, the mean value was 1.11 (sd = 0.11) and a \log_2 transformation was performed to obtain a normalized distribution (Figure 2).

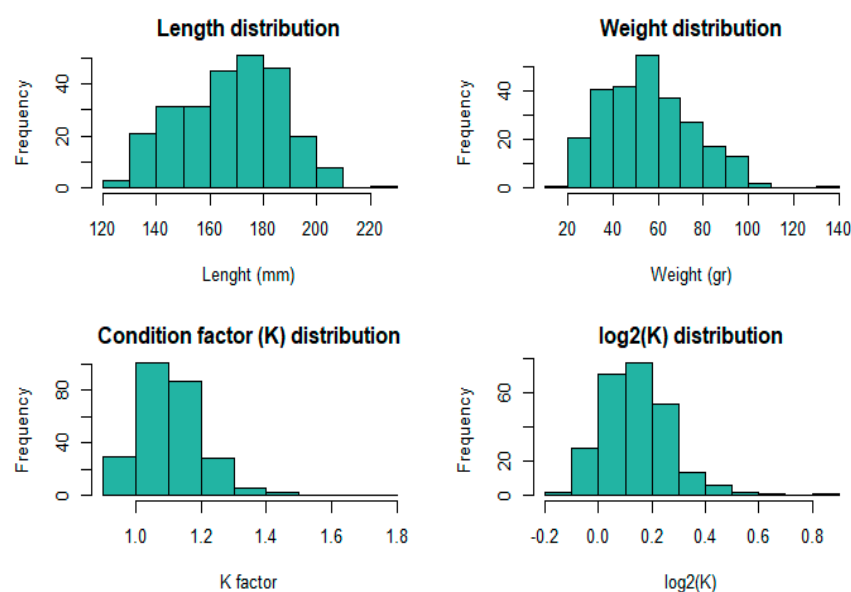


Figure 2. Histograms of distributions for length, weight, K factor and \log_2 transformed K factor ($n = 257$).

3.3. Genetic Diversity

For the high coverage dataset, the expected (H_e) and observed (H_o) heterozygosity were 0.34 and 0.31 respectively; and the same metrics for the shallow coverage scenario were 0.23 and 0.22. Differences in subpopulations detection were visualized when dimensionality reduction was used through PCA (Figure 3). The first and second principal components for the SbfI-SphI SNP set accounted for 10.39% and 8.34% of the observed variation, while for the PstI-NlaIII dataset, the proportions for PC1 and PC2 were 9.11% and 4.62% respectively.

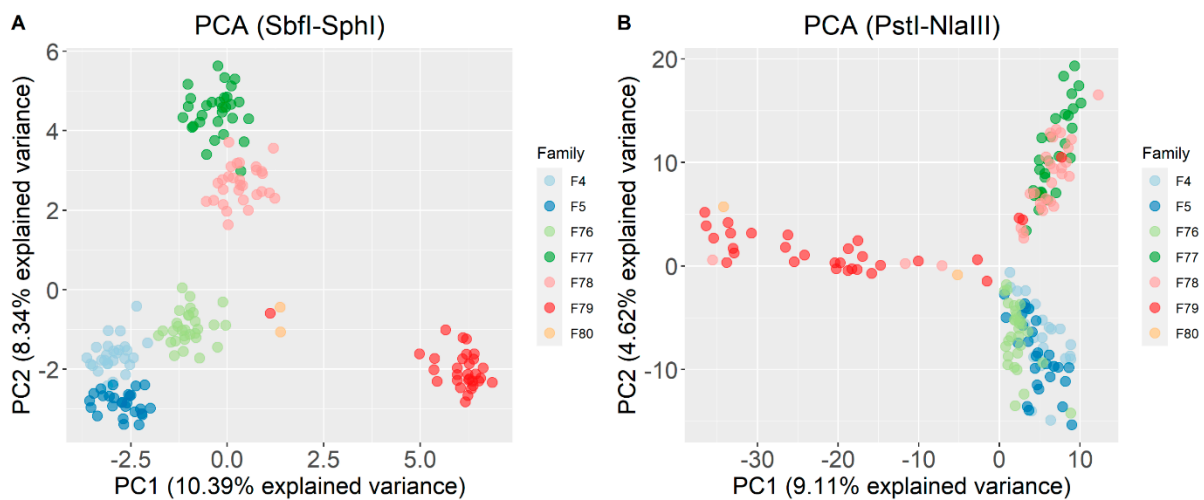


Figure 3. Principal component analysis for (A) SbfI-SphI and (B) PstI-NlaIII genotyping scenarios. The represented population is the intersection ($n = 175$) of the individuals that were genotyped in both scenarios.

DAPC provided further insights of the formed genetic clusters (Figure 4). The optimal number of clusters (K) was suggested to be seven for SbfI-SphI and five for PstI-NlaIII. In general, clustering according to the shallow coverage dataset was less efficient and only provided a rough discrimination between the different families. In the analyses of the high coverage SNP sets, two individuals (belonging to families F77 and F79) failed to group with their original families and were assigned to closely related groups (Figure 4A). Finally, the genetic clusters representing families F4 and F5, which had the same sire, always appeared to group closely (SbfI-SphI) or were indistinguishable (PstI-NlaIII).

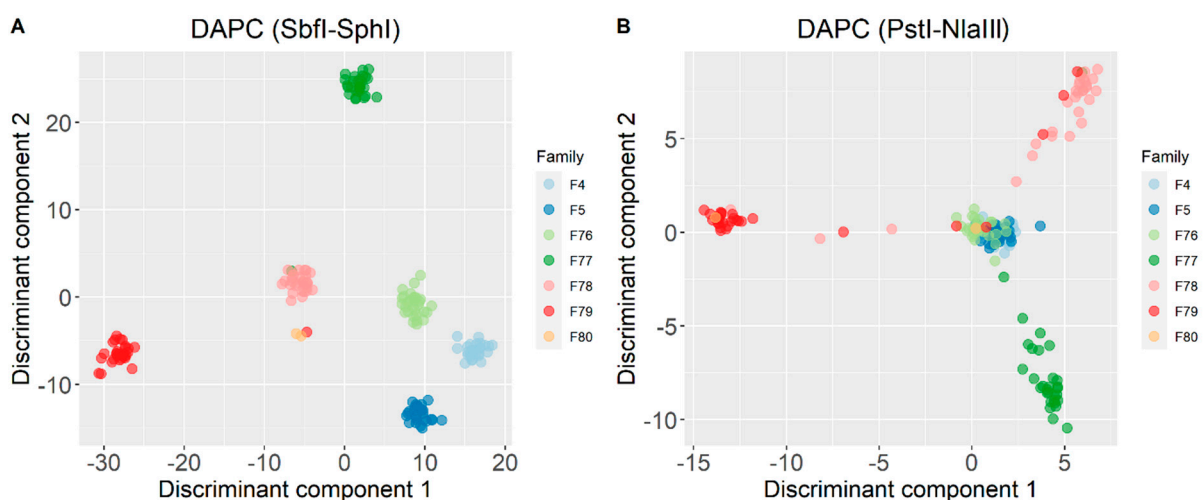


Figure 4. Discriminant analysis of principal components (DAPC) for (A) SbfI-SphI and (B) PstI-NlaIII genotyping scenarios. The represented population is the intersection ($n = 175$) of the individuals that were genotyped in both scenarios.

A cross-validation scheme was employed to assess the accuracy of family reassignment (DAPC) using the two datasets. The relative frequency (observed minimum) of assignment to the original (according to pedigree records) family was 95.83% for SbfI-SphI and 68.75% for PstI-NlaIII. Frequencies of family assignments as calculated during cross-validation are visualized in Figure 5 for both SNP sets. Furthermore, the analysis performed on PstI-NlaIII dataset demonstrated that reassignment accuracy remained at the same levels for MAF filtering thresholds spanning from 0.05 to 0.25 (Figure S1).

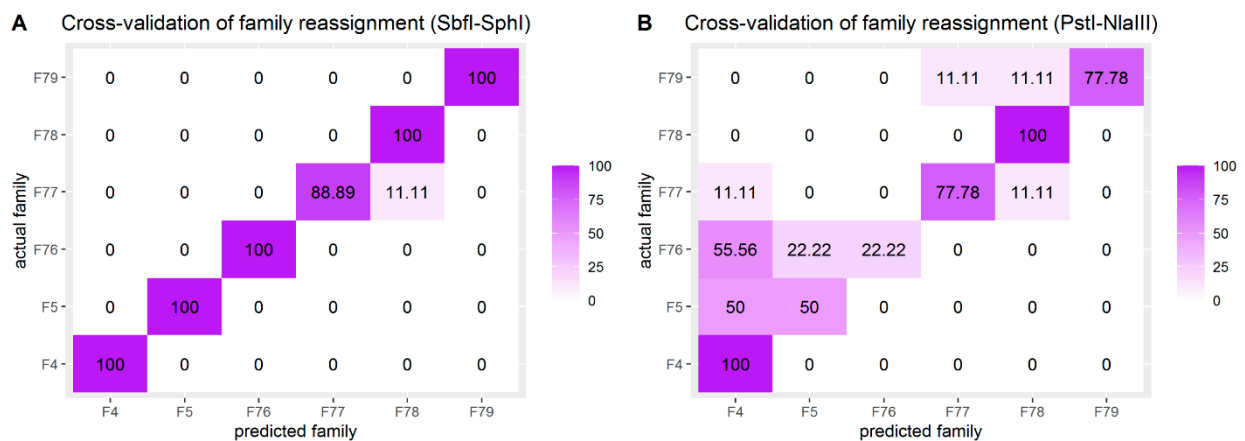


Figure 5. Heatmaps visualizing the relative frequencies (%) of family predictions according to the DAPC cross-validation scheme for (A) SbfI-SphI and (B) PstI-NlaIII.

Mean genomic relatedness (see Genetic Diversity and Kinship in Materials and Methods) among full siblings was found to be 0.23 (sd = 0.09) for SbfI-SphI and 0.19 (sd = 0.16) for PstI-NlaIII, and the same metric for non-sibling pairs was -0.05 (sd = 0.06) for SbfI-SphI and -0.05 (sd = 0.11) for PstI-NlaIII. The density plots in Figure 6 visualize the overlapping of GRs distributions between full-sibling and non-sibling pairs, which was found to be 3.70% for SbfI-SphI and 16.08% for PstI-NlaIII.

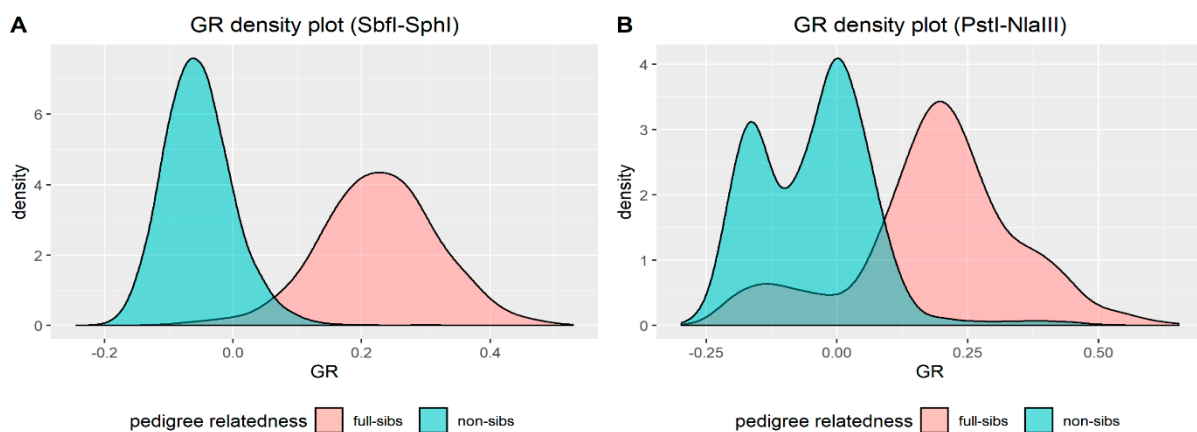


Figure 6. Density plots of genomic relatedness (GR) among full siblings and non-siblings for: (A) SbfI-SphI and (B) PstI-NlaIII.

3.4. Association Scans for Phenotypic Traits

No significant QTL peaks were observed for body length or \log_2 transformed condition factor (K) when using genotypic information from the high coverage dataset (Figure 7).

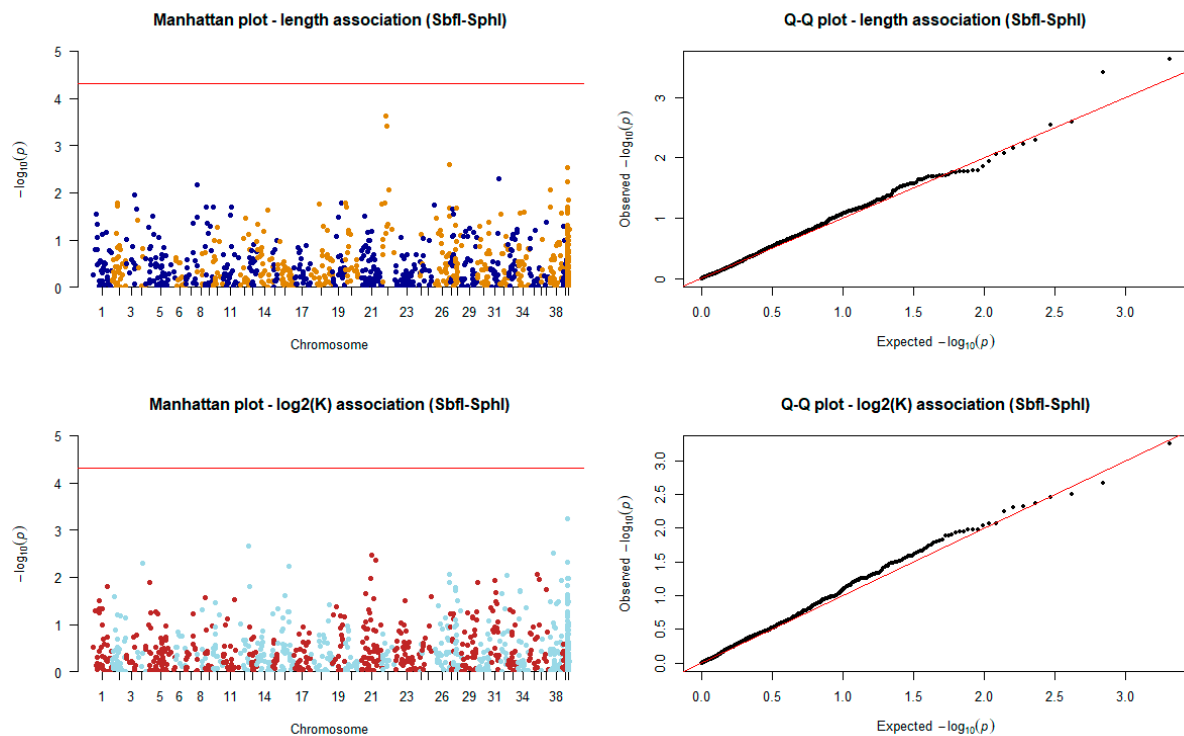


Figure 7. Manhattan and quantile–quantile plots of the association tests for length and \log_2K in the SbfI-SphI scenario ($n = 253$). The red horizontal line indicates the Bonferroni error rate-adjusted significance level.

Association of markers derived from the shallow coverage strategy with body length failed to identify QTLs. Nevertheless, SNPs from the same dataset were found to be significant in the case of \log_2K . As shown in Figure 8, a total of eight SNPs were identified. More specifically, three SNP markers passed the Bonferroni adjusted significance threshold, and another five SNPs were found to be significantly associated with the \log_2 transformed condition factor after adjusting the obtained p -values with the BH method.

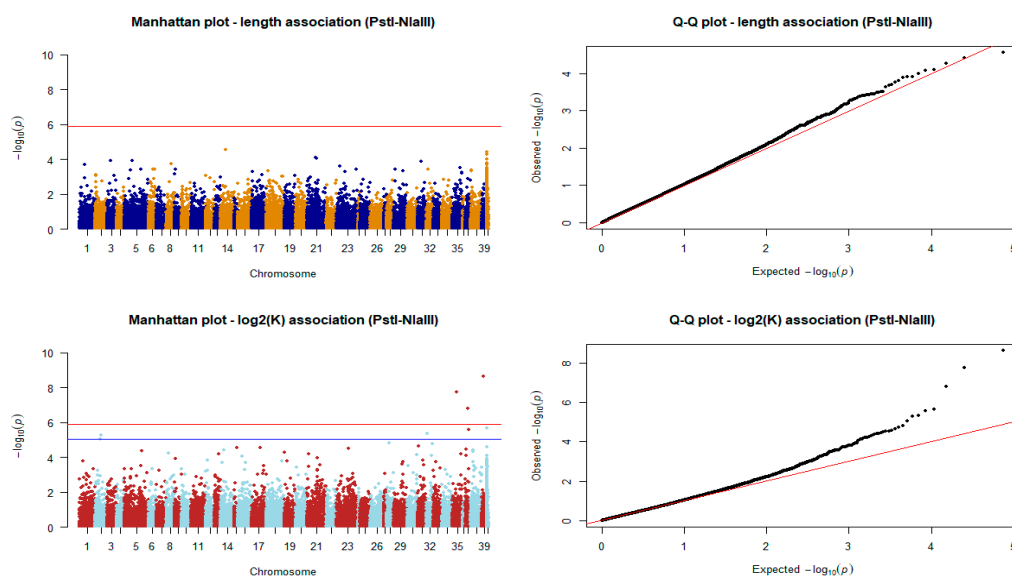


Figure 8. Manhattan and quantile–quantile plots of the association tests for length and \log_2K in the PstI-NlaIII scenario ($n = 179$). The red horizontal line indicates the Bonferroni error rate-adjusted significance level. The blue line indicates the threshold of the significant markers after BH adjustment of p -values.

Finally, we highlight 11 genes based on physical distance (less than 20 kb) from the chromosomal position of each variant (Table 1) and found that three of these SNPs were located in genes, while two others were located within a distance of less than 1000 bp from genes.

Table 1. SNPs significantly associated with the logarithmically transformed condition factor and nearby candidate genes. * SNP located within 1 kb of gene, ** SNP located in gene.

Linkage Group	SNP Position (bp)	Unadjusted <i>p</i> -Value	BH Adjusted <i>p</i> -Value	Annotated or Predicted Genes within 20 kb on Either Side
NC_036876.1	6,904,331	2.317911×10^{-9}	8.859984×10^{-5}	LOC111960292 **
NC_036872.1	16,812,304	1.710431×10^{-8}	3.268976×10^{-4}	nkx2.7, LOC111957894, LOC111957491
NC_036874.1	10,834,663	1.568166×10^{-7}	1.998052×10^{-3}	apom *, LOC111959084, LOC111958785
NW_019945418.1	7585	2.161829×10^{-6}	2.030243×10^{-2}	-
NC_036874.1	13,345,514	2.655718×10^{-6}	2.030243×10^{-2}	-
NC_036869.1	2,707,853	4.399173×10^{-6}	2.761334×10^{-2}	LOC111955246 *, pex1
NC_036839.1	21,199,964	5.056859×10^{-6}	2.761334×10^{-2}	klf7 **
NC_036839.1	17,537,842	8.506600×10^{-6}	4.064454×10^{-2}	LOC111972823 **

4. Discussion

In the current study, both high and low coverage ddRAD genotyping scenarios provided genotypic information that is useful for different applications in a wide range of breeding related tasks. Reduced representation sequencing can be a valuable tool for selective breeding practices in emerging aquaculture species, such as Arctic charr [14]. Genomic information is useful for important aspects of fish breeding that include, but are not limited to estimating relationships between individuals, assessing genetic diversity in populations and even investigating the genetic basis of phenotypic variation [41]. However, the cost of genotyping is a limiting factor for the sustainability of breeding programs. Available, reduced representation methods include strategies that yield for the same cost either deep coverage but low density of genetic markers, or high genotyping density with shallow coverage—and thus, a greater proportion of missing data. Encouragingly, as suggested by both simulated data and empirical studies, low coverage genotyping by sequencing appears to be an attractive option for selective breeding practices [12,42].

As expected by RADinitio library simulations, the PstI-NlaIII library resulted in significantly higher number of SNP biallelic genotypes than SbfI-SphI. The actual counts of filtered variants were lower than estimated in silico with the deviation being larger and more profound for the low coverage scenario. This could be attributed to a number of factors that among others include the low depth in the case of PstI-NlaIII (~ 3X) and the fact that the reference genome was derived from a distant population and might not be adequately representative of the nucleus of the national Swedish breeding program.

4.1. Genetic Diversity

The deep coverage strategy outperformed shallow coverage genotyping. A comparison of heterozygosity metrics indicates that the latter underestimates the proportion of heterozygotes in the population, possibly due to missing information in the marker set [10]. Observed heterozygosity (H_o) was found to be lower than expected heterozygosity (H_e) in both cases, reflecting the fact that the studied population was derived from a closed-nucleus breeding program [43].

Furthermore, genotyping in the SbfI-SphI scenario resulted in better and more robust detection of subpopulations and clustering at the family level in the PCA and DAPC analytical procedures. This finding was further supported by a cross-validation scheme where

the estimated accuracy of family reassignment was higher for deep coverage, suggesting that high quality genotypic information is a priority over marker density for genetic diversity and population structure analysis. Despite providing a much more defined view on genetic clusters, DAPC for the SbfI-SphI scenario resulted in two individuals not grouping with their original families, but this could be explained by putative errors in the pedigree records.

It should be highlighted that in our study we tried to decipher the population structures and genetic clustering of full-sibling families in a closed breeding nucleus, a task that is more challenging and demanding than studying genetic diversity and differentiation among distant populations [44]. Even though differing MAF thresholds could potentially affect population structure inference [45,46], setting stricter MAF filtering thresholds for the PstI-NlaIII dataset in our study did not affect the clustering performance. However, the total number of SNP markers decreased, which could consequently result in decreased computational intensity.

4.2. Kinship Investigation

Kinship estimations among full-sibling pairs were higher for the high coverage dataset and provided better discrimination from the GR values of non-sibling pairs, despite incorporating remarkably less genotypic information than the low coverage dataset. As with genetic diversity analyses, this might have been due to the higher SNP calling uncertainty entailed by shallow coverage genotyping-by-sequencing approaches. Overall, both genotypic datasets underestimated the average relationship between full-sibling pairs, thereby performing worse compared to similar studies in livestock that used SNP arrays [47]. The above was probably to be expected in the case of HC dataset, since the number of SNPs (~1000) is considered low for providing the resolution needed for a genomic relationship matrix. Notably, genomic relationship matrices are most commonly constructed from SNPs in the range of tens of thousands [48]. On the other hand, a high-density SNP dataset even at low coverage is expected to result in a more accurate genomic relationship matrix and better discern relationships between full siblings. More specifically, low coverage genotyping by sequencing has been shown to be effective in estimating genomic relationships amongst full siblings both in Atlantic salmon [8] and in Arctic charr [14]. However, compared to the above studies the library preparation protocol of our study differed. In particular, opposed to the previous studies, the size selection during library preparation was performed manually which might have rendered the detection of SNPs in the low coverage dataset more challenging due to the larger number of sampled sites. As a result, the LC dataset in our study had a mean coverage of 3X compared to 5X of the aforementioned studies. Nevertheless, both datasets in our study were able to successfully separate full siblings from more distantly related animals as shown from their respective distributions of genomic relationships.

4.3. Association Analysis

Our association study for length and \log_2 transformed condition factor (K) confirmed that the density of the marker set is crucial when conducting genome-wide regression studies in aquatic animal populations [49–51]. All SNPs of the high coverage set failed to reach the genome-wide adjusted significance threshold for both studied traits. SNPs from the low coverage dataset failed to significantly associate with body length but eight of them were identified as significantly associated with \log_2 K. Further analysis showed that some of those loci are located in genomic regions containing genes that are affiliated with development and metabolism. Among them, a SNP was located 570 bp upstream of the *apolipoprotein M* (*apom*) gene, which has been previously associated with obesity in mice [52,53] and humans [54,55]. Moreover, two of the significant SNPs were located in or near genes that code for transcription factors such as *klf7* and *nkx2.7*. Genetic studies on QTLs of condition factor in Arctic charr have been previously conducted [56–58], but to our knowledge the current article is the first to present genetic associations using genomic

approaches. The authors acknowledge though, that the association study was performed on small sample sizes, and further validation regarding the putative role of the above SNPs would be required. Nevertheless, it should be noted that the studied sample consisted of full-sibling families from a closed breeding nucleus, meaning that the probability of detecting QTLs is higher compared to scenarios where distant related individuals are used.

5. Conclusions

Both deep and shallow coverage genotyping by sequencing strategies provides valuable genomic information that can be incorporated in fish breeding programs. While deep coverage results in high quality but less dense genotypes, shallow coverage provides higher genotypic density with a greater proportion of missing data and higher uncertainty. The first strategy performed better in genetic diversity and population structure analyses, while low coverage genotyping proved to be more informative regarding genome-wide regression for association, with phenotypic traits demonstrating the potential of both approaches in aquaculture breeding schemes.

Supplementary Materials: The following are available online at <https://www.mdpi.com/2076-2615/11/3/899/s1>: File S1 containing the barcode combinations that were used for the SbfI-SphI dataset. File S2 containing the barcode combinations that were used for the PstI-NlaIII dataset. Figure S1 depicting the family predictions using DAPC for varying levels of MAF.

Author Contributions: Conceptualization, C.P.; methodology, C.P.; software, F.P. and C.P.; validation, F.P. and C.P.; formal analysis, F.P.; investigation, F.P. and C.P.; resources, C.P.; data curation, F.P.; writing—original draft preparation, F.P.; writing—review and editing, F.P. and C.P.; visualization, F.P.; supervision, C.P.; project administration, C.P.; funding acquisition, C.P. All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge support from FORMAS under grant agreement 2018-00869 (NextGenCharr) and the Kolarctic funding body under the ARCTAQUA project (grant agreement 4/2018/095/KO4058). Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The facility is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation.

Institutional Review Board Statement: The study was carried out in accordance with Swedish legislation for conducting animal research as described in the Animal Welfare Act 2018:1192 (ethics permit: 5.2.18-09859/2019).

Data Availability Statement: The sequence reads in the form of fastq files have been deposited to the National Centre for Biotechnology Information (NCBI) and are publicly available under project ID PRJNA705295.

Acknowledgments: We would like to thank the VBCN personnel for assisting with the phenotypic recordings. The authors would like to thank two anonymous reviewers for their constructive comments and suggestions for improving the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Houston, R.D.; Bean, T.P.; Macqueen, D.J.; Gundappa, M.K.; Jin, Y.H.; Jenkins, T.L.; Selly, S.L.C.; Martin, S.A.M.; Stevens, J.R.; Santos, E.M.; et al. Harnessing Genomics to Fast-Track Genetic Improvement in Aquaculture. *Nat. Rev. Genet.* **2020**, *21*, 389–409. [[CrossRef](#)] [[PubMed](#)]
2. You, X.; Shan, X.; Shi, Q. Research Advances in the Genomics and Applications for Molecular Breeding of Aquaculture Animals. *Aquaculture* **2020**, *526*, 735357. [[CrossRef](#)]
3. Robledo, D.; Palaiokostas, C.; Bargelloni, L.; Martínez, P.; Houston, R. Applications of Genotyping by Sequencing in Aquaculture Breeding and Genetics. *Rev. Aquac.* **2017**, *10*, 1–13. [[CrossRef](#)] [[PubMed](#)]
4. Baird, N.A.; Etter, P.D.; Atwood, T.S.; Currey, M.C.; Shiver, A.L.; Lewis, Z.A.; Selker, E.U.; Cresko, W.A.; Johnson, E. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE* **2008**, *3*, e3376. [[CrossRef](#)]
5. Andrews, K.R.; Good, J.M.; Miller, M.R.; Luikart, G.; Hohenlohe, P.A. Harnessing the Power of RADseq for Ecological and Evolutionary Genomics. *Nat. Rev. Genet.* **2016**, *17*, 81–92. [[CrossRef](#)] [[PubMed](#)]

6. Peterson, B.K.; Weber, J.N.; Kay, E.H.; Fisher, H.S.; Hoekstra, H.E. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE* **2012**, *7*, e37135. [[CrossRef](#)] [[PubMed](#)]
7. Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S.; Mitchell, S.E. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* **2011**, *6*, e19379. [[CrossRef](#)] [[PubMed](#)]
8. Dodds, K.G.; McEwan, J.C.; Brauning, R.; Anderson, R.M.; van Stijn, T.C.; Kristjánsson, T.; Clarke, S.M. Construction of Relatedness Matrices Using Genotyping-by-Sequencing Data. *BMC Genom.* **2015**, *16*, 1047. [[CrossRef](#)] [[PubMed](#)]
9. Bilton, T.P.; McEwan, J.C.; Clarke, S.M.; Brauning, R.; van Stijn, T.C.; Rowe, S.J.; Dodds, K.G. Linkage Disequilibrium Estimation in Low Coverage High-Throughput Sequencing Data. *Genetics* **2018**, *209*, 389–400. [[CrossRef](#)] [[PubMed](#)]
10. Bilton, T.P.; Schofield, M.R.; Black, M.A.; Chagné, D.; Wilcox, P.L.; Dodds, K.G. Accounting for Errors in Low Coverage High-Throughput Sequencing Data When Constructing Genetic Maps Using Biparental Outcrossed Populations. *Genetics* **2018**, *209*, 65–76. [[CrossRef](#)] [[PubMed](#)]
11. Bilton, T.P.; Chappell, A.J.; Clarke, S.M.; Brauning, R.; Dodds, K.G.; McEwan, J.C.; Rowe, S.J. Using Genotyping-by-sequencing to Predict Gender in Animals. *Anim. Genet.* **2019**, *50*, 307–310. [[CrossRef](#)]
12. Dodds, K.G.; McEwan, J.C.; Brauning, R.; van Stijn, T.C.; Rowe, S.J.; McEwan, K.M.; Clarke, S.M. Exclusion and Genomic Relatedness Methods for Assignment of Parentage Using Genotyping-by-Sequencing Data. *G3 Genes Genomes Genet.* **2019**, *9*, 3239–3247. [[CrossRef](#)]
13. Faville, M.J.; Ganesh, S.; Cao, M.; Jahufer, M.Z.Z.; Bilton, T.P.; Easton, H.S.; Ryan, D.L.; Trethewey, J.A.K.; Rolston, M.P.; Griffiths, A.G.; et al. Predictive Ability of Genomic Selection Models in a Multi-Population Perennial Ryegrass Training Set Using Genotyping-by-Sequencing. *Theor. Appl. Genet.* **2018**, *131*, 703–720. [[CrossRef](#)]
14. Palaiokostas, C.; Clarke, S.M.; Jeuthe, H.; Brauning, R.; Bilton, T.P.; Dodds, K.G.; McEwan, J.C.; De Koning, D.-J. Application of Low Coverage Genotyping by Sequencing in Selectively Bred Arctic Charr (*Salvelinus alpinus*). *G3 Genes | Genomes | Genetics* **2020**. [[CrossRef](#)]
15. Carlberg, H.; Nilsson, J.; Brännäs, E.; Alanära, A. An Evaluation of 30years of Selective Breeding in the Arctic Charr (*Salvelinus alpinus* L.) and Its Implications for Feeding Management. *Aquaculture* **2018**, *495*, 428–434. [[CrossRef](#)]
16. Eriksson, L.O.; Alanära, A.; Nilsson, J.; Brännäs, E. The Arctic Charr Story: Development of Subarctic Freshwater Fish Farming in Sweden. *Hydrobiologia* **2010**, *650*, 265–274. [[CrossRef](#)]
17. Kajungiro, R.A.; Palaiokostas, C.; Pinto, F.A.L.; Mmochi, A.J.; Mtolera, M.; Houston, R.D.; de Koning, D.J. Population Structure and Genetic Diversity of Nile Tilapia (*Oreochromis niloticus*) Strains Cultured in Tanzania. *Front. Genet.* **2019**, *10*. [[CrossRef](#)]
18. Moses, M.; Mtolera, M.S.P.; Chauka, L.J.; Lopes, F.A.; de Koning, D.J.; Houston, R.D.; Palaiokostas, C. Characterizing the Genetic Structure of Introduced Nile Tilapia (*Oreochromis niloticus*) Strains in Tanzania Using Double Digest RAD Sequencing. *Aquac. Int.* **2019**. [[CrossRef](#)]
19. Torati, L.S.; Taggart, J.B.; Varela, E.S.; Araripe, J.; Wehner, S.; Migaud, H. Genetic Diversity and Structure in Arapaima Gigas Populations from Amazon and Araguaia-Tocantins River Basins. *BMC Genet.* **2019**, *20*. [[CrossRef](#)] [[PubMed](#)]
20. Hosoya, S.; Kikuchi, K.; Nagashima, H.; Onodera, J.; Sugimoto, K.; Satoh, K.; Matsuzaki, K.; Yasugi, M.; Nagano, A.J.; Kumagaya, A.; et al. Assessment of Genetic Diversity in Coho Salmon (*Oncorhynchus kisutch*) Populations with No Family Records Using DdRAD-Seq. *BMC Research Notes* **2018**, *11*, 1–5. [[CrossRef](#)] [[PubMed](#)]
21. Kai, W.; Nomura, K.; Fujiwara, A.; Nakamura, Y.; Yasuike, M.; Ojima, N.; Masaoka, T.; Ozaki, A.; Kazeto, Y.; Gen, K.; et al. A DdRAD-Based Genetic Map and Its Integration with the Genome Assembly of Japanese Eel (*Anguilla japonica*) Provides Insights into Genome Evolution after the Teleost-Specific Genome Duplication. *BMC Genom.* **2014**, *15*, 233. [[CrossRef](#)]
22. Oral, M.; Colléter, J.; Bekaert, M.; Taggart, J.B.; Palaiokostas, C.; McAndrew, B.J.; Vandeputte, M.; Chatain, B.; Kuhl, H.; Reinhardt, R.; et al. Gene-Centromere Mapping in Meiotic Gynogenetic European Seabass. *BMC Genom.* **2017**, *18*, 449. [[CrossRef](#)] [[PubMed](#)]
23. Manousaki, T.; Tsakogiannis, A.; Taggart, J.B.; Palaiokostas, C.; Tsaparis, D.; Lagnel, J.; Chatziplis, D.; Magoulas, A.; Papan-droulakis, N.; Mylonas, C.C.; et al. Exploring a Nonmodel Teleost Genome Through RAD Sequencing—Linkage Mapping in Common Pandora, *Pagellus erythrinus* and Comparative Genomic Analysis. *G3 Genes | Genomes | Genetics* **2016**, *6*, 509–519. [[CrossRef](#)] [[PubMed](#)]
24. Shi, Y.; Zhou, Z.; Liu, B.; Kong, S.; Chen, B.; Bai, H.; Li, L.; Pu, F.; Xu, P. Construction of a High-Density Genetic Linkage Map and QTL Mapping for Growth-Related Traits in Takifugu Bimaculatus. *Mar. Biotechnol.* **2020**, *22*, 130–144. [[CrossRef](#)]
25. Barria, A.; Christensen, K.A.; Yoshida, G.M.; Correa, K.; Jedlicki, A.; Lhorente, J.P.; Davidson, W.S.; Yáñez, J.M. Genomic Predictions and Genome-Wide Association Study of Resistance Against *Piscirickettsia salmonis* in Coho Salmon (*Oncorhynchus kisutch*) Using ddRAD Sequencing. *G3 Genes Genomes Genet.* **2018**, *8*, 1183–1194. [[CrossRef](#)] [[PubMed](#)]
26. Brown, J.K.; Taggart, J.B.; Bekaert, M.; Wehner, S.; Palaiokostas, C.; Setiawan, A.N.; Symonds, J.E.; Penman, D.J. Mapping the Sex Determination Locus in the Hāpuku (*Polyprion oxygeneios*) Using ddRAD Sequencing. *BMC Genom.* **2016**, *17*, 448. [[CrossRef](#)]
27. Jiang, D.L.; Gu, X.H.; Li, B.J.; Zhu, Z.X.; Qin, H.; ning Meng, Z.; Lin, H.R.; Xia, J.H. Identifying a Long QTL Cluster Across ChrLG18 Associated with Salt Tolerance in Tilapia Using GWAS and QTL-Seq. *Mar. Biotechnol.* **2019**, *21*, 250–261. [[CrossRef](#)] [[PubMed](#)]
28. Kyriakis, D.; Kanterakis, A.; Manousaki, T.; Tsakogiannis, A.; Tsagris, M.; Tsamardinos, I.; Papaharisis, L.; Chatziplis, D.; Potamias, G.; Tsigenopoulos, C.S. Scanning of Genetic Variants and Genetic Mapping of Phenotypic Traits in Gilthead Sea Bream Through ddRAD Sequencing. *Front. Genet.* **2019**, *10*. [[CrossRef](#)]

29. Taslima, K.; Wehner, S.; Taggart, J.B.; de Verdal, H.; Benzie, J.A.H.; Bekaert, M.; McAndrew, B.J.; Penman, D.J. Sex Determination in the GIFT Strain of Tilapia Is Controlled by a Locus in Linkage Group 23. *BMC Genet.* **2020**, *21*, 49. [CrossRef]
30. Taslima, K.; Davie, A.; McAndrew, B.J.; Penman, D.J. DNA Sampling from Mucus in the Nile Tilapia, *Oreochromis niloticus*: Minimally Invasive Sampling for Aquaculture-Related Genetics Research. *Aquac. Res.* **2016**, *47*, 4032–4037. [CrossRef]
31. Rivera-Colón, A.G.; Rochette, N.C.; Catchen, J.M. Simulation with RADinitio Improves RADseq Experimental Design and Sheds Light on Sources of Missing Data. *Mol. Ecol. Resour.* **2021**, *21*, 363–378. [CrossRef]
32. Palaiokostas, C.; Bekaert, M.; Khan, M.G.; Taggart, J.B.; Gharbi, K.; McAndrew, B.J.; Penman, D.J. A Novel Sex-Determining QTL in Nile Tilapia (*Oreochromis niloticus*). *BMC Genom.* **2015**, *16*, 171. [CrossRef] [PubMed]
33. Langmead, B.; Salzberg, S.L. Fast Gapped-Read Alignment with Bowtie 2. *Nat Methods* **2012**, *9*, 357–359. [CrossRef]
34. Rochette, N.C.; Rivera-Colón, A.G.; Catchen, J.M. Stacks 2: Analytical Methods for Paired-End Sequencing Improve RADseq-Based Population Genomics. *Mol. Ecol.* **2019**, *28*, 4737–4754. [CrossRef]
35. Jombart, T. Adegnet: A R Package for the Multivariate Analysis of Genetic Markers. *Bioinformatics* **2008**, *24*, 1403–1405. [CrossRef] [PubMed]
36. Jombart, T.; Devillard, S.; Balloux, F. Discriminant Analysis of Principal Components: A New Method for the Analysis of Genetically Structured Populations. *BMC Genet.* **2010**, *11*, 94. [CrossRef] [PubMed]
37. Endelman, J.B. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *Plant Genome* **2011**, *4*. [CrossRef]
38. Endelman, J.B.; Jannink, J.-L. Shrinkage Estimation of the Realized Relationship Matrix. *G3 Genes | Genomes | Genetics* **2012**, *2*, 1405–1413. [CrossRef] [PubMed]
39. Poland, J.; Endelman, J.; Dawson, J.; Rutkoski, J.; Wu, S.; Manes, Y.; Dreisigacker, S.; Crossa, J.; Sánchez-Villeda, H.; Sorrells, M.; et al. Genomic Selection in Wheat Breeding Using Genotyping-by-Sequencing. *Plant Genome* **2012**, *5*. [CrossRef]
40. Perdry, H.; Dandine-Roulland, C.; Bandyopadhyay, D.; Kettner, L. Pack Package ‘Gaston’: Genetic Data Handling (QC, GRM, LD, PCA) and Linear Mixed Models. Available online: <https://cran.r-project.org/web/packages/gaston/gaston.pdf>.
41. Peñaloza, C.; Manousaki, T.; Franch, R.; Tsakogiannis, A.; Sonesson, A.; Aslam, M.L.; Allal, F.; Bargelloni, L.; Houston, R.D.; Tsigenopoulos, C.S. Development and Validation of a Combined Species SNP Array for the European Seabass (*Dicentrarchus labrax*) and Gilthead Seabream (*Sparus aurata*). *bioRxiv* **2020**, 2020.12.17.423305. [CrossRef]
42. Gorjanc, G.; Cleveland, M.A.; Houston, R.D.; Hickey, J.M. Potential of Genotyping-by-Sequencing for Genomic Selection in Livestock Populations. *Genet. Sel. Evol.* **2015**, *47*, 12. [CrossRef]
43. Nilsson, J.; Brännäs, E.; Eriksson, L.-O. The Swedish Arctic Charr Breeding Programme. *Hydrobiologia* **2010**, *650*, 275–282. [CrossRef]
44. Nyinondi, C.S.; Mtolera, M.S.P.; Mmochi, A.J.; Pinto, F.A.L.; Houston, R.D.; de Koning, D.J.; Palaiokostas, C. Assessing the Genetic Diversity of Farmed and Wild Rufiji Tilapia (*Oreochromis urolepis urolepis*) Populations Using DdRAD Sequencing. *Ecol. Evol.* **2020**, *10*, 10044–10056. [CrossRef]
45. Linck, E.; Battey, C.J. Minor Allele Frequency Thresholds Strongly Affect Population Structure Inference with Genomic Data Sets. *Mol. Ecol. Resour.* **2019**, *19*, 639–647. [CrossRef] [PubMed]
46. Selechnik, D.; Richardson, M.F.; Hess, M.K.; Hess, A.S.; Dodds, K.G.; Martin, M.; Chan, T.C.; Cardilini, A.P.A.; Sherman, C.D.H.; Shine, R.; et al. Inherent Population Structure Determines the Importance of Filtering Parameters for Reduced Representation Sequencing Analyses. *bioRxiv* **2020**, 2020.11.14.383240. [CrossRef]
47. Lourenco, D.A.L.; Fragomeni, B.O.; Tsuruta, S.; Aguilar, I.; Zumbach, B.; Hawken, R.J.; Legarra, A.; Misztal, I. Accuracy of Estimated Breeding Values with Genomic Information on Males, Females, or Both: An Example on Broiler Chicken. *Genet. Sel. Evol.* **2015**, *47*, 56. [CrossRef] [PubMed]
48. VanRaden, P.M. Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [CrossRef] [PubMed]
49. Al-Tobasei, R.; Ali, A.; Garcia, A.L.S.; Lourenco, D.; Leeds, T.; Salem, M. Genomic Predictions for Fillet Yield and Firmness in Rainbow Trout Using Reduced-Density SNP Panels. *BMC Genom.* **2021**, *22*, 92. [CrossRef] [PubMed]
50. Kriaridou, C.; Tsairidou, S.; Houston, R.D.; Robledo, D. Genomic Prediction Using Low Density Marker Panels in Aquaculture: Performance Across Species, Traits, and Genotyping Platforms. *Front. Genet.* **2020**, *11*. [CrossRef]
51. Wang, Q.; Yu, Y.; Yuan, J.; Zhang, X.; Huang, H.; Li, F.; Xiang, J. Effects of Marker Density and Population Structure on the Genomic Prediction Accuracy for Growth Trait in Pacific White Shrimp *Litopenaeus Vannamei*. *BMC Genet.* **2017**, *18*, 45. [CrossRef]
52. Christoffersen, C.; Federspiel, C.K.; Borup, A.; Christensen, P.M.; Madsen, A.N.; Heine, M.; Nielsen, C.H.; Kjaer, A.; Holst, B.; Heeren, J.; et al. The Apolipoprotein M/S1P Axis Controls Triglyceride Metabolism and Brown Fat Activity. *Cell Rep.* **2018**, *22*, 175–188. [CrossRef]
53. Yang, L.; Li, T.; Zhao, S.; Zhang, S. Expression of Apolipoprotein M and Its Association with Adiponectin in an Obese Mouse Model. *Exp. Ther. Med.* **2019**, *18*, 1685–1692. [CrossRef]
54. Liu, D.; Pan, J.-M.; Pei, X.; Li, J.-S. Interaction Between Apolipoprotein M Gene Single-Nucleotide Polymorphisms and Obesity and Its Effect on Type 2 Diabetes Mellitus Susceptibility. *Sci. Rep.* **2020**, *10*, 7859. [CrossRef]
55. Sramkova, V.; Berend, S.; Siklova, M.; Caspar-Bauguil, S.; Carayol, J.; Bonnel, S.; Marques, M.; Decaunes, P.; Kolditz, C.-I.; Dahlman, I.; et al. Apolipoprotein M: A Novel Adipokine Decreasing with Obesity and Upregulated by Calorie Restriction. *Am. J. Clin. Nutr.* **2019**, *109*, 1499–1510. [CrossRef] [PubMed]

-
56. Küttner, E.; Moghadam, H.K.; Skúlason, S.; Danzmann, R.G.; Ferguson, M.M. Genetic Architecture of Body Weight, Condition Factor and Age of Sexual Maturation in Icelandic Arctic Charr (*Salvelinus alpinus*). *Mol. Genet Genom.* **2011**, *286*, 67–79. [[CrossRef](#)] [[PubMed](#)]
 57. Norman, J.D.; Danzmann, R.G.; Glebe, B.; Ferguson, M.M. The Genetic Basis of Salinity Tolerance Traits in Arctic Charr (*Salvelinus alpinus*). *BMC Genet.* **2011**, *12*, 81. [[CrossRef](#)] [[PubMed](#)]
 58. Moghadam, H.K.; Poissant, J.; Fotherby, H.; Haidle, L.; Ferguson, M.M.; Danzmann, R.G. Quantitative Trait Loci for Body Weight, Condition Factor and Age at Sexual Maturation in Arctic Charr (*Salvelinus alpinus*): Comparative Analysis with Rainbow Trout (*Oncorhynchus mykiss*) and Atlantic Salmon (*Salmo salar*). *Mol. Genet Genom.* **2007**, *277*, 647–661. [[CrossRef](#)] [[PubMed](#)]