# AN R-BASED PIPELINE FOR GENOTYPE CALLING AND PARENTAGE ASSIGNMENT OF TRIPLOID OFFSPRING TO THEIR DIPLOID PARENTS

Mathieu Besson, Julien Roche, Pierre Patrice, Pierrick Haffray, François Allal, Marc Vandeputte, Florence Phocas

# AN R-BASED PIPELINE FOR GENOTYPE CALLING AND PARENTAGE ASSIGNMENT OF TRIPLOID OFFSPRING TO THEIR DIPLOID PARENTS

M. Besson[a*], J. Roche[a], P. Patrice[a], P. Haffray[a], F. Allal[b], M. Vandeputte[b,c], F. Phocas[c]

[a] SYSAAF (French Poultry and Aquaculture Breeders Technical Centre), 35042 Rennes, France
[b] MARBEC, University of Montpellier, CNRS, Ifremer, IRD, 34250 Palavas-les-Flots, France
[d] Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France
[*] Presenting author: mathieu.besson@inrae.fr

## Introduction

Triploids, which are individuals bearing three sets of chromosomes, are commonly used in fish breeding. Triploids are sterile which allow (1) higher growth after maturity and feed efficiency, (2) better flesh quality and (3) lower risk of genetic introgression into wild populations. In rainbow trout where triploids account for a significant proportion of production, selective breeding programs are all performed on diploid lines expecting transfert of the genetic progress to their triploid progenies.

. However, evaluating only diploid performance is suboptimal as genetic correlations for the same trait between diploids and triploids may differ from unity. Hence, maximizing genetic gain in diploid lines requires the evaluation of breeding values on triploid sibs of the diploid candidates. In mixed-family designs of breeding programs, this implies correctly genotyping triploids and recovering their pedigree, which is not currently possible with open source tools. Here, we present a pipeline to genotype and assign triploids in fish obtained by second polar body retention after fertilization (the technique applied in salmonids) to their diploid parents based on a newly developed R package, GenoTriplo and an adaptation of APIS (Griot et al., 2020) software.

## Material and Methods

We genotyped 1,232 triploid offspring obtained from 190 diploid dams and 98 diploid sires of rainbow trout on a Thermo Fisher SNP chip array with 57501 SNPs, from which we kept the 38,033 highest quality markers. Allele signals were obtained via Axiom Analysis Suite software as the luminescence of probesets A and B ($S_A$ and $S_B$) for each marker and individual. GenoTriplo was designed in two steps, clustering and genotype calling. For the clustering, each individual was represented by a couple of coordinates (x the contrast and y the signal strength) for each marker.

$$x = \text{Contrast} = \log2\left(\frac{S_A}{S_B}\right)$$

$$y = \text{Signal Strength} = \frac{\log2(S_A) + \log2(S_B)}{2}$$

Rmixmod R package was used to find clusters of individuals sharing the same genotype for a given marker. In the iterative algorithm, the initial number of clusters was set to 8, i.e. twice to the maximum number of possible genotypes. Clusters with close contrast values were merged to keep only four or less clusters. Five iterations were realized and we kept the iteration with the highest likelihood value. For the genotype calling, we identified the more extreme cluster based on its contrast (x) mean value and set it as homozygous (AAA if mean(x)>0 or BBB if mean(x)<0). Then, the remaining clusters were ordered by contrasts. Three main criteria were involved to improve cluster precision and to discriminate low-quality markers. (1) Genotype was declared as missing if the probability of the individual to belong to the cluster was below 0.85. (2) Genotype was also set missing if the distance between the cluster center and the individual exceeds $2.8*SD_{cluster}$. (3) Genotypes of all individuals of the same cluster were all set missing if the standard deviation of the cluster exceed $0.28*(1+0.5*abs(Mean_{cluster}))$. There were three categories of markers: PolyHighRes markers with 3 alleles observed and 4 genotypes, NoMinorHom markers with one homozygous genotype missing and CRbelowThreshold markers with call rate below 90%. Finally, we adapted APIS to enable parentage assignment of triploids.

To do so, we created three new likelihood tables, one for each possible offspring genotype (AAA, ABB, ABC – this latter being relevant only for microsatellite markers). Similar tables were produced to perform assignment by exclusion. To test whether this pipeline was efficient, we used APIS to assign the 1,232 triploid offspring genotyped with GenoTriplo. The assignment was done using the exclusion method with the 1,000 best markers selected on the highest values for minor allele frequency and call rate.

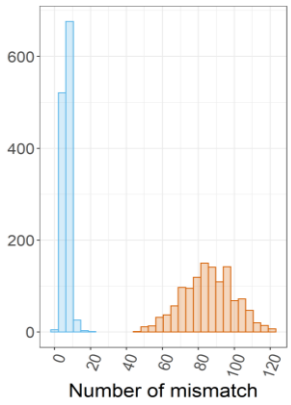


**Figure 1**. Number of offspring as a function of mismatches for the best parental couple (blue) and the second-best parental couple (red).

## Results

Thanks to the 8 clusters initially targeted, most of the rare genotypes were detected. This resulted in minimizing the proportion of CRbelowThreshold markers and maximizing the proportion of informative NoMinorHom and PolyHighRes markers (Table 1) in comparison to Grashei et al. (2020)' method. Furthermore, all offspring were assigned to their parents. The best couple category had a maximum of 19 mismatches for 1,000 markers and a mean number of mismatches of 6.9 which is low (<1% of mismatch). The second-best couple had a minimum of 47 mismatches for a mean of 85.6 (Figure 1).

**Table 1**. Number of markers per category in diploid parents and triploid offspring datasets.

| Population and method | PolyHighRes | NoMinorHom | CRbelowThreshold |
|---|---|---|---|
| Diploid parents using Axas | 31,096 | 2,855 | 3,485 |
| Triploid offspring using GenoTriplo | 21,715 | 6,233 | 4,734 |

## Discussion

Initially targeting eight clusters instead of four ones is a major difference between GenoTriplo and the method proposed by Grashei et al. (2020). It enabled the algorithm to identify clusters with few individuals that would have been included in a bigger cluster and which would have then been discarded due their distance to the cluster's center. Consequently, GenoTriplo maximizes the number of informative PolyHighRes markers found. Meanwhile, we showed that the accuracy of GenoTriplo allowed to assign all our triploid offspring to their diploids parents. Even though the true parents were unknown the difference in the numbers of mismatches between the best and the second-best couple was large enough to confirm the success of the assignment. Hence, this pipeline is a major breakthrough towards the selection of diploid lines based on performances of their triploid sibs.

## Acknowledgements

The HYPOTEMP project ( n°PFEA470019FA1000016) was funded by the French Government and the European Union (EMFF, European Maritime and Fisheries Fund) at the "Appels à projets Innovants" managed by the France Agrimer Office.

## References

Grashei, K.E., Ødegård, J., Meuwissen, T.H.E., 2020. Genotype calling of triploid offspring from diploid parents. Genetics Selection Evolution 52, 15.

Griot, R., Allal, F., Brard-Fudulea, S., Morvezen, R., Haffray, P., Phocas, F., Vandeputte, M., 2020. APIS: An auto-adaptive parentage inference software that tolerates missing parents. Mol. Ecol. Resour. 20, 579–590.