Genetics of Aquatic Organisms 7(3), GA676
https://doi.org/10.4194/GA676

RESEARCH PAPER

Cross-species Application of a SNP array: Application for Analysis of Genetic Diversity in Chum Salmon (*Oncorhynchus keta*) Populations

Min Sung Kim[1], Ju Ri Kwak[1], Jae Yong Han[1], Ji Been Park[1], Ju Yeong Kang[1], Woo-Jai Lee[1,*]

[1]BLUGEN, Busan, Republic of Korea

How to Cite

Kim, M.S., Kwak, J.R., Han, J.Y., Park, J.B., Kang, J.Y., Lee, W.-J. (2023). Cross-species Application of a SNP array: Application for Analysis of Genetic Diversity in Chum Salmon (*Oncorhynchus keta*) Populations. *Genetics of Aquatic Organisms, 7(3), GA676*. https://doi.org/10.4194/GA676

Article History
Received 08 August 2023
Accepted 18 October 2023
First Online 15 November 2023

Corresponding Author
Tel.: +827087719090
E-mail: woojailee@blugen.co.kr

Keywords
Chum salmon
Oncorhynchus keta
Single nucleotide polymorphisms
Cross-species application
Genetic diversity

Abstract

We tested the cross-species application of the Coho salmon SNP array on two populations of Chum salmon from South Korea. Of the 203,077 SNPs, 1,838 (~1%) only represented common SNPs between the two populations. Diversity analysis showed salmon in the ND region formed a narrow cluster, whereas salmon in the TW region formed another cluster more widely distributed than the ND cluster. It seems obvious that the cross-species application of a SNP array can be limited in terms of the available SNPs for most studies such as breeding purposes and target gene identifications, but still worth trying for local population studies.

Introduction

Single nucleotide polymorphism (SNP) arrays containing hundreds of thousands of SNPs are commonly used in genetic studies of model species. However, in non-model organisms, the discovery of useful SNPs can be time- consuming and methodologically complex (Fountain *et al*., 2021). Relatively few arrays have been developed for non-model species because of the high costs associated with their development. Therefore, once a SNP chip is developed for one species, it is often adapted and applied to closely related species; this is known as the cross-species application of SNP arrays. By leveraging existing SNP arrays, researchers can overcome the challenge of developing high-cost arrays specific to each non-model species (Miller *et al*., 2012). Several commercially available SNP arrays for major model species such as Coho and Atlantic salmon species have been published (Thermo or other citations), but not much information available for the applicability of the published arrays for the genomic studies of a different species. Both Coho and Chum salmon are pacific salmon species and it is worth trying to investigate the applicability of a SNP array from Coho salmon to a closely related Chum salmon species.

Chum salmon (*Oncorhynchus keta*) has a wide geographic distribution ranging from the Arctic to California in North America and from Siberia to Japan and Korea in East Asia (Salo 1991). The Republic of Korea

is situated in the southernmost part of the Chum salmon migration range (Kim *et al*., 2007; Kwon *et al*., 2014), and Chum salmon are known to spawn in rivers, including the Bukcheon and Namdaecheon streams in Gangwon-do, Osipcheon in Samcheok, Taehwagang in Ulsan, during the fall season (October- November) each year (Kwon *et al*., 2014). While the release of young Chum salmon for a restocking program has been carried out annually in each region's rivers, the rate of salmon return has been less than 1% since the 2000s (Kim *et al*., 2007). It is helpful to distinguish between salmon populations and acquire a comprehensive understanding of their biological characteristics to improve the recovery rate of the released salmon resources and to effectively manage them in marine ecosystems.

The aim of this study was to investigate the cross-species applicability of a Coho salmon chip and to get more insights into the genetic differences between two Chum salmon populations, Chum salmon from Namdaecheon, where the largest natural salmon hatching takes place in Korea, and Chum salmon from Taehwagang, using a 200 K SNP array designed for Coho salmon. We assessed the potential of cross-species applications for genetic diversity analysis using a SNP array from closely related species.

Materials and Methods

We collected a total of 96 tissue samples of Chum salmon, 48 samples from the Namdaecheon Stream (ND) and another 48 samples from Taehwagang (TW) respectively in South Korea (Figure 1), and Genomic DNA was extracted using the DNeasy® 96 Blood and Tissue Kit (QIAGEN, Valencia, CA, USA), following the manufacturer's standard protocols. The DNA quality was confirmed using electrophoresis on a 1% agarose gel and quantified using a NanoDrop spectrophotometer.

Coho salmon (*O. kisutch*) from Thermo Fisher (Product #550862). Genotyping was performed using a commercially available Coho salmon chip.

An Affymetrix GeneTitan Multi-Channel Instrument was used for staining, washing, and scanning the array signals, following the manufacturer's protocol. Genotype calls were conducted using Axiom Analysis Suite software v3.1 (Thermo Scientific), as recommended by the Axiom Analysis user guide. QC of genotyping was performed using PLINK v1.09 and assessed separately for each population (Barria *et al*., 2018). All subsequent analyses were performed using the common SNPs in both populations after QC.

PCA and admixture analyses on the genotyping data from the 96 individuals were conducted using R

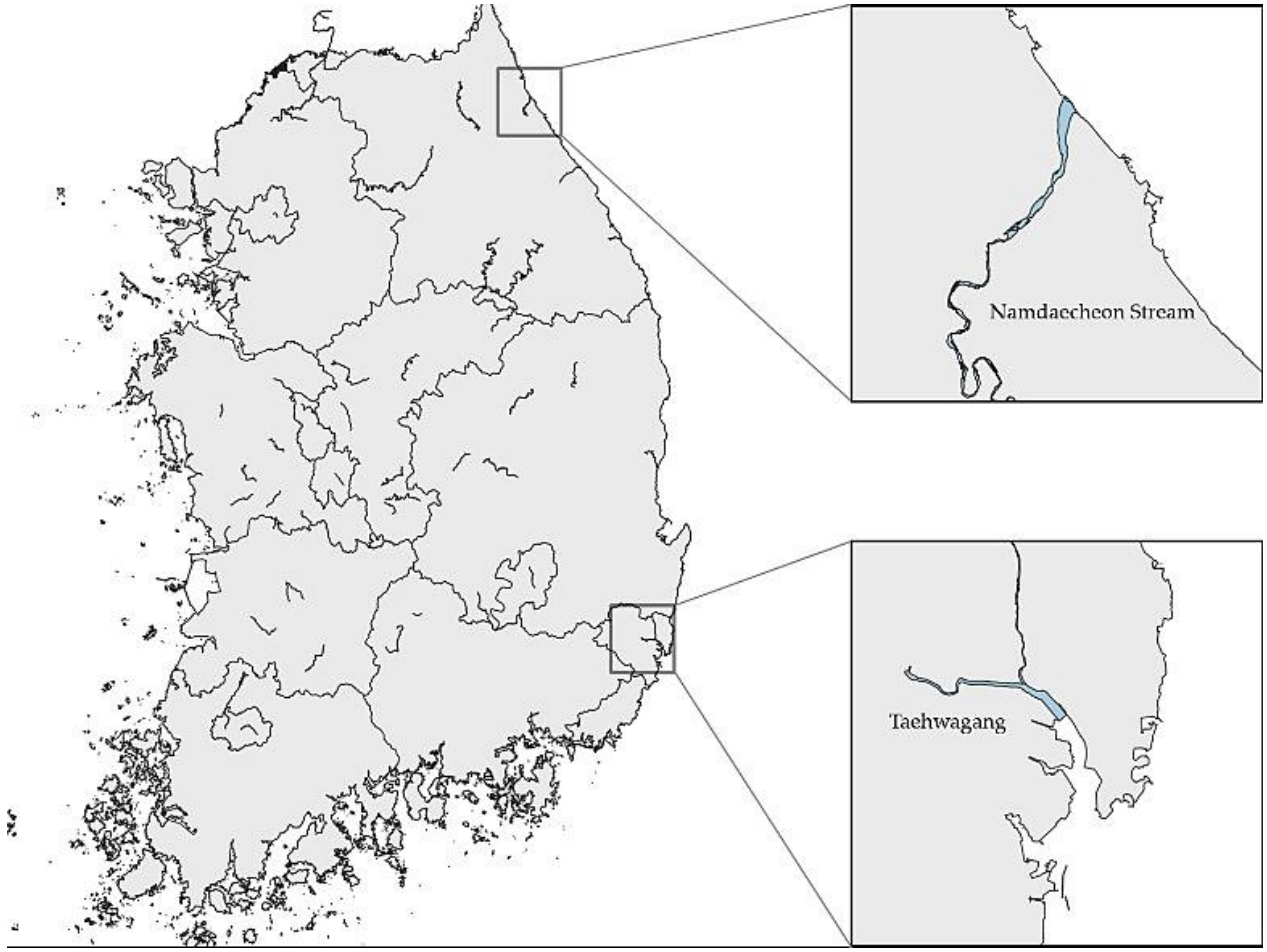

Figure 1. Location of the sampling area.

v1.09 (Alexander *et al.*, 2009; Jombart 2008), and the resulting first- and second-component values were plotted on a 2-D scatterplot.

Results

Genotype analysis was performed using a 200 K Coho salmon SNP array. Of the 203,077 SNPs, 112,976 SNPs were selected based on the categories recommended by Thermo Fisher (PolyHighResolution, NoMinorHomozygous, and MonoHighResolution). Subsequently, we filtered out the selected SNPs with a missing genotyping rate > 0.05, SNP missing rate > 0.05, and Minor Allele Frequency (MAF) < 0.05. We confirmed 1,838 (~1%) SNPs that all 96 samples met the genotyping criteria, and 111,138 (99%) SNPs were excluded. (Table 1). And we utilized a total of 1,838 SNPs that were commonly identified for subsequent analysis.

Principal component analysis (PCA) and admixture analyses were performed to investigate the genetic relationships between the two populations. Based on the PCA results, the two populations were classified according to the first principal component (PC1), which explained approximately 10.6% of the genetic variation, and the second principal component (PC2), which explained approximately 7% of the variation (Figure 2a). The ND population showed a relatively clustered distribution of individuals, whereas the TW population showed a wide distribution range, with some overlapping individuals (Figure 2a). In the admixture analysis, the optimal number of clusters was determined to be two, and the membership coefficients for each cluster were calculated for all individuals. The ND population showed a clear structural difference, with a high membership coefficient (above 0.99) in Cluster 1, whereas the TW population showed membership coefficients of 0.28 and 0.72 in Clusters 1 and 2, respectively (Figure 2b).

Discussion

SNPs are high-quality and convenient genotyping platforms for population genetic studies (Zhou *et al*., 2021). Recently, a high-density 200 K SNP genotyping array was developed and validated in Coho salmon population (Messmer *et al*., 2018). However, a SNP array for Chum salmon used in this study has not yet been developed because developing a SNP array is time-consuming, costly, and involves complicated data interpretation. Thus, applying commercial SNP arrays across closely related different species is a cost-effective and efficient approach for genotyping SNPs that requires minimal equipment and expertise (Miller *et al*., 2012).

Furthermore, cross-species application provides a potential method for rapid marker development if genomic resources for closely related relatives are available (Messmer *et al*., 2018). Based on studies conducted by Ogden *et al*. (2012) and Hoffman *et al*. (2013), which involved cross-species microarray analysis between distantly related species, such as oryx and modern domesticated bovines, Antarctic fur seals, and domestic dogs, it has been demonstrated that genetic analysis can be successfully conducted across such species. Considering the estimated genetic divergence of approximately seven million years between Coho and Chum salmon, our study further supports the feasibility of utilizing cross-species microarray analysis to investigate genetic variations between these two species. Therefore, this study was conducted using a 200 K SNP genotyping array designed from Coho salmon, for Chum salmon populations which is evolutionarily similar to Coho salmon.

Of the 203,077 SNPs analyzed in the 96 salmon individuals, 1,838 were filtered and used for subsequent analysis. It is likely that approximately 1% of the SNP loci were useful and analyzed. Compared with previous studies using cross-species microarray analysis, our findings demonstrated overall consistency with similar results. For example, Ogden *et al*. (2012) genotyped Scimitar-horned and Arabian oryx using the Illumina BovineSNP50 BeadChip and reported 148 and 149 polymorphic SNPs (0.27%, 0.28%), respectively. Likewise, SNP genotyping in dromedary (Bertolini *et al*., 2017) was performed using an Illumina Ovine SNP 600K BeadChip, which reported 14,179 SNPs (2.3%).

In the first stage of quality control (QC), 112,976 SNPs were selected based on the recommended categories. SNPs with a missing rate of less than 95% were removed; however, no SNPs were excluded. The final step of the QC process was to remove all SNPs with very low MAF. Filtering less than 95% of the MAF using the 112,976 SNP resulted in 1,838 high-quality SNPs for downstream analysis (111,138 SNPs excluded). All SNPs

Table 1. Counts of selected samples and markers in this study.

Description	No. of samples	No. of SNPs
Total number of samples	96	
Samples with a missing genotyping rate >0.05	0	
Selected samples	96	
Total number of SNP		112,976
SNP missing rate <0.05		0
MAF (Minor Allele Frequency) <0.05		111,138
Selected SNPs for analysis		1,838

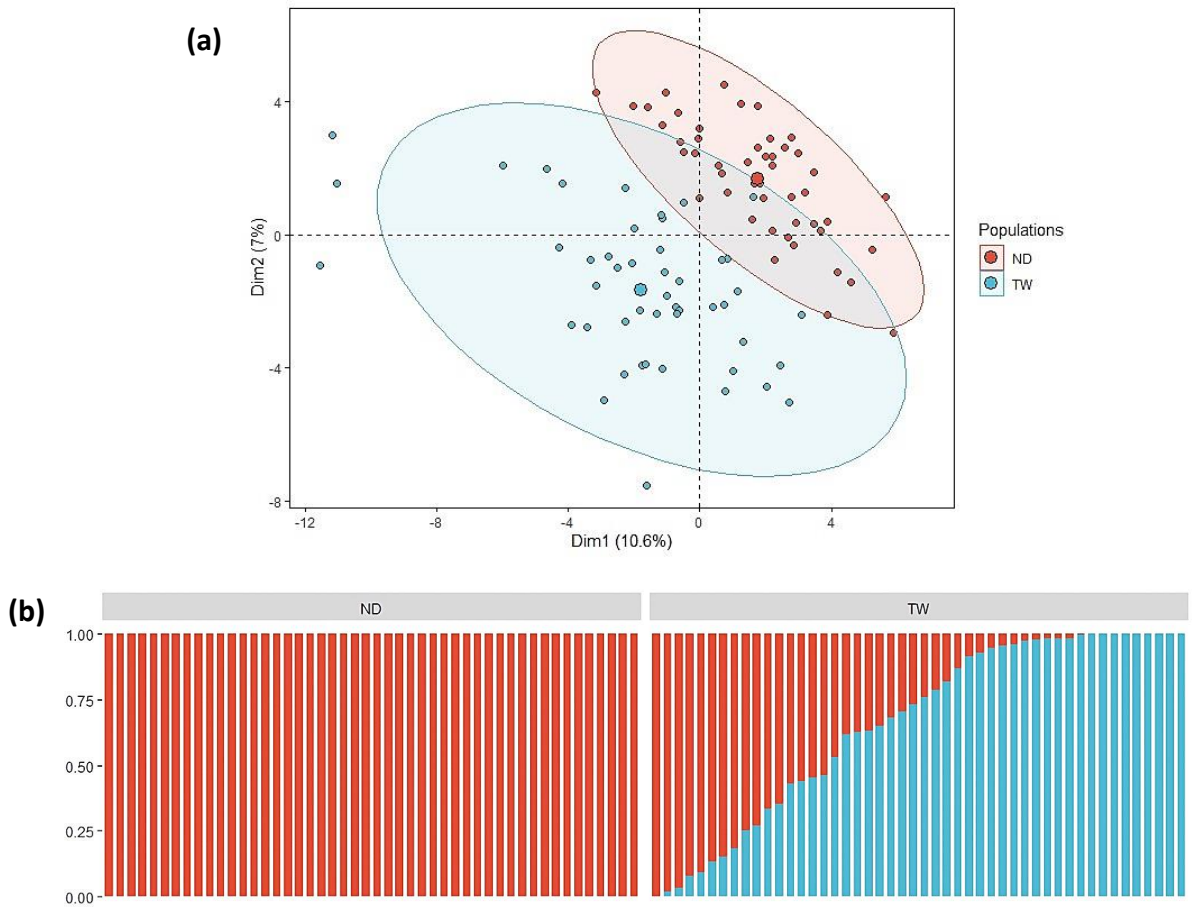

Figure 2. (a) Principal component analysis (PCA) and (b) Admixture analysis based on 1,838 SNPs. Color and sign code: red, populations from Namdaecheon Stream (ND); blue, populations from Taehwagang (TW).

passed the missing rate filter; however, a significant portion of the SNPs were filtered out based on the MAF criterion. This indicates that the filtered SNPs are specific to Coho salmon and are not present in Chum salmon, suggesting independent evolution or lineage-specific mutations at these SNP positions.

PCA and admixture analysis showed that the ND population formed a single narrow cluster with a uniform genetic structure. In contrast, the TW population showed no clear clusters and was widely dispersed, possibly due to the fact that Ulsan Taehwagang salmon has been affected by the restocking program. Since 2000, a small-scale release program has been underway in Taehwagang to enhance salmon resources, which involves the release of young salmon captured and produced in the Yangyang Namdaecheon stream (Kim *et al*., 2007). Therefore, some individuals appeared to share genetic variation with the ND population.

In the present study, we assessed the cross-species application of SNP arrays and conducted a genetic diversity analysis of domestic salmon populations. By utilizing a SNP array that was genetically similar to Coho salmon, we observed the differentiation of the Chum salmon population. Here we present the results of population genetic analysis using a SNP array designed from a different species and this would be a good case study for researchers considering a cross-species application of SNP arrays. These results provide valuable data for biodiversity conservation and management of the Chum salmon population in Korea.

Ethical Statement

This manuscript is not related to any form of research misconduct, including double publication, plagiarism/self-plagiarism, fabrication, and falsification of research at the point of submitting this manuscript to Genetics of Aquatic Organisms. If this research manuscript is withdrawn due to the violation of Research Ethics Statement, I will waive all relevant rights and benefits.

Author Contribution

Conceptualization, M.S.K.; data collection and curation, J.R.K. and J.B.P.; data analysis, M.S.K. and J.Y.H.; writing original draft preparation, J.R.K., J.B.P. and J.Y.K; supervision and project administration, W.J.L.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

This research was supported by Korea Fisheries Institute of Marine Science & Technology Promotion (KIMST) funded by the Ministry of Oceans and Fisheries (20220570).

References

Alexander, D. H., Novenbre, J., Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655-64. https://doi.org/10.1101/gr.094052.109

Barria, A., López, M. E., Yoshida, G., Carvalheiro, R., Lhorente, J. P., & Yáñez, J. M. (2018). Population Genomic Structure and Genome-Wide Linkage Disequilibrium in Farmed Atlantic Salmon (*Salmo salar* L.) Using Dense SNP Genotypes. *Frontiers in genetics*, 9, 649. https://doi.org/10.3389/fgene.2018.00649

Bertolini, F., Elbeltagy, A., Rothschild, M. (2017). Evaluation of the application of bovine, ovine and caprine SNP chips to dromedary genotyping. *Livestock Research Rural Development*, 29(2), article 31. https://www.lrrd.cipav.org.co/lrrd29/2/bert29031.html

Fountain, E. D., Zhou, L. C., Karklus, A., Liu, Q. X., Meyers, J., Fontanilla, I. K., Rafael, E. F., Yu, J. Y., Zhang, Q., Zhu, X. L., Pei, E. L., Yuan, Y. H., Banes, G. L. (2021). Cross-species application of Illumina iScan microarrays for cost-effective, high-throughput SNP discovery. *Frontiers in Ecology and Evolution*, 9, 629252. https://doi.org/10.3389/fevo.2021.629252

Hoffman, J. I., Thorne. M. A. S., McEwing, R., Forcada, J., Ogden, R. (2013). Cross-amplification and validation of SNPs conserved over 44 million years between seals and dogs. *PLoS One*, 8(7), 1–10. https://doi.org/10.1371/journal.pone.0068365

Jombart, T. (2008). Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403-1405. https://doi.org/10.1093/bioinformatics/btn129

Kim, G. E., Kim, C. G., Lee, Y. H. (2007). Genetic similarity-dissimilarity among Korea Chum salmons of each stream and their relationship with Japan salmons. *The Sea: Journal of the Korean Society of Oceanography*, 12(2), 94-101. https://www.researchgate.net/publication/264031176_Genetic_Similarity-dissimilarity_Among_Korea_Chum_Salmons_of_Each_Stream_and_Their_Relationship_with_Japan_salmons

Kim, S. A., Lee, C. S., Kang, S. K. (2007). Present status and future prospect in salmon research in Korea. *The Sea: Journal of the Korean Society of Oceanography*, 12(2), 57-60. https://koreascience.kr/article/JAKO200721138097353.page

Kohn, M. H., Murphy, W. J., Ostrander, E. A., Wayne, R. K. (2006). Genomics and conservation genetics. *Trends in ecology & evolution*, 21(11), 629-637. https://doi.org/10.1016/j.tree.2006.08.001

Kwon, O. N., Kim, J. K., Yoon, M. G., Kim, D. H., Hong, K. E. (2014). Marine prey selectivity of released juvenile chum salmon (Oncorhynchus keta) during arly marine migration in Korean waters. *Journal of Fisheries and Marine Sciences Education*, 26(2), 421-429. https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE02400808

Malhi, R. S., Trask, J. S., Shattuck, M., Johnson, J., Chakraborty, D., Kanthaswamy, S., Ramakrishnan, U., Smith, D. G. (2011). Genotyping single nucleotide polymorphisms (SNPs) across species in Old World Monkeys. *American Journal of Primatology*, 73(10), 1031-1040. http://doi.org/10.1002/ajp.20969

Messmer, A. M., Leong, J. S., Rondeau, E. B., Mueller, A., Despins, C. A., Minkley, D. R., Kent, M. P., Lien, S., Boyce, B., Morrison, D., Fast, M. D., Norman, J. D., Danzmann, R. G., Koop, B. F. (2018). A 200K SNP chip reveals a novel Pacific salmon louse genotype linked to differential efficacy of emamectin benzoate. *Marine Genomics*, 40, 45-57. http://doi.org/10.1016/j.margen.2018.03.005

Miller, J. M., Kijas, J. W., Heaton, M. P., McEwan, J. C., Coltman, D. W. (2012). Consistent divergence times and allele sharing measured from cross-species application of SNP chips developed for three domestic species. *Molecular Ecology Resources*, 12(6), 1145-1150. https://doi.org/10.1111/1755-0998.12017

Ogden, R., Baird, J., Senn, H., McEwing, R. (2012). The use of cross-species genome-wide arrays to discover SNP markers for conservation genetics: A case study from Arabian and scimitar-horned oryx. *Conservation Genetics Resources*, 4, 471–473. https://doi.org/10.1007/s12686-011-9577-2

Salo, E. O. (1991). Life history of chum salmon (Oncorhynchus keta). UBC Press.

Zhou, Q., Chen, Y. D., Lu, S., Liu, Y., Xu, W. T., Li, Y. Z., Wang, L., Wang, N., Yang, Y. M., Chen, S. L. (2021). Development of a 50K SNP array for Japanese flounder and its application in genomic selection for disease resistance. *Engineering*, 7(3), 406-411. https://doi.org/10.1016/j.eng.2020.06.017 s