BALLARD FISH & OYSTER CO. LLC DBA
CHERRYSTONE
AQUA-FARMS

P: (757) 331 1208

W: www.clamandoyster.com

A: 1588 Townfield Dr. Cape Charles, VA 23310

VIRGINIA FISHERY RESOURCE GRANT PROGRAM

FINAL REPORT

FRG 2024-01

Making genetic resources easily accessible for the hard clam (*Mercenaria mercenaria*) aquaculture industry in Virginia

Submitted by:

Alexandra McCarty
Ballard Fish & Oyster Company, LLC
Cherrystone Aqua-Farms

Project Partners:

Karen Hudson, Virginia Institute of Marine Science
Jan McDowell, Virginia Institute of Marine Science
Kimberly Reece, Virginia Institute of Marine Science
Richard Snyder, VIMS Eastern Shore Laboratory

Ann Gallivan, JC Walker Brothers Inc
Susan Miller, Bagwell Enterprises Inc

To:

Fisheries Resource Grant Program
Marine Advisory Program
Virginia Institute of Marine Science
Gloucester Point, VA

April 18, 2025

Introduction

The hard clam (*Mercenaria mercenaria*) aquaculture industry in Virginia leads the nation in hard clam production and is the largest contributor to Virginia's shellfish aquaculture economic value (Hudson, 2019). Due to the higher salinity requirements of hard clams, aquaculture production mainly occurs on the Eastern Shore of Virginia (ESVA) and in the lower Chesapeake Bay. Cherrystone Aqua-Farms, located in Cape Charles on the ESVA, was founded in 1895 and has been producing aquaculture clams since 1983. We are a vertically integrated company, meaning we produce clam seed in our hatchery from our broodstock lines and then grow product until harvest on numerous leases located in the Bay and seaside of Virginia. Currently, we are the largest producer of hard clams on the East Coast and employ over 200 direct employees and hundreds more through cooperative growing programs. To put it lightly, clam aquaculture is the livelihood of our company.

Clam aquaculture companies in Virginia, including Cherrystone, conduct separate breeding efforts and maintain proprietary broodstock lines in-house. Company success starts with seed production, where proprietary broodstock lines must produce enough viable seed to meet company production and sales goals. Hatcheries strive to maintain genetic diversity and minimize inbreeding within their broodstock lines to ensure their product thrives and survives, as inbreeding can lead to reduced fitness (growth and reproduction) and mortality (e.g. Evans et al., 41 2004, Ibarra et al., 1995, Wada and Komaru, 1994, Deng et al., 2005, Zheng et al., 2012). The negative impacts from selective breeding in a hatchery can be detrimental to a business and an entire industry if seed is supplied from a few hatcheries. Due to the importance and proprietary nature of these lines, companies have been unwilling to share genetic material or trade secrets with each other or with outside individuals (i.e. extension specialists, scientists).

A recent collaborative project in 2023 led by PI McCarty (in her former role as a Shellfish Aquaculture Research and Extension Specialist for VIMS Marine Advisory Program, McCarty at al., 2025) gained industry trust to move forward with assessing the genetic health of their lines. This project demonstrated the utility of genetic tools to the Virginia hard clam aquaculture industry by analyzing genetic diversity and inbreeding metrics of commercial hard clams sourced from retail stores in Virginia in relation to wild hard clam samples from the east coast of the United States (previously presented to industry, Ropp et al., 2023). Results were presented to industry in a comprehensive report with supporting resources during an in-person workshop. Industry responded with requests for additional expertise and resources to directly assess the health of their broodstock lines.

This project developed the genetic resources necessary to assess the genetic health (diversity and inbreeding) for hard clam broodstocks. Until this project, no such resources existed, and genetic assessments were nearly impossible for companies to conduct independently. First, we worked with experts at VIMS and specialists at a contract service organization (Center for Aquaculture Technologies in San Diego, CA) to develop an affordable low-density genotyping tool that can be used to address questions regarding genetic diversity and inbreeding. Second, willing industry members in Virginia used the newly developed genotyping tool to assess the genetic health of their proprietary broodstock lines.

Methods

i. Low-density genotype tool development

We (Cherrystone Aqua-Farms) facilitated the development of a low-density genotyping array using previously generated genomic data. Previous efforts sequenced DNA from three commercial hard clam companies in Virginia using the DArTseq method (DArT Pty Ltd, Canberra, Australia). Sequence data for the commercial groups were combined with sequence data previously generated for 16 wild populations sampled along the East Coast of the United States (Prince Edward Island, Canada to Florida, Ropp et al., 2023). Single nucleotide polymorphisms (SNPs) were called from the combined sequence data (3 commercial stocks and 16 wild populations). Raw SNPs were filtered to ensure only high-quality markers remained. Filtered SNPs (2,435 SNPs) were given to experts at the Center for Aquaculture Technologies (CAT, San Diego, CA) for panel development. Details regarding sampling design, genotyping, and SNP filtering can be found in a recent publication by McCarty and project partners from the Virginia Institute of Marine Science (2025).

Experts at CAT selected a subset of markers from the filtered SNP file to include in a low-density genotyping panel. Markers representative of the three cultured populations (i.e. hatchery-spawned) were prioritized to ensure the panel would be useful to the hard clam industry. Details regarding marker selection, design, and panel validation are proprietary to CAT.

ii. Assess genetic health of industry broodstock lines

Participating industry members (JC Walker Brothers Inc., Bagwell Enterprises Inc., Cherrystone Aqua-Farms) sampled 80 hard clams from each broodstock line they wanted to assess. For each hard clam sample, a piece of mantle tissue was preserved in 95% ethanol in pre-labeled microcentrifuge tubes. Tubes were stored in a refrigerator for at least 3 days to ensure tissue fixation. After (at least) 3 days in the refrigerator, ethanol was removed from each tube. Fixed tissue samples were delivered to PI McCarty at Cherrystone Aqua-Farms. All sampling materials and a tissue sampling and preservation protocol (see Appendix for protocol) were delivered to each company and discussed before sampling commenced.

PI McCarty shipped all clam samples (80 clams from 12 groups = 960 total samples) to CAT in San Diego. A total of 10 industry broodstock lines and two wild populations, one from the bayside (Pocomoke Sound) and one from the seaside (Metompkin Bay), were sampled. At CAT, DNA was extracted for each sample and genotyped on the developed low-density genotyping array. A genetic report was delivered to each company with other company samples masked to ensure confidentiality was maintained. Each report discussed genetic diversity and inbreeding metrics of each company's broodstock line(s) in relation to the two wild populations.

Each company was given contact information for the specialists at CAT for any follow-up questions. PI McCarty reached out to participating companies soliciting feedback and asking if there were any outstanding questions warranting further meetings, projects, or collaborative efforts.

Results

i. Low-density genotype tool development

The final low-density panel contained a total of 451 markers. Marker frequencies were estimated across all 19 populations (16 wild and 3 cultured) to check the relevance of the panel for wild populations. All 451 markers performed well across all populations, except Florida. For the Florida population, ~110 of the 450 markers were monomorphic in the population. This means that there was no variation at these markers (i.e. the same nucleotide), and thus aren't informative. Even so, specialists at CAT communicated that the panel has enough markers to still be useful for Florida populations (~350 markers relevant markers). Results from the panel development are included in the Appendix.

ii. Assess genetic health of industry broodstock lines

Genomic DNA was extracted successfully from all fixed hard clam samples at CAT and all samples were run on the low-density genotyping panel. Individual reports were returned to each company with other companies masked. For each company, the analyses and report narrative describe the genetic diversity metrics and population structure of their broodstock lines relative to the two wild populations. Measurements in the analysis and report include:

1. Diversity: allele frequencies, average number of alleles/polymorphic markers, expected and observed heterozygosity's, Nei's genetic diversity, Weir ad Cockerham's pairwise F_{ST} estimates, contemporary effective population size (Ne) estimates using the linkage disequilibrium (LD) method (Waples & Do, 2010),
2. Inbreeding: inbreeding coefficients (F_{IS}),
3. Relatedness: Lynch-Ritland moment estimator,
4. Population structure: Principle Component Analysis (PCA), STRUCTURE.

Due to the proprietary nature of the broodstock analyzed in these reports, they will not be included. However, an example of a sample genetic report from CAT is included in the Appendix for those interested in the context.

After reports were received, industry members were given contact information for the specialists at CAT (available in the results report from CAT, see Appendix). Aside from thank yous, there was no mention of how useful this project was for each company or if they have moved forward with further analyses. As for Cherrystone, we requested the raw genotype data from CAT and have analyzed it further to answer additional questions. We plan to send additional samples for genotyping this year and plan to implement the use of this tool into our regular operations to ensure we are managing our broodstock appropriately.

Conclusions and Recommendations

i. Implications of this project in Virginia

Virginia is home to the largest hard clam industry in the Unites States, which makes all players direct competitors with one another. While we all battle similar challenges, the success of our individual companies is our top priority. This hinders our ability, willingness, and level of collaboration. However, when there is a common need, industry is willing to work together. Neary all hard clam hatcheries in VA agreed to participate in this project, reflecting the

importance of this topic. The one company that did not participate in the project later communicated a desire and interest to use the resources once the project was complete. This project serves as an example of how the hard clam industry can work together while still maintaining our individuality within the industry.

For Cherrystone (not speaking for the other project partners because no feedback was received), this project was extremely beneficial. It was the first time we investigated the genetics of our animals, which is important since we have been selecting and maintaining our broodstock lines for decades. Underlying genetics can highlight breeding or broodstock management practices that are working or not working, which can be informative for determining best practices to ensure continued hatchery success. Also, knowing the genetics of our broodstock lines can alleviate rearing and husbandry requirements if broodstock lines are actually not genetically different. With genotyping data from this project, we performed some additional (preliminary) analyses to investigate family and parental representation within each of our broodstock lines. We also used the data to look for potential genetic signatures of selection in our different growing areas. In the future, we are going to genotype additional animals to ensure our spawning practices are maintaining adequate levels of diversity. Cherrystone plans to use this tool regularly moving forward when selecting our broodstock lines, as incorporating genetic data into breeding decisions will result in larger and quicker genetic gains.

While no follow-up was received from industry project partners likely due to the complex dynamic of being direct competitors, this project is considered a major success for industry. This project answered the question of: what is the level of genetic diversity and inbreeding of my proprietary broodstock lines? Industry members now have the resources they need (knowledge of the genetic basis of their lines, genotyping tool, and other project resources) to incorporate genetic assessments into their operations. While implementing these assessments into regular operations takes background genetic knowledge, this project connected industry members to experts at the Center for Aquaculture Technologies who can directly advise in a confidential manner. As Cherrystone continues using the tool and learning more about the genetics of hard clams on the ESVA, I envision new questions surfacing among industry members requiring additional collaborative efforts among industry, scientific professionals, and genetic service contractors.

ii. Relevance of this project to the broader hard clam industry

This study created a relevant, affordable (~$11.50/sample for DNA extraction and genotyping at time of project, prices subject to change) low-density genetic tool that is publicly available for all to use. The panel was created prioritizing the cultured hard clam samples, which is important because the hard clam industry uses seed produced in hatcheries. Hatchery spawning, rearing, and field grow out create specific genetic signatures (i.e. domesticated selection, see McCarty et al., 2025 for more details) that are not present in wild populations. While tool development prioritized industry hard clam samples, it also proved useful for wild populations from Canada to Florida, expanding the utility of this tool. Until now, accessible genetic resources for the hard clam did not exist or were only attainable by going through scientific experts, which often breached industry comfort levels. Now, private hard clam companies and industry members have

access to a relevant genetic resource that they can use at their own free will, putting them in control and eliminating any potential breach of confidentiality.

The resources generated during this project provide a blueprint for other industry members looking to assess the genetic health of their broodstock (see Appendix):

- Tissue sampling and preservation protocol - details the supplies needed and steps to follow for sampling and preserving hard clam tissue for genetic analysis.
- Sample Genetic Overview Report provided by CAT.

Armed with these resources and access to the newly developed affordable genotyping tool, the hard clam industry is now able to implement genetic assessments into their operations.

References

Deng, Y., Liu, X., Zhang, G., and Guo, X. 2005. Inbreeding depression and maternal effects on early performance of Pacific abalone. North American Journal of Aquaculture, 67: 231–236.

Evans, F., Matson, S., Brake, J., and Langdon, C. 2004. The effects of inbreeding on performance traits of adult Pacific oyster (*Crassostrea gigas*). *Aquaculture*, 230: 89–98.

Hudson, K. 2019. *Virginia Shellfish Aquaculture Situation and Outlook Report: Results of the 2018 Virginia Shellfish Aquaculture CroFp Reporting Survey*. https://doi.org/10.25773/jc19-y847

Ibarra, A.M., Cruz, P., and Romero, B.A. 1995. Effects of inbreeding on growth and survival of self-fertilized catarina scallop larvae, *Argopecten circularis*. *Aquaculture*, 134: 37–47.

McCarty, A.J., McDowell, J.R., Reece, K.S., Snyder, R.A., Hudson, K. 2025. Application of genetic technology in support of the northern hard clam (*Mercenaria mercenaria*) aquaculture industry. *Frontiers in Aquaculture* doi: 10.3389/faquc.2025.1556440

Ropp, A.J., Reece, K.S., Snyder, R.A., Song, J., Biesack, E.E., McDowell, J.R., 2023. Fine-scale population structure of the northern hard clam (*Mercenaria mercenaria*) revealed by genome-wide SNP markers. *Evol Appl* 16, 1422–1437. https://doi.org/10.1111/eva.13577

Wada, K.T. and Komaru, A. 1994. Effect of selection for shell coloration on growth rate and 643 mortality in the Japanese pearl oyster, *Pinctada fucata martensii*. Aquaculture, 125: 59–65.

Waples, R.S., Do, C. 2010. Linkage disequilibrium estimates for contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. Evol Appl 3(3):244-262. doi: 10.1111/j.1752-4571.2009.00104.x

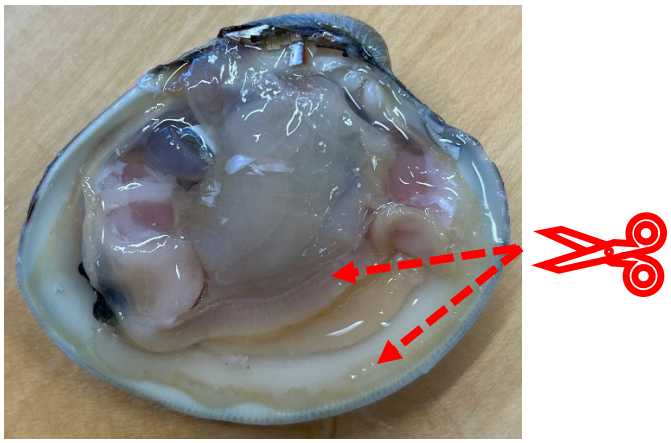
Zheng, H., Li, L., & Zhang, G. 2012. Inbreeding depression for fitness-related traits and purging the 655 genetic load in the hermaphroditic bay scallop *Argopecten irradians irradians* (Mollusca: Bivalvia). 656 *Aquaculture*, *366–367*, 27–33. https://doi.org/10.1016/j.aquaculture.2012.08.029

Appendix

1. Tissue Sampling and Preservation Protocol
2. Results from the low-density hard clam panel development and panel validation by the Center for Aquaculture Technologies (CAT)
3. Example Genetic Overview Report from CAT

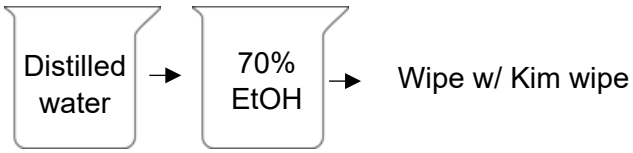
Supplies provided/needed:

- Centrifuge tubes pre-labeled and filled with 1 mL 95% ethanol
- Cardboard tube storage boxes
- 70% ethanol
- Distilled water
- Kimwipes
- Plastic beakers (2)
- Tube rack
- Sampling tools (forceps, scissors)
- Sterile gloves
- Squirt bottle
- Sucking knife (not provided)



For each broodstock line, sample 80 clams according to the protocol below:

1. Shuck clam.
2. Use tweezers and scissors to hold and snip a piece of the mantle tissue (see diagram above) roughly the size of this box: ☐
3. Place tissue in pre-labeled centrifuge tube filled with 1 mL of 95% ethanol. Tissue should be submerged completely by ethanol.
4. Place centrifuge tube in box.
5. Sterilize tools (tweezers and scissors) between samples by rinsing in beakers of:



6. Store box with tissue samples in refrigerator for at least 3 days.
7. After 3 days, remove nearly all the ethanol from the tubes for shipment.
8. Drop-off boxes with samples and log of box organization to Lexy McCarty.

Lexy McCarty
Cherrystone Aqua-Farms
Cell: 240-344-3796

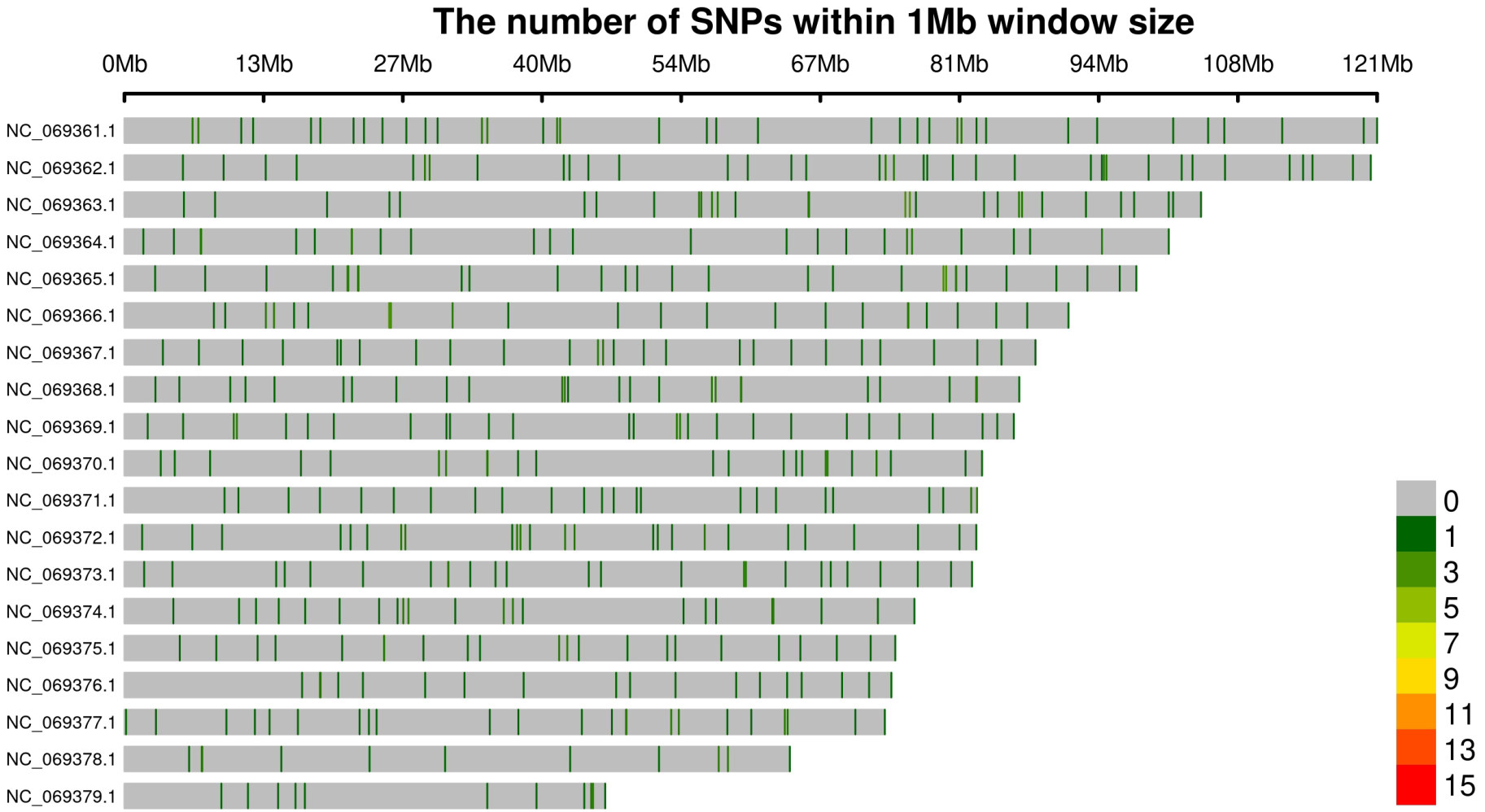
Results from the low-density hard clam panel development and validation

Panel designed and validated by
the Center for Aquaculture Technologies (San Diego, CA).

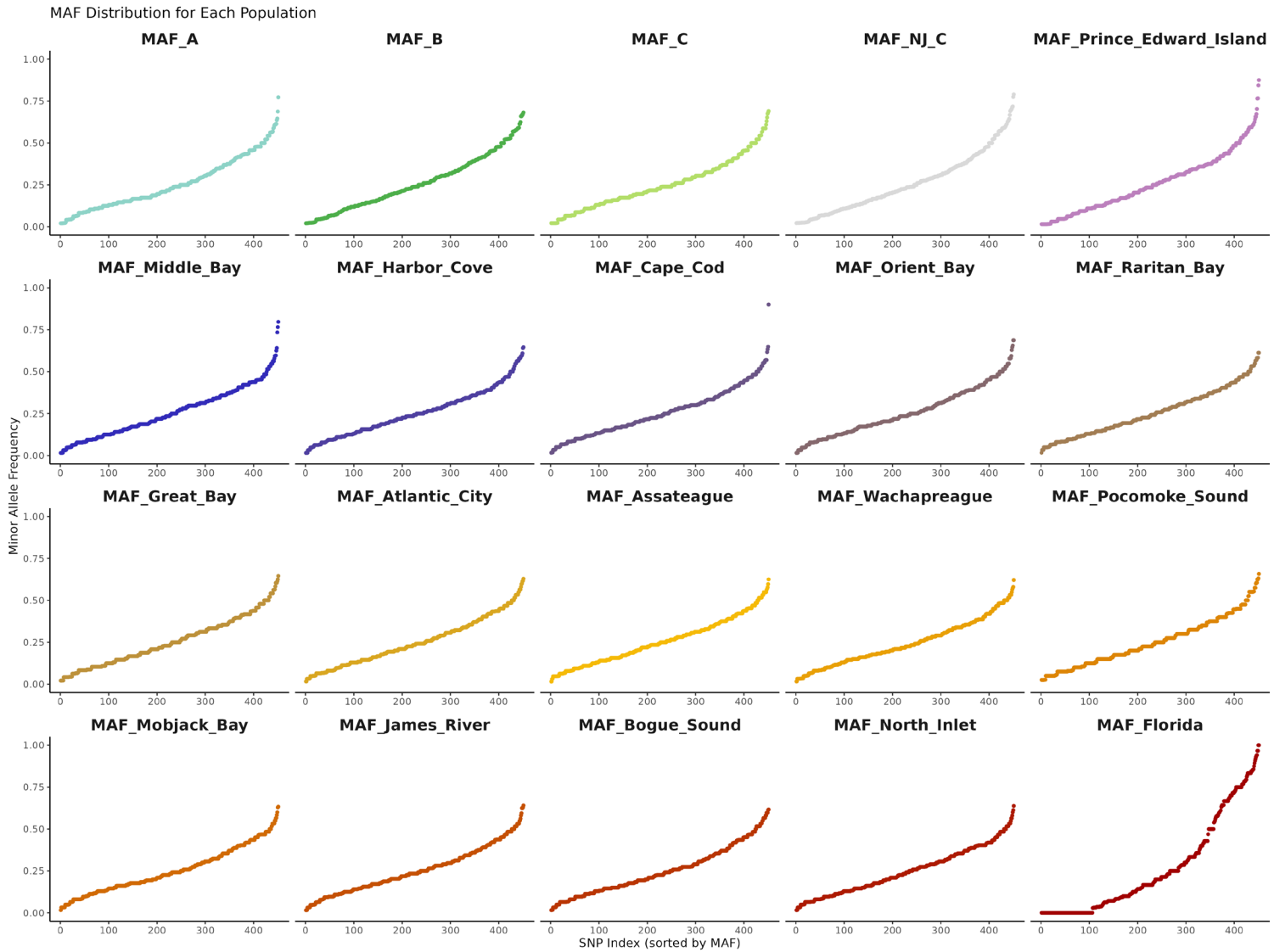
For questions:

- Fill out the online contact form: https://aquatechcenter.com/contact/
- Call (+01-858-450-2972)
- Email info@aquatechcenter.com

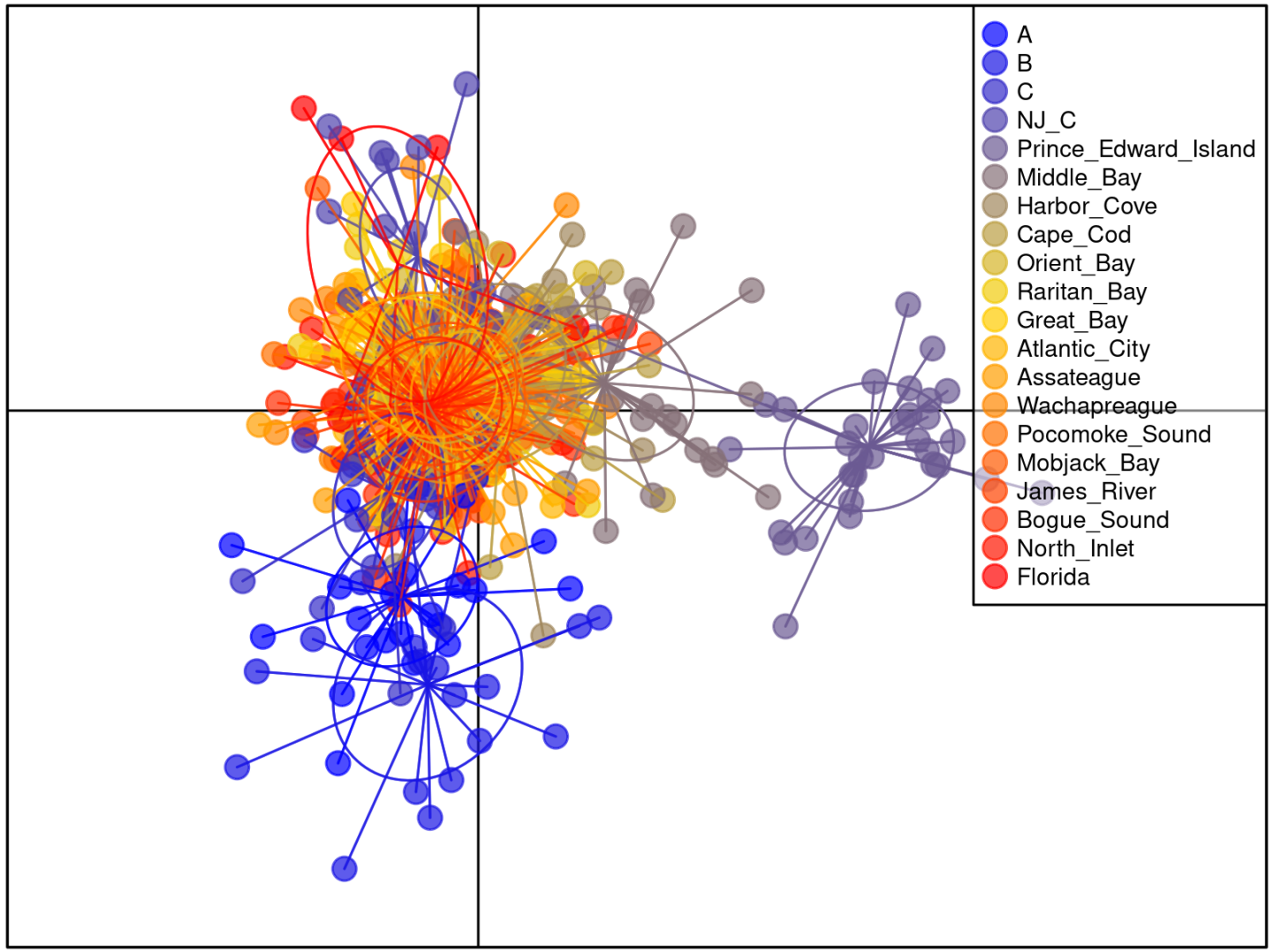
SNP Distribution



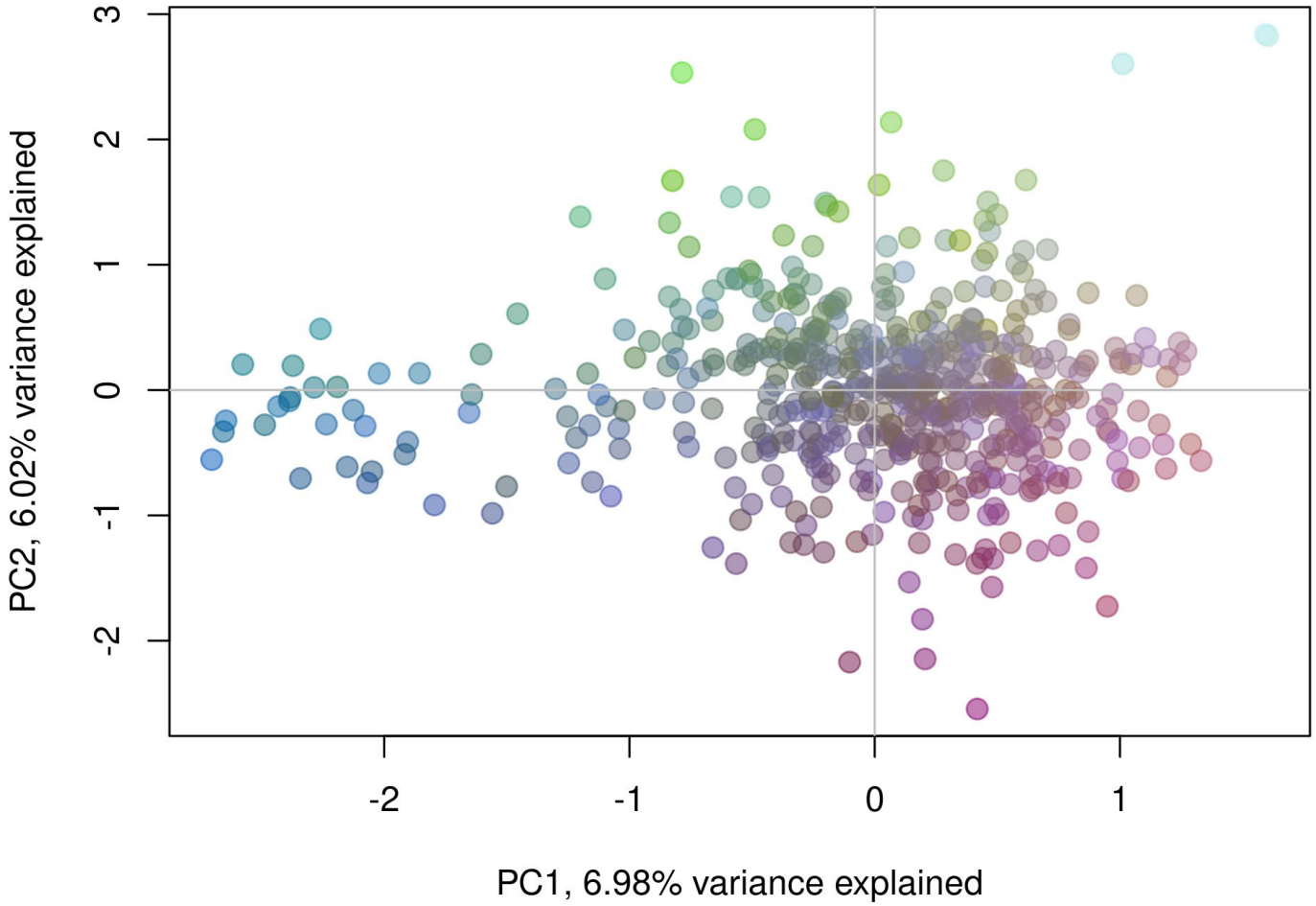
MAF distribution by population

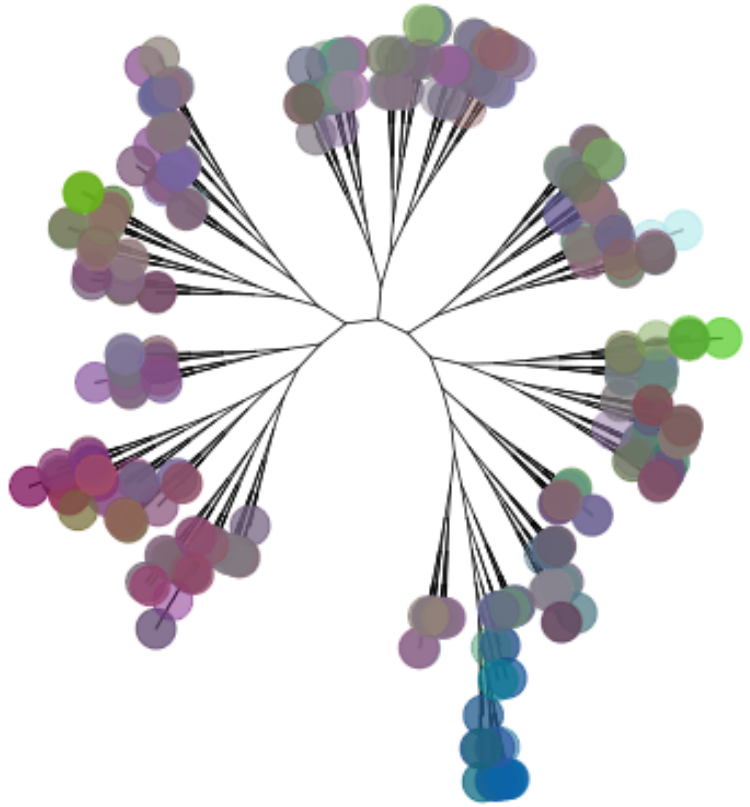


Population Structure (DAPC)

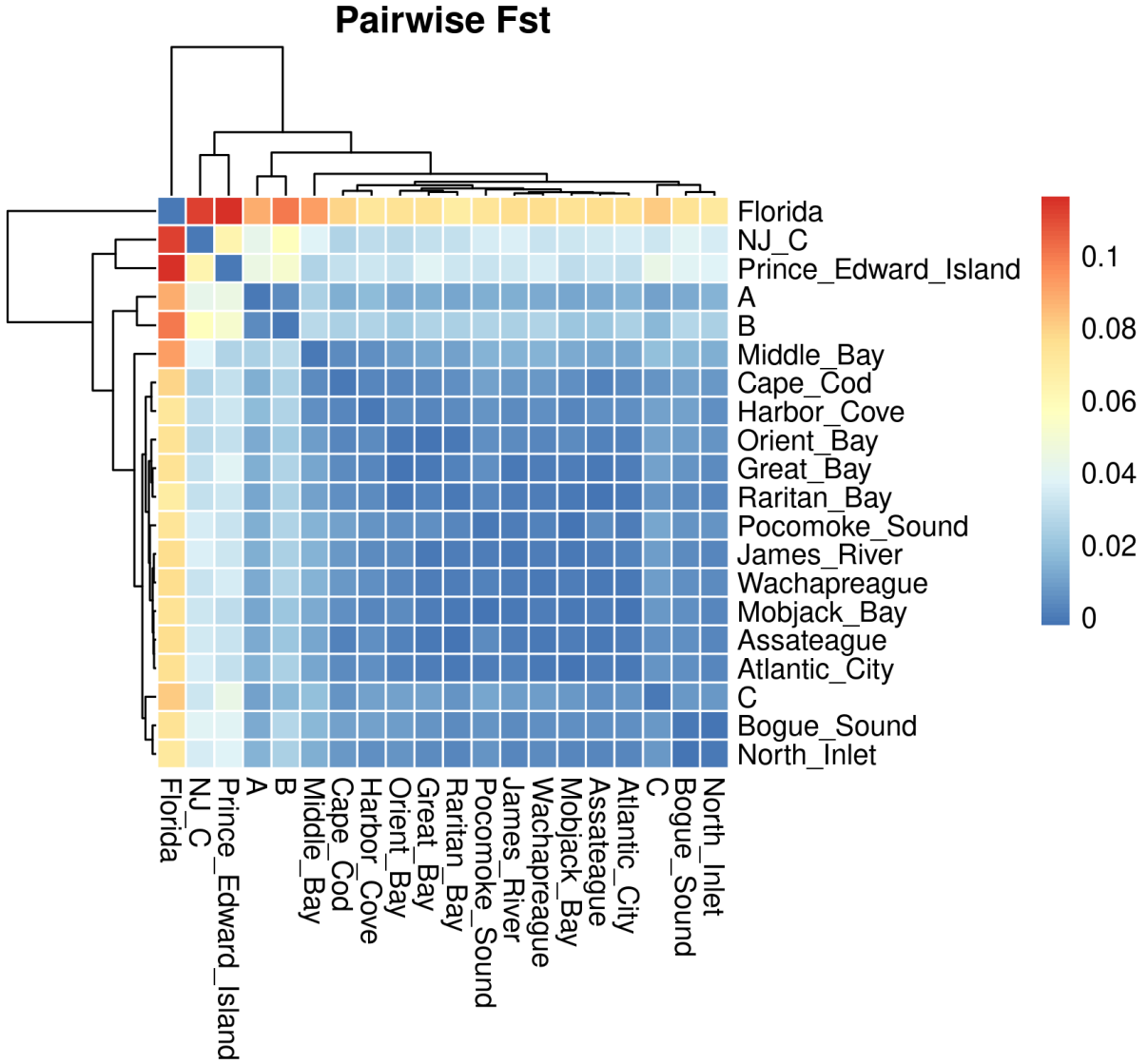


Population Structure (PCA and tree)





Population Differentiation(F_{ST})



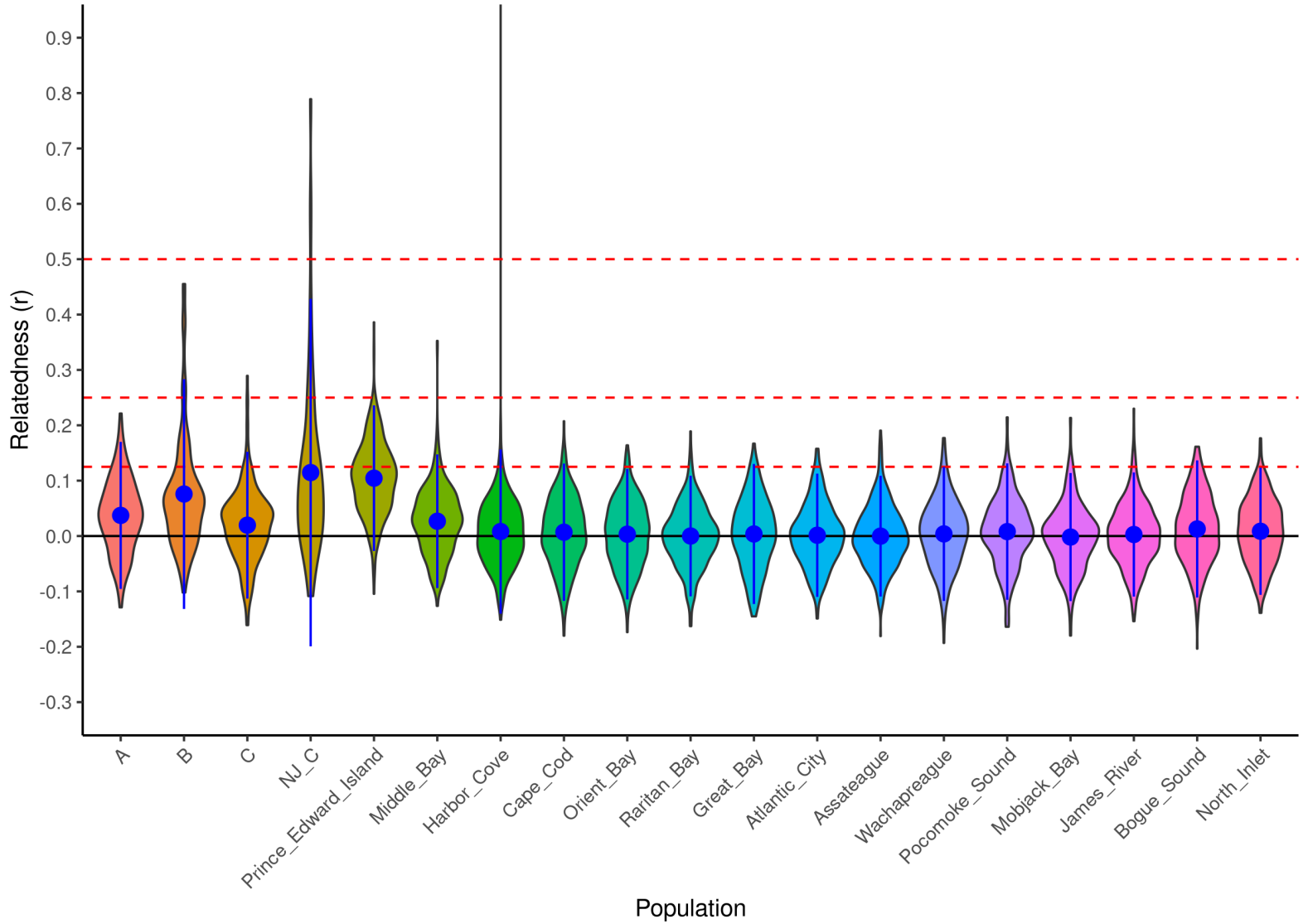
Population Diversity (selected 451 SNPs)

Population	*# of Loci Genotyped*	*# Polymorphic Loci*	*% Polymorphic Loci*	*Ave. # Alleles per Locus*	*Nei's Diversity Score*	*Hexp*	*Hobs*
A	451	450	99.77827	2.00	0.34	0.33	0.28
B	451	448	99.33481	1.99	0.34	0.33	0.27
C	451	451	100.00000	2.00	0.34	0.33	0.27
NJ_C	451	445	98.66962	1.99	0.33	0.32	0.24
Prince_Edward_Island	451	451	100.00000	2.00	0.33	0.32	0.27
Middle_Bay	451	451	100.00000	2.00	0.35	0.34	0.29
Harbor_Cove	451	451	100.00000	2.00	0.35	0.34	0.28
Cape_Cod	451	451	100.00000	2.00	0.35	0.34	0.27
Orient_Bay	451	451	100.00000	2.00	0.35	0.34	0.27
Raritan_Bay	451	451	100.00000	2.00	0.35	0.34	0.29
Great_Bay	451	451	100.00000	2.00	0.35	0.34	0.29
Atlantic_City	451	451	100.00000	2.00	0.34	0.34	0.29
Assateague	451	451	100.00000	2.00	0.35	0.34	0.29
Wachapreague	451	451	100.00000	2.00	0.34	0.33	0.27
Pocomoke_Sound	451	451	100.00000	2.00	0.34	0.33	0.28
Mobjack_Bay	451	451	100.00000	2.00	0.35	0.34	0.29
James_River	451	451	100.00000	2.00	0.35	0.34	0.28
Bogue_Sound	451	451	100.00000	2.00	0.34	0.34	0.27
North_Inlet	451	451	100.00000	2.00	0.34	0.34	0.28
Florida	451	327	72.50554	1.73	0.29	0.26	0.15

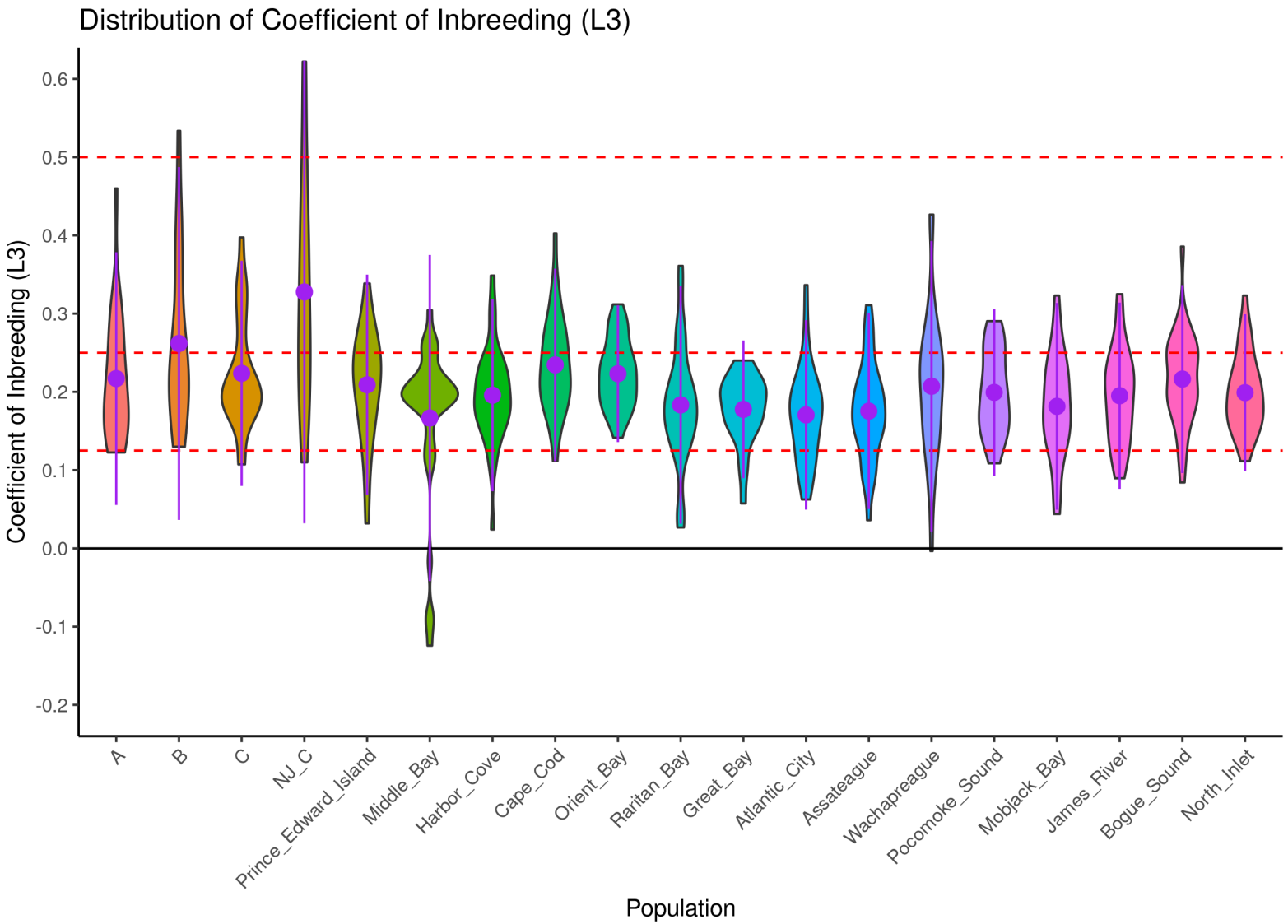
Population Diversity (filtered 2420 SNPs)

Population	*# of Loci Genotyped*	*# Polymorphic Loci*	*% Polymorphic Loci*	*Ave. # Alleles per Locus*	*Nei's Diversity Score*	*Hexp*	*Hobs*
A	2,420	1,994	82.39669	1.82	0.22	0.21	0.17
B	2,420	1,856	76.69421	1.77	0.21	0.21	0.17
C	2,420	2,038	84.21488	1.84	0.22	0.21	0.17
NJ_C	2,420	1,730	71.48760	1.71	0.21	0.20	0.15
Prince_Edward_Island	2,420	1,874	77.43802	1.77	0.21	0.20	0.17
Middle_Bay	2,420	2,127	87.89256	1.88	0.22	0.22	0.19
Harbor_Cove	2,420	2,195	90.70248	1.91	0.22	0.22	0.17
Cape_Cod	2,420	2,188	90.41322	1.90	0.22	0.22	0.17
Orient_Bay	2,420	2,200	90.90909	1.91	0.22	0.22	0.17
Raritan_Bay	2,420	2,166	89.50413	1.90	0.22	0.21	0.18
Great_Bay	2,420	2,085	86.15702	1.86	0.22	0.21	0.18
Atlantic_City	2,420	2,147	88.71901	1.89	0.22	0.21	0.18
Assateague	2,420	2,152	88.92562	1.89	0.22	0.22	0.18
Wachapreague	2,420	2,165	89.46281	1.89	0.21	0.21	0.17
Pocomoke_Sound	2,420	2,011	83.09917	1.83	0.22	0.21	0.17
Mobjack_Bay	2,420	2,145	88.63636	1.89	0.22	0.21	0.18
James_River	2,420	2,160	89.25620	1.89	0.22	0.21	0.18
Bogue_Sound	2,420	2,125	87.80992	1.88	0.21	0.21	0.17
North_Inlet	2,420	2,131	88.05785	1.88	0.21	0.21	0.17
Florida	2,420	1,290	53.30579	1.53	0.22	0.19	0.12

Relatedness (Ritland)



Inbreeding (Trioml L3 estimator)



Title: Genetic overview for “name of the client”

Date: 22-July-2020

Report: G##-20-##

Example for genetic overview using synthetic data (simulation based)

Client: Jane Doe

jdoe@yahoo.com

Test Services	Sample Type	Sample Number	Species
• *DNA Extraction, SNP Genotyping* • *Genetic overview analysis*	Pleopods	800	*Litopenaeus vannamei*

Service

The request from the client was to perform diversity analyses on samples from 4 lines; AA, BB, CC and DD of *L. vannamei* (n=800 total; 200 each) genotyped across an established panel of bi-allelic single nucleotide polymorphism (SNP) markers assembled for *L. vannamei*. Resulting SNP genotype data for samples were subjected to analyses to determine levels of diversity and inbreeding within lines and assess population differentiation between lines.

Executive Summary

Single nucleotide polymorphism (SNP) genotype data were generated for samples across 4 lines of shrimp; AA, BB, CC, and DD (200 samples each) and analyzed for diversity, inbreeding, and genetic structure. Analyses revealed relatively high levels of diversity across two of the lines; AA and BB, but eroded populations in the case of CC and DD especially the last one mentioned. Overall, these results indicate that the lines AA and BB have sufficient diversity to support a breeding program, although a different strategy must be taken for CC and DD, including combining all of them, infusing with new genes or discarding animals form there. The analyses did reveal that all lines AA and BB exhibit small levels of inbreeding, but these were considered statistically non-significant; while the levels are not deemed a significant concern for a breeding program, it is recommended that these lines be monitored at regular intervals to assess changes to these levels in subsequent generations. On the other hand, lines CC and DD displayed concerned levels of inbreeding. Population differentiation analyses (F_{ST}, PCA, and STRUCTURE) indicated low levels of differentiation between the two more diverse lines, AA and BB suggesting that can be combined. Even though CC and DD seem to be different among them, when doing a deeper analysis few individuals are contributing to small clusters or families. Diversity measures, along with population structure estimates, can be used moving forward to monitor status and guide decisions regarding the lines (to maximize diversity and manage genetic improvements), and assess genetic characteristics of new lines that may be acquired in the future.

Methods

Samples: A list of 800 samples was provided by the client for the project. Samples were indicated as belonging to 4 separate lines of shrimp (n= 200/line). For the sake of succinctness, client provided line names were abbreviated as follows: Line AA; 200 samples, 100 males and 100 females; Line BB; 200 samples, 100 males and 100 females; Line CC; 200 samples, 100 males and 100 females and Line DD; 200 samples, 100 males and 100 females.

Genotyping: Total gDNA was extracted from all samples using CAT established protocols. Resulting DNAs were screened for bi-allelic variation across CATs established panel of SNP loci for *L. vannamei*. After quality control, a total of 192 SNP loci were included in analyses. The mean per locus call rate was 98.2% across all genotyped samples. No samples failed genotyping QC (call rate < 80%) and the mean per sample call rate across all loci was 98.6%.

Diversity and Population Structure: Standard diversity measures were generated including allele frequencies, average number of alleles/polymorphic markers, expected and observed heterozygosities and Nei's genetic diversity. Relatedness values were calculated using the Lynch-Ritland moment estimator. The level of genetic differentiation was calculated using Weir and Cockerham's pairwise F_{ST} estimates between all lines. The significance for the pairwise F_{ST} estimates was tested using 5,000 permutations. Principal Component Analyses (PCA) were conducted to visualize patterns of genetic variation among lines and among all combined genotyped samples. The model-based Bayesian clustering program STRUCTURE was also used to examine population structure in each line and over all lines. Values of K = 1 to 8 (# of putative populations) with 5 repetitions of each K were tested with a burn-in rate of 50,000 followed by 50,000 MCMC iterations.

Results

Diversity: The diversity statistics generated indicate that 2 lines; AA and BB have a high level of within group genetic variability, while the others, CC and DD have very limited variability (Table 1).

Table 1. Number of loci typed, Nei's unbiased genetic diversity, mean number of alleles over all loci, percentage of polymorphic loci, and observed and expected heterozygosity for each line and all lines combined.

Group	Number of SNPs	Nei	Mean number of alleles	Ratio of polymorphic loci	He	Ho	ΔF
Overall	192	0.3233	1.97	96.9%	0.3233	0.2131	n/a
AA	192	0.2415	1.86	84.9%	0.242	0.2466	n/a
BB	192	0.2422	1.81	79.7%	0.242	0.2469	n/a
CC	192	0.2106	1.66	64.6%	0.211	0.2214	n/a
DD	192	0.1395	1.41	38.5%	0.139	0.1376	n/a

Most loci typed were polymorphic in AA and BB (79.7% - 84.9%), resulting in a mean number of alleles per locus in the range of 1.81 to 1.86 (approaching the maximum value of 2). Heterozygosities in these two lines were also high and translated into relatively high scores for Nei's genetic diversity index. However, all the genetic parameters are low for BB and CC, indicating a limited genetic diversity, particularly for DD with only 38.5 % of polymorphic alleles and 1.41 mean number of alleles. The genetic diversity scores ranged between 0.1395 and 0.2422 (for a bi-allelic set of SNPs, Nei's Diversity score has a maximum value of 0.5). The diversity scores for the diverse lines AA and BB, fall on the higher end of the range of observed values for other cultured L. vannamei populations where diversity scores range from 0.21 to 0.32 (average diversity of 0.28). Nei's diversity score provides an estimate of the extent of genetic variability in the population and more specifically provides the probability that, at a single locus, any two alleles chosen at random from the population will be different from each other. In highly related populations, this probability (or diversity score) is low due to increased levels of genetic similarity, which is the case for CC and DD. In contrast, the values obtained for the lines AA and BB screened in this study indicate that the probability of different alleles is high.

Inbreeding/pedigree: In this example, there is not data to calculate a per generation change in estimated inbreeding (ΔFIS, table 1), which gives an indication as to how fast deleterious levels of inbreeding may be accumulating in a population. That is an advantage of the GO+ package in which, lines with any levels of inbreeding should continue to be monitored at regular intervals to assess the rate at which these estimates are changing.

***Relatedness*:** We conducted further analyses to assess levels of relatedness among lines (Figure 1). A relatedness value of 0.5 is what is expected from individuals that are full-sibs or parent-offspring, while an r-value of 0.25 is indicative of half-siblings and an r-value of 0.125 is indicative of a first-cousin level of relatedness. Mean r values comparisons display a high level of relatedness between lines AA and BB as well as between CC and DD. However, the other pair-wise comparisons, shows negative values, indicating that there is not significant relatedness among these lines.

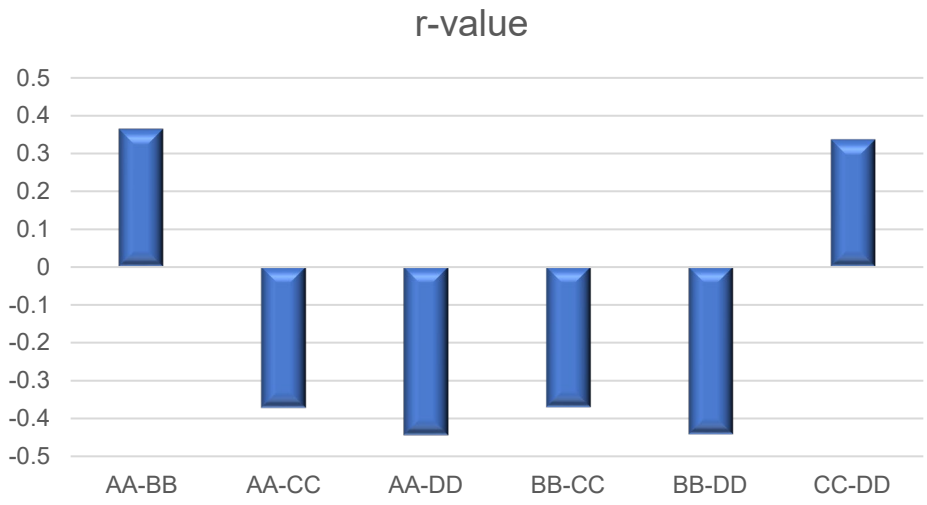


Figure 1. Relatedness estimates (r) among lines. Lines AA and BB as well as lines CC-DD are the most related comparisons. The other combinations are not.

Because a moment estimator was used for relatedness calculations, the mean r values consider positive and negative values. Thus, r mean values are useful for contrasts among lines as previously described, but a better representation of the pair-wise comparisons is the distribution of those values (Figures 2 and 3).

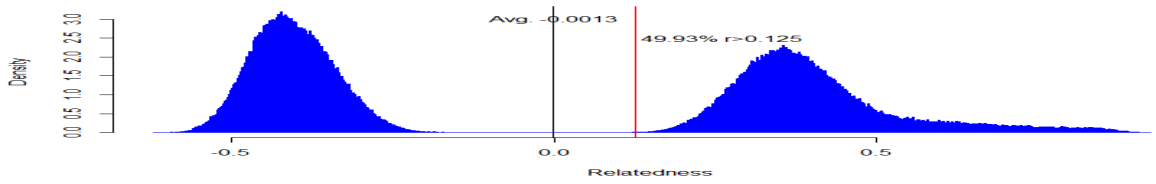


Figure 2. Relatedness (r) distribution of pairwise comparisons for the overall population containing the four lines together (n=800). A bimodal distribution showing half of the comparisons to the left below the lower level of relatedness (r=0.125, half cousins) and the other half to the right over this number. By contrast, the average r value for the overall population is r=-0.0013, which would indicate no relatedness. This emphasizes the usefulness of this type of representation.

The individual histograms of relatedness coefficients (r) for each line is shown in Figure 3. This analyses confirmed that lines CC and DD are more identical by descendant with a higher percentage of individuals over the lowest value for consanguinity (19.94 % and 26.41 % respectively) compared with the low-related and more diverse AA and BB (5.99 % and 7.43 % respectively).

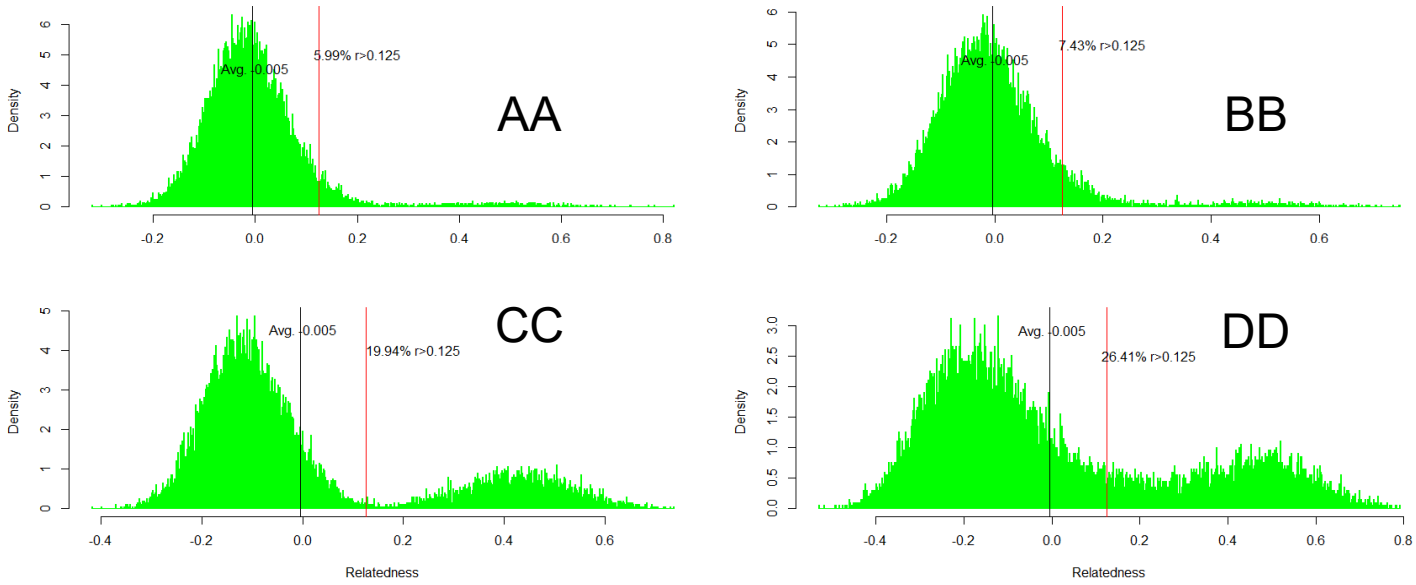
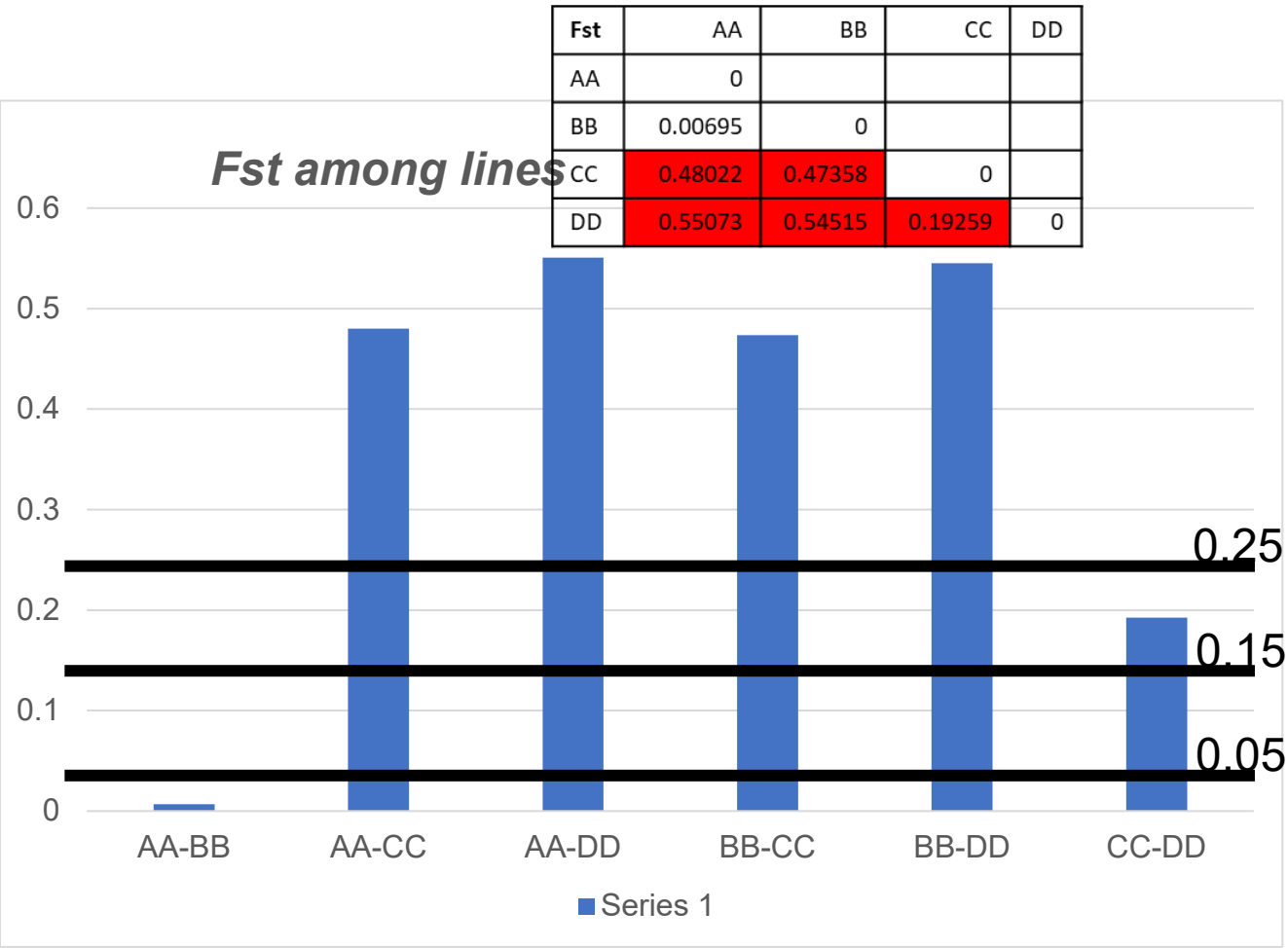


Figure 3. Relatedness (r) histograms of pairwise comparisons for each individual line; AA, BB, CC and DD. Normal distributions with low percentage of comparisons over the lower level of relatedness (r=0.125, half cousins, red line) are shown for lines AA and BB. Bimodal distributions displaying both a portion of non-related comparisons and more than ≈ 20 % over the concerning value are shown for lines CC and DD.

***Structure of the population**:* Analysis of F_{ST} values revealed great genetic differentiation among the lines, except between AA and BB. Pair-wise F_{ST} values are presented in Figure 4 and these values ranged from a low of 0.00695 to a very high of 0.55073. The lowest F_{ST} estimates indicate that lines AA and BB were the most genetically similar to one another. In contrast, the rest of the comparisons are very different among them.



Fst	AA	BB	CC	DD
AA	0			
BB	0.00695	0		
CC	0.48022	0.47358	0	
DD	0.55073	0.54515	0.19259	0

Figure 4. Pairwise estimates of F_{ST}, an estimate of population structure, for each line. Values in the top table and bars in the main graph. Black lines indicate the limit for levels of genetic differentiation as following: ≤ 0.05, little; between 0.05 – 0.15, moderate; between 0.15 – 0.25, great and over 0.25, very great. Genetic differentiation among all lines was very high and measured more than than 0.25 for all comparisons, except for AA-BB. The lowest F_{ST} estimates showed that lines AA and BB were the most genetically similar to one another; while the rest of the values over 0.19 indicates great – very great differentiation among the other lines.

Principal Components Analysis (PCA) were performed to provide visual representations of genetic differentiation among all genotyped individuals from the 4 lines (Figure 5). As indicated by the PCA plots, there are three (3) different clusters and lines AA and BB, although very diverse are grouping in the same cluster. This points out the low level of differentiation between these two lines and in practical terms, it can be merge into one diverse population. These results agree with the F_{ST} estimates previously shown and with the following STRUCTURE plots.

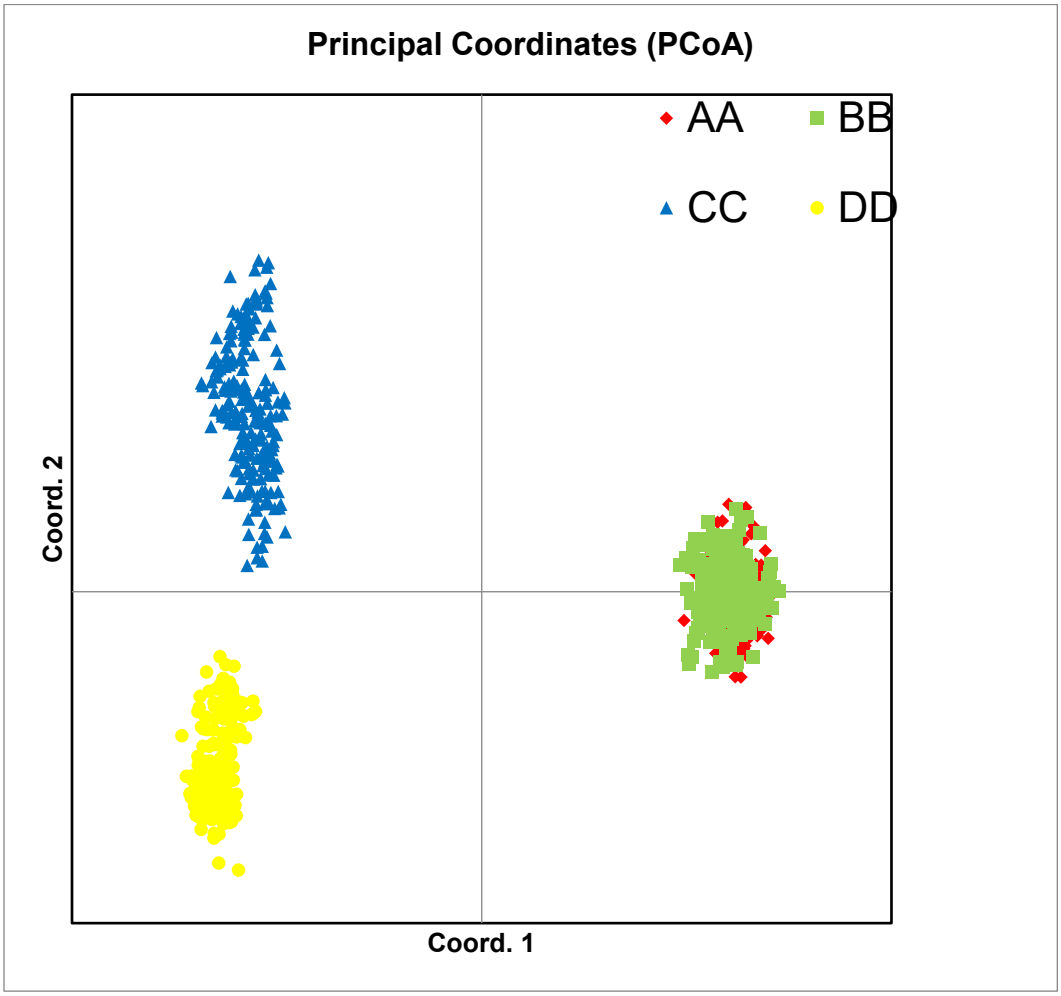


Figure 5. Results of the Principal Components Analysis (PCA) showing three different clusters, indicating high differentiation among line, except for AA and BB that are quite similar and are group in one cluster.

To further explore the structure within each line, the same analysis was performed using only the samples of each line separately. Interestingly as shown in Figure 6, the individuals in lines AA and BB are largely overlapped, but few micro-clusters appear for lines CC and DD. This might indicate that only few families are contributing in lines CC an DD (only 4 clusters), which explain the lack of diversity for most of the genetic parameters previously mentioned.

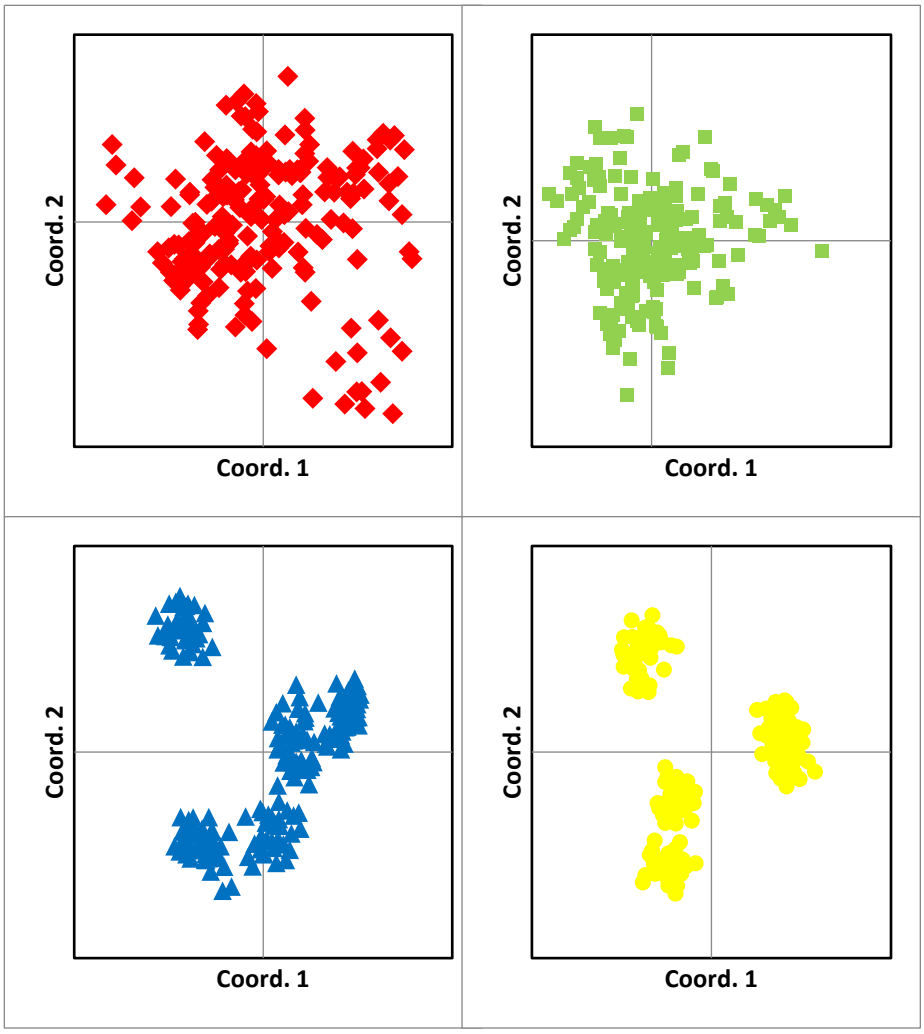


Figure 6. Results of the Principal Components Analysis (PCA) withing the four lines; AA in red; BB in green; CC in blue and DD in yellow. Separation in clusters of similitude is shown in lines CC and DD.

STRUCTURE plots were generated and use different colors to represent population substructure generated assuming K = 2 – 8 populations (Figure 7). Largely, the STRUCTURE results support the findings of the other population analyses conducted in this study. Analyses indicated K= 2 or K=4 populations as a best statistical fit with the data.

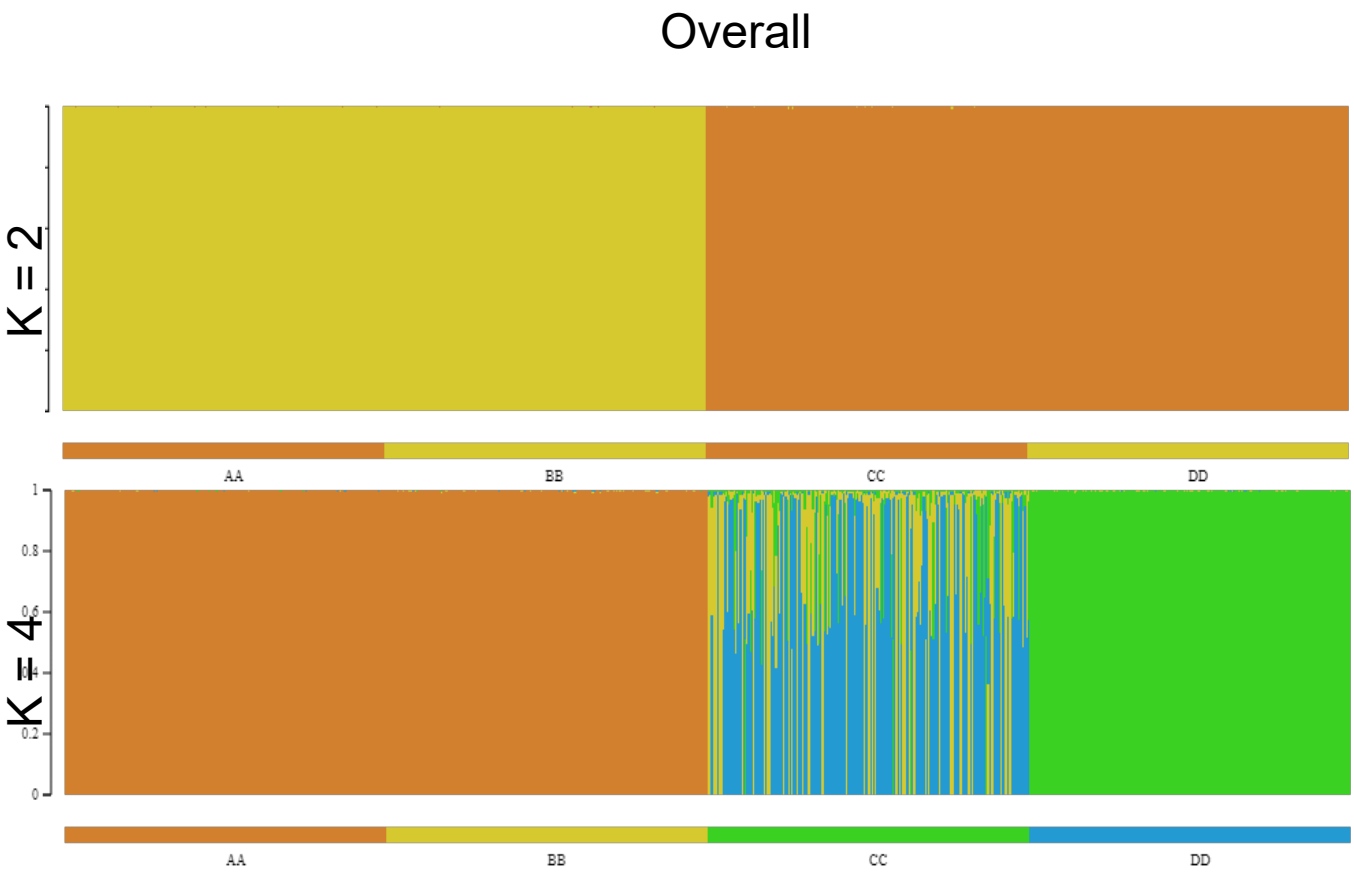


Figure 7. STRUCTURE barplot representing overall population structure K = 2 and K=4 populations. The likely genetic makeup of each individual shrimp is represented by a single vertical bar. Analyses indicated K= 2 and/or K=4 populations as a best statistical fit with the data.

Conclusions

- Overall diversity measures indicate that lines AA and BB contain high levels of genetic diversity and appear to be healthy from the breeding perspective. It is the contrary for the lines CC and DD that are very eroded and need to be carefully managed in a breeding strategy.
- The distribution of relatedness coefficients (r) show that lines CC and DD have many related individuals within each one, while in lines AA and BB this proportion is small and typical in culture systems. As with all breeding programs, it is recommended that lines be monitored at regular intervals to assess the rate at which inbreeding, and relatedness is changing.
- FST, PCA and STRUCTURE plots analyses indicate differentiation among all lines, except between AA and BB that can be merge in the farming practice. The sub-structure in small clusters withing lines CC and DD offers a plausible explanation for the limited diversity and high relatedness in these mentioned lines.
- Diversity measures, along with population structure results, can be used moving forward to monitor status and guide decisions regarding the lines (to maximize diversity and manage genetic improvements), and assess genetic characteristics of new lines that may be acquired in the future.