Proc. Assoc. Advmt. Anim. Breed. Genet. 22:445-448

RECONSTRUCTION OF UNKNOWN DONOR GENOME FROM CHIMERIC PEARL SAC TISSUE AND HOST GENOTYPES IN *PINCTADA MAXIMA*

M.S. Khatkar[1,2], D.B. Jones[3], P.B. Toole[3,4], H.W. Raadsma[1,2], D.R. Jerry[2,3] and K.R. Zenger[2,3]

[1] Sydney School of Veterinary Science, Faculty of Science, The University of Sydney, Camden, NSW, Australia
[2]ARC Hub for Advanced Prawn Breeding, James Cook University, Townsville QLD, Australia
[3] Centre for Sustainable Tropical Fisheries and Aquaculture, College of Science and Engineering, James Cook University, Queensland, Australia
[4] Atlas South Sea Pearls Ltd., Western Australia, Australia

SUMMARY

Commercial pearl production involves two oysters, the seeded individual (host), and a sacrificed oyster (donor) from which a piece of tissue is utilised during the seeding process. During commercial seeding, it is often difficult to keep track of individual pairing information of host and donor oysters. Here we describe a method for reconstructing donor genotypes from host genotypes and allele frequencies generated from chimeric pearl sac tissue (a mixture of host and donor tissues) at pearl harvest. Using simulation of expected genotype frequencies and genotypes from true samples, we demonstrate that donor genotypes can be constructed with high accuracy. Best results were observed when the error rate of allele frequencies is low, SNPs have a high minor allelic frequency, and when the proportion of the donor tissue in the pearl sac tissue is greater than 20 percent.

INTRODUCTION

Commercial pearl production involves two oysters, the seeded individual (host), and a sacrificed oyster (donor) which produces a piece of tissue (also termed the 'saibo' tissue). This saibo tissue is implanted into the host during the seeding process (Figure 1). Both donor and host oyster tissues are known to be actively expressed during the production of a pearl (Arnaud-Haond et al., 2007), and the quality of pearls produced are influenced by the genomes of both the host and donor oysters (Jerry *et al.*, 2012), (Tayale *et al.*, 2012). Hence, for selective breeding and dissecting the contribution of host and donor oysters, and their interaction on various pearl quality traits, the identification and recording of both animals is critical. In many commercial pearling farms, the tracking of donor oysters is not maintained routinely due to additional management complexity and expense (Jerry *et al.*, 2012) (Jones *et al.*, 2014). However, the identification and validation of donor oysters during pearl harvest at the end of a four year pearl production cycle is integral for conducting genetic studies and making breeding decisions. Under the assumption of the donor genome being present in the pearl sac tissue at harvest, it should be possible to differentiate reciprocal host and donor genomes (and reconstruct genotypes) by using either the known host or donor genotype and pearl-sac allele frequencies.

Here we present a method for reconstructing donor genotypes from host genotypes and allele frequencies from pearl sac tissue (a mixture of host and donor tissues). Using simulations, we demonstrate the effect of genotyping error and variable proportions of host and donor tissue observed within pearl sac samples on the accuracy of reconstructing unknown donor genotypes.

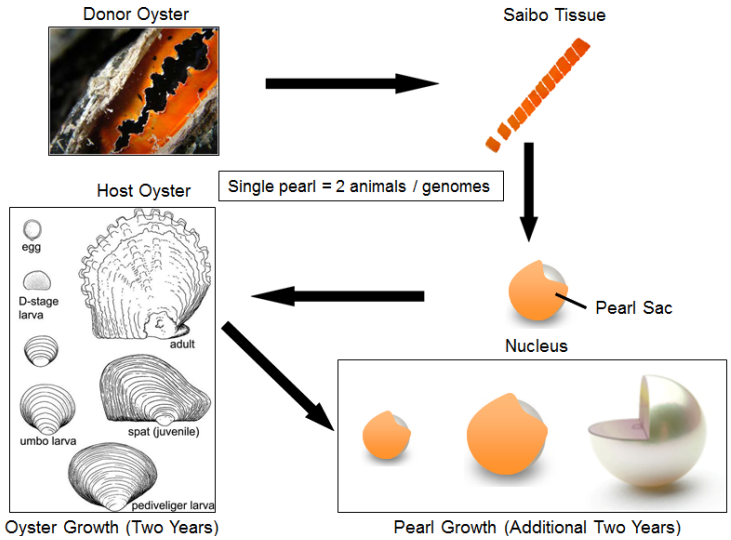

Figure 1. Commercial pearl production via seeding

MATERIAL AND METHODS

Method of reconstruction of donor genotypes from allele frequency of pearl sac tissue and genotype of host. Let **H** be a vector of genotypes of n SNPs from a host coded as 0 for one homozygotes (AA), 1 for heterozygotes (AB), and 2 for other homozygotes (BB), and **P** a vector between 0 and 1 representing allele frequency (frequency of B allele) of n SNPs of pearl sac tissue. Then reconstruction of the donor genotypes from the allele frequency of the pearl sac tissue and genotypes of the host was done in three steps. First the proportion of host tissue in the pearl sac tissue was estimated as the regression coefficient of host genotypes on allele frequencies of pearl sac tissue from a linear regression model i.e. $2\mathbf{P}=b\mathbf{H}+e$, then allele frequencies of the donor was estimated by $\mathbf{D} = (2\mathbf{P}-b\mathbf{H})/(1-b)$. Finally, the donor genotypes were reconstructed by classifying the frequency estimates **D** of the donor into the nearest genotypic class 0 (AA), 1 (BB) and 2 (BB). The classification error of donor genotypes was calculated from a confusion matrix between the estimated genotypes and the actual genotypes of the donor. A total 1,000 randomly selected pairs of animals were evaluated for each scenario.

Simulation. The working and utility of the above procedure was tested by sampling and simulation. The genotypes of a host genome were generated on 935 SNPs by randomly sampling the genotypes of one animal from a real genotypic dataset on pearl oyster (*Pinctada maxima*) recently generated by genotyping 329 animals with a high-density DArTseq Diversity Arrays SNP panel (Kilian *et al.*, 2012). Similarly genotypes of one donor were generated by sampling another animal from the same dataset. These 935 SNPs had a minimum of 0.3 minor allelic frequency (MAF) in the panel of genotyped animals. Lower thresholds for MAF were also explored. The pearl sac tissue was assumed as a mixture of host and donor tissues in various proportion. The allele frequency for the pearl sac tissue was generated by mixing the donor and the host genotypes in ten different proportions by assuming 5, 15, 25, 35, 45, 55, 65, 75, 85, 95% of host tissue and the remaining respective proportion of the donor tissue. In addition, to accommodate some laboratory error in estimating the allele frequencies, a continuous uniform distribution for error rate in an interval of -0.06 and 0.06 was

generated and added to the allele frequencies computed above (while keeping the resultant allele frequencies within a bound of 0 and 1). Other uniform distributions with higher error rates, and normal distributions *N(0, σ2)* with four different error distributions (σ = 0.01, 0.02, 0.05, 0.10) were also evaluated.

RESULTS AND DISCUSSION

The accuracy of reconstructed donor genotypes from pearl sac tissue consisting of different proportion of host and donor tissue, each computed from 1,000 host and donor pairs, are presented in Table 1. These results are based on a uniform error distribution with an interval of -0.06 and 0.06. (Figure 1). The estimated proportion of host genome was close to the actual proportion for most of the scenarios except the scenario 1 and 2 where proportion of the host tissue was small. Nevertheless, the median mis-classification rate for donor genotypes was zero except for the last two scenarios where the proportion of host tissue in the pearl sac was very high (> 85 %). Other laboratory error rates in the estimate of allele frequency and using SNPs with lower MAF were also explored. However, the higher error rates in the estimation of allele frequency resulted in higher misclassification rate of donor genotypes. Similarly using SNPs with lower MAF thresholds also resulted in higher misclassification rate (results not shown).

Table 1. Accuracy of reconstructed donor genotypes from Pearl sac consisted of different proportion of host and donor tissue

Proportion of host in pearl sac tissue	Median estimated host proportion	QR estimated host proportion	Median correlation between estimated and actual genotype	QR correlation between estimated and actual genotype	Median mis-classification rate (%)	QR mis-classification rate (%)
0.05	0.10	0.02-0.20	1	0.98-1	0	0-0
0.15	0.19	0.12-0.29	1	0.98-1	0	0-0
0.25	0.29	0.22-0.38	1	0.98-1	0	0-0
0.35	0.38	0.33-0.46	1	0.98-1	0	0-0
0.45	0.48	0.43-0.54	1	0.98-1	0	0-0
0.55	0.57	0.53-0.62	0.99	0.98-1	0	0-0
0.65	0.66	0.63-0.70	0.99	0.98-0.99	0	0-0
0.75	0.76	0.74-0.79	0.98	0.97-0.99	0	0-0
0.85	0.85	0.84-0.87	0.96	0.95-0.96	0	0-2.8
0.95	0.95	0.94-0.95	0.74	0.71-0.75	40.3	37.9-43.0

QR is 0.1 and 0.9 quantile based on 1,000 host and donor pairs

The error rates in constructing the donor genotypes of SNPs with low MAF were higher, and hence we recommend the use of high MAF SNPs. Pedigree information can also be used to correct some of the incorrectly constructed genotypes and genotypes of low MAF SNPs of the donors. These corrected genotypes can then finally be used for GWAS and genomic selection. Genetic relationship between donor and host tissue, especially close relationship such as full-sibs, may also affect the accuracy of reconstruction of donor genotypes and warrants further investigations.

Overall these results suggested that donor genotypes can be constructed with a very high certainty if the error rate in the estimation of allele frequency is low, high MAF SNPs are used, and the proportion of the host tissue in the pearl sac tissue is not very high.

To validate this method further we are now preparing synthetic pools by mixing DNA of pairs of animals in different proportions. In addition we will sample a number of trio samples consisting of host, donor and pearl sac tissues. The procedure of taking samples of pearl sac tissue may also affect the proportion of host and donor genome; a few different procedures of sampling such as slicing the interior pearl sac tissue with a scalpel and using a sterile dental swap to collect cells without an incision on the pearl sac will be evaluated to minimise the proportion of host tissue in pearl sac tissue. These samples will be genotyped with low-density DArTseq Diversity Arrays SNP panel (Kilian *et al.*, 2012). The analysis of these samples will provide estimate of donor tissue in the pearl sac, estimation of laboratory error rate in the allelic frequency of pearl sac, and finally ability to reconstruct donor genotypes.

CONCLUSION

The method presented here provides a way to reconstruct the genotypes of the donor from allelic frequency data on pearl sac tissue and genotypes of the host. The results also suggested that some error in the estimation of allele frequency can be tolerated. However, when a very high proportion of the host tissue is present in the pearl sac tissue, it difficult to reconstruct donor genomes and introduced high error in the estimated genotypes of donor.

REFRENCES

Arnaud-Haond, S., Goyard, E., Vonau, V., Herbaut, C., Prou, J., & Saulnier, D. (2007). Mar Biotechnol (NY), 9(1), 113-116.

Jerry, D. R., Kvingedal, R., Lind, C. E., Evans, B. S., Taylor, J. J. U., & Safari, A. E. (2012). Aquaculture, 338–341, 66-71.

Jones, D. B., Jerry, D. R., Khatkar, M. S., Moser, G., Raadsma, H. W., Taylor, J. J., & Zenger, K. R. (2014). Aquaculture, 434, 476-485.

Kilian, A., P. Wenzl, E. Huttner, J. Carling, L. Xia, H. Blois, V. Caig, K. Heller-Uszynska, D. Jaccoud, C. Hopper, M. Aschenbrenner-Kilian, M. Evers, K. Peng, C. Cayla, P. Hok and G. Uszynski (2012).. Methods Mol Biol, 888, 67-89.

Tayale, A., Gueguen, Y., Treguier, C., Le Grand, J., Cochennec-Laureau, N., Montagnani, C., & Ky, C.-L. (2012). Aquatic Living Resources, 25(3), 269-280.