Proceedings, 10th World Congress of Genetics Applied to Livestock Production

RAD Sequencing reveals genome-wide heterozygote deficiency in pair crosses of the Chilean mussel *Mytilus* spp.

C. Peñaloza[1], S. C. Bishop[1], J. Toro[2] and R. D. Houston[1]
[1]University of Edinburgh, Edinburgh, UK, [2]Austral University, Valdivia, Chile

ABSTRACT: Mussels are increasingly important aquaculture species. Selective breeding utilizing genomic resources is at a formative stage. We develop genome-wide markers and investigate their inheritance in two Chilean mussel families using a restriction-site associated DNA (RAD) sequencing approach. A total of 4,537 SNP markers were discovered. RAD sequencing analysis indicated that mussels are among the most polymorphic species studied to date, with an average frequency of 1 SNP each 25bp. Family-based sequencing allowed us to examine the inheritance pattern of markers. The majority of SNPs (72%) deviate from Mendelian segregation ratios, showing a trend towards a deficiency of heterozygote genotypes in the offspring. Additionally, many alleles were observed in the mussel offspring without being present in their parents. Our results indicate that the inheritance of markers has yet to be fully understood before they can be applied for the genetic improvement of mussels.
Keywords: Chilean mussel; Heterozygote deficiency; RAD-Sequencing

Introduction

Mussels are bivalve shellfish which represent an increasing segment of the global aquaculture industry (FAO (2010)). A limiting factor for mussel farming is the availability of spat. Spat are young mussels, which are traditionally collected from the wild to sustain mussel production. Abundance of spat varies among years, mainly due to variable environmental factors. This variability in spat supply leads to variability in production yields. To improve the stability and sustainability of mussel farming, hatchery production of mussel spat is increasingly becoming available. Hatcheries not only benefit mussel aquaculture by providing a more reliable source of juveniles, they also allow the development of genetic improvement through selective breeding programs.

Selective breeding including the use of genomic resources is at a formative stage. To support these programs, genome-wide genetic markers are required. Although the development of such markers is currently facilitated by next generation sequencing technologies, mussels exhibit certain genetic properties that could potentially limit the development and application of genomic technologies. Genetic markers usually follow Mendel's laws of segregation and independent assortment. However, the properties of genetic markers in mussels are not well understood and may contravene these expectations.

Over the past 40 years of mussel genetic research, numerous studies have revealed a common trend: the number of individuals that are heterozygous for a locus are significantly lower than expected under Hardy-Weinberg equilibrium (in population studies) or Mendelian segregation ratios (in family-based studies). This phenomenon is known as heterozygote deficiency, and has also been widely observed in other bivalve species such as clams, scallops and oysters (revised by Plutchak et al. (2006)). Several hypotheses to explain heterozygote deficiency in mussels (and bivalves in general) have been proposed, including inbreeding, technical artefacts (e.g. null-alleles), aneuploidy, parental imprinting, Wahlund effects and natural selection acting directly or indirectly on the marker loci (Zouros and Foltz (1984); Mallet et al. (1985)). Yet no clear consensus has emerged. As heterozygote deficiency in bivalve species has hindered to some extent the development of linkage maps and paternity analyses, further understanding of its origin is required before molecular tools are to be applied in bivalve breeding.

To date, heterozygote deficiencies in mussels have been reported using a limited number of allozyme or microsatellite markers. To better characterize these unusual segregation distortions, and therefore potentially provide an insight into their origin, this phenomenon must be explored on a genome-wide scale. In this respect, valuable information may be obtained by genomic surveys of genetic variants and their transmission. The aims of this study were to apply restriction site associated DNA (RAD) sequencing a powerful method for SNP discovery and genotyping, to create new genomic resources (high density SNP markers) for the Chilean mussel, and to examine genome-wide patterns of familial inheritance.

Materials and Methods

Mussel families. In 2008, approximately one hundred fifty full-sibling families were created by mating parents collected at random from three wild populations from the Chilean coast: Chaihuín (CH) (39°56' S; 73°35' W), Yaldad (YA) (43º 08' S; 73º 44' W) and Punta Arenas (53°08' S; 70°53' W). Standard methods for artificial fertilization and rearing were followed. From this set of families, two were chosen for library preparation, family 1 (from CH) and family 2 (from YA), and four were sampled to extend the tests of Mendelian inheritance.

RAD-tag library preparation. Genomic DNA was purified from a total of 96 individuals (The two parents and 46 offspring from each of the full-sibling families) using a modified CTAB extraction method. DNA was quanti-

fied using a Picogreen assay.

RAD libraries were prepared following protocols described elsewhere (Baird et al. (2008)). Briefly, the genomic DNA from each of the 96 individuals was digested separately with a SbfI restriction enzyme. Each individual sample was barcoded with a unique combination of 16 P1 and 6 P2 adapter sequences. All samples were multiplexed and sequenced in a single lane of an Illumina HiSeq 2000 instrument.

SNP Discovery and Genotyping. Raw Illumina reads were demultiplexed according to the barcode combination assigned to each individual during the RAD library preparation. Low quality reads (Q<20) and those that lacked a correct barcode were removed from further analyses. Within each mussel family, reads from all individuals were aligned against each other, allowing a maximum number of two mismatches between reads and requiring a minimum depth of two reads. A consensus sequence (RAD-tag) was defined for each cluster of reads. Within mussel families, SNP genotypes were determined by aligning all sequence reads for each individual mussel against the assembled RAD-tags; SNPs were called in each family only if one of both parents were heterozygous for the polymorphism.

Data Analysis. An average SNP frequency (SNPs per bp) was estimated for each assembled RAD-tag. In addition, the offspring genotypes at each SNP locus were tested against expected Mendelian segregation ratios using a chi-square goodness of fit test.

Assembly validation. The accuracy of SNPs and genotypes determined by RAD-Sequencing were validated by Sanger technology. A subset of three RAD-tags (~300 bp long each) were chosen for the design of primer pairs *Myt-12396*, *Myt-2828* and *Myt-9369*. After the standardization of PCR primers, they were used to amplify the DNA of 5 individual mussels. PCR products were purified and sequenced with reverse and forward primers to check for consistency. The concordance of results between the two sequencing methodologies was examined.

Extended family testing for Mendelian inheritance. The analysis of Mendelian inheritance was extended to four additional mussel families. The genomic DNA of each additional parent pair and 7 of their respective offspring were PCR amplified with primer pairs *Myt-12396*, *Myt-2828* and *Myt-9369*. Each PCR product was purified and Sanger sequenced with both forward and reverse primers. The generated Sanger reads were aligned within each mussel family for variant identification and evaluation of transmission patterns.

Results and Discussion

The RAD-Sequencing of both mussel families generated a total of 81 million valid reads. The RAD data was assembled into 4,113 RAD-tags (i.e. a consensus sequence upstream or downstream a SbfI restriction site), which were screened for SNPs. In family 1, 4,537 SNPs were discovered and genotyped. However, the dam from the family 2 had significantly fewer reads than average, causing a very high level of missing genotype data and, consequently, biasing SNP calling within the family. Hence, family 2 was retained for SNP frequency estimations, but excluded from Mendelian analysis. The correspondence between the RAD-tags and the Sanger sequences were high. This indicates that the results obtained using the Illumina HiSeq 2000 platform and the RAD SNP calling protocol are reliable. Hence, the assembled sequences reveal the true genetic variation contained within the mussel genome.

Three main observations derived from the characterization of the RAD sequences and the analysis of inheritance patterns in mussel families. First, the Chilean mussel shows exceptionally high levels of genetic polymorphism. A histogram illustrating the distribution of SNP frequencies in RAD-tags reveals a positively skewed distribution, with a mode of 1 SNP per 25 bp (Fig. 1). Typically genomic coding regions show lower SNP frequencies than noncoding regions (Strachan and Read 1999). However, we compared SNP frequencies between putative coding and noncoding RAD-tags, and both sequence types showed equally high frequencies (data not shown), suggesting that the genomic content of mussels, as a whole, is highly polymorphic. Similar frequencies have been observed in other bivalve species such as the Pacific oyster (1 SNP per 40 bp) and the Baltic clam (1 SNP per 19 bp) (Pante et al. (2012); Sauvage et al. (2007)). The remarkably high level of polymorphism found in our study provides evidence supporting bivalves as one of the most genetically diverse animal groups studied to date.

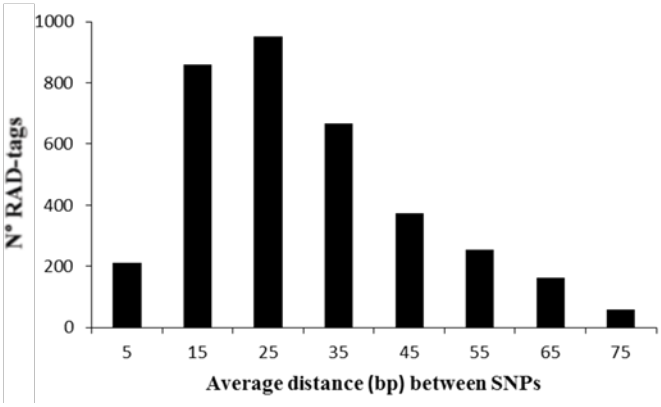

Figure 1. Distribution of SNP frequencies in the RAD-tags.

Second, most of the discovered SNP markers deviate from Mendel's Law of Segregation. In family 1 we evaluated whether the distribution of genotypes in the progeny followed Mendelian expectations. The chi-square analyses indicated that 3,280 markers (72%) were distorted. The majority of these distortions were caused by a deficiency of heterozygotes in the progeny (Fig. 2). Although heterozygote deficiencies at marker loci have been reported previously, our results additionally reveal that it is a widespread

genomic phenomenon, which may potentially be of biological significance. The fact that we used an integrated platform for SNP discovery and genotyping (i.e. a genotyping–by-sequencing approach) allowed us to directly identify polymorphisms from the mussels' RAD sequences, and thus helped reduce technical artefacts – one of the most accepted hypothesis for heterozygote deficiencies – as important sources of segregation distortions. The genome-wide segregation analysis in the Chilean mussel indicates that despite significantly reducing these sources, the phenomenon still remains. This evidence suggests that heterozygote deficiencies in mussels might be of biological origin, thus endogenous sources of deviations (e.g. novel inheritance mechanisms) should be further explored.

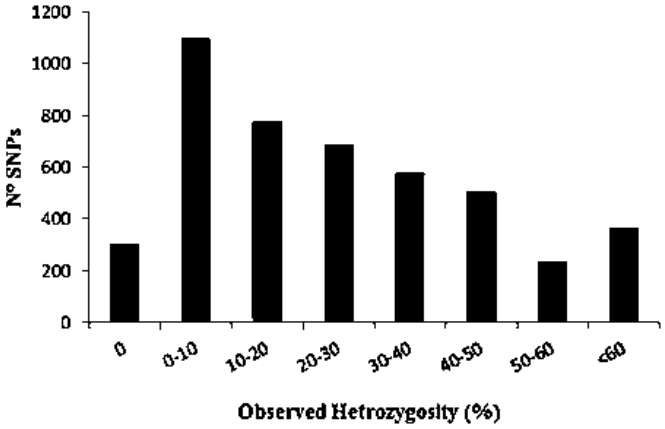

Figure 2. Histogram indicating SNP heterozygosity in the mussel offspring. The expected heterozygosity in the offspring is 50 % for all SNPs.

Third, several alleles were identified in the offspring that were not present in their parents. The results from the inheritance analysis performed in four additional families confirm the observation that mussel offspring exhibit multiple Mendelian inconsistencies. One possible explanation is that some degree of cross-contamination of individuals among families occurred. However, several security measures were taken to ensure family integrity throughout the experiment. Therefore, the likelihood of a mixture is believed to be minimal. Another explanation for the presence of these unexpected alleles is *de novo* mutations. Mutations can be caused by errors in DNA replication during cell proliferation. Since mussels fertilize externally by releasing thousands ($\sim 10^8$) of each male and female gametes to the water column, a high number of *de novo* mutations are expected to arise due to the high fecundity of the species. Interestingly, many previous family studies (e.g., Macavoy et al. (2008) and Del Rio-Portilla and Beaumont (2000)) have also reported offspring harbouring unexpected alleles/genotypes. Nevertheless in bivalve genetic research these individuals have been traditionally treated as contaminants and therefore discarded from further analysis. The surprisingly high level of unexpected alleles in the Chilean mussel F1 generation raises the fundamental questions of how and where in the genome they originate. If these non-inherited alleles are appearing in a single generation in relevant regulatory or coding DNA regions, then the overall expression of genes (or gene variants) may be compromised, which in turn may have a significant influence on the phenotype of individuals. This can potentially impact the effectiveness of selective breeding programs, particularly if molecular markers are to be applied for genetic improvement.

Experiments are underway to examine marker inheritance in additional mussel crosses and to investigate possible causes of these highly unusual segregation patterns.

Conclusions

The high density SNP markers developed in this study greatly advance genomic resources for a mussel species, which could potentially be of use for linkage mapping and parentage analysis, among others. However, the significant number of markers showing segregation ratio distortions and the frequent presence of non-parental alleles in offspring demonstrates that the inheritance of genetic markers has yet to be fully understood. This is an issue to be addressed before they can be applied to the genetic improvement of bivalve species.

Acknowledgements

The authors acknowledge staff at Edinburgh Genomic facility for assistance with RAD and ABI sequencing. The authors acknowledge funding from Beca Chile (CONICYT), Fondecyt 1120419, and the BBSRC.

Literature Cited

Baird, N., Etter, P., Atwood, T. et al. (2008). PLoS One. 3.

Del Rio-Portilla, M. A. and Beaumont (2000). J. Exp. Mar. Biol. Ecol. 254: 1-17.

FAO 2010. The State of World Fisheries and Aquaculture, Rome.

Macavoy, E. S., Wood, A. R. and Gardner, J. P. A. (2008). Aquaculture. 274: 41-48.

Mallet, A. L., Zouros, E., Gartner-Kepkay, K. E. et al. (1985). Marine Biology. 87: 165-172.

Pante, E., Rohfritsch, A., Becquet, V. et al. (2012). PLoS One. 7.

Plutchak, L. L., Simmons, R. E. and Woodruff, D. S. (2006). J. Molluscan Stud. 72: 337-348.

Sauvage, C., Bierne, N., Lapègue, S. et al. (2007). Gene. 406: 13-22.

Strachan, T. and Read, A. P. (1999). 9.2.3 The frequency and spectrum of mutations in coding DNA differs from that in noncoding DNA in: Hum. Mol. Genet. 2nd ed. Wiley-Liss, New York.

Zouros, E. and Foltz, D. W. (1984). Malacologia. 25: 583-591.