A genome-wide association study for host resistance to Ostreid Herpes Virus in Pacific oysters (*Crassostrea gigas*)


1. *The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian EH25 9RG, United Kingdom*
2. *Centre for Environment Fisheries and Aquaculture Science (Cefas) Weymouth Laboratory, Dorset DT4 8UB, United Kingdom*
3. *Department of Biosciences, Ecological Genetics Research Unit, University of Helsinki, Helsinki, Finland*

*alejandro.gutierrez@roslin.ed.ac.uk* (Corresponding Author)

**Introduction**

The ostreid herpesvirus (OsHV-1-μvar) has been suggested as the main cause of mass mortalities of Pacific oyster (*Crassostrea gigas*) worldwide (Segarra *et al.* 2010). Given that Pacific oyster accounts for 98% of global oyster production, estimated at ~0.6 M tons in 2015, this disease is a significant problem for global aquaculture. Improving host resistance to OsHV-1 via selective breeding has become a major target to help prevent outbreaks. A large additive genetic component was described for survival during an OsHV-1 infection, with estimated heritability ranging from 0.21 to 0.63 (Dégremont *et al.* 2015; Azéma *et al.* 2017; Camara *et al.* 2017), suggesting that the resistance could be improved by selective breeding.

SNP (single-nucleotide polymorphism) arrays are important tools for high resolution genetic analysis of complex traits in farmed animal species. In the past few years, many genomic resources for *C. gigas* have been developed including a reference genome assembly (Zhang *et al.* 2012), and a moderate number of genetic markers, such as microsatellites and SNPs (Sekino *et al.* 2003; Wang *et al.* 2015). Importantly, the recent development of medium - high density SNP arrays for oysters (Gutierrez *et al.* 2017; Qi *et al.* 2017) raises the possibility of performing genome-wide association analyses (GWAS) to investigate the genetic architecture of traits of economic interest, and for the application of genomic selection to expedite selective breeding for genetic improvement.

The aim of this study was to investigate the genetic architecture of resistance to OsHV-1 infection in *C. gigas* using a large immersion challenge experiment followed by a GWAS to investigate loci associated with the trait, and their relative contribution to the heritability of the trait.

**Material and methods**

Oyster used in this study were obtained from multiple crosses of parents provided by Guernsey Sea Farms (UK) and reared at Cefas facilities. 1,000 oysters selected for disease challenge at 8 months of age, these included 165 samples from the three nuclear families and 835 samples originating from a mass spawning of several parental oysters, and were selected to include a representative range of oyster sizes. An aliquot of the oyster herpes virus OsHV-1 μvar was added to the water tank at a concentration of 2.49x10⁷ copy numbers/ml with continuous flow. The challenge lasted for 21 days, by which time mortality rate had returned...
to baseline levels, and mortalities and survivors were snap-frozen and stored for DNA extraction. Survival was measured as a binary trait; i.e. 0 (mortality) or 1 (survival). The viral count of all samples, either at point of mortality or at the end of the trial, was determined by qPCR. The obtained copy number was then divided by the weight of the animal to obtain a measure of the viral load. Viral load values were then transformed to log scale for normalization.

Genotyping was carried out at Edinburgh Genomics (UK) using the recently developed Affymetrix SNP array for oysters (Gutierrez et al. 2017). Genotypes were obtained using the Axiom Analysis Suite v2.0.0.35, following the “best practices workflow” with a sample and SNP call threshold of 90%, which resulted in genome-wide genotypes for 854 of the samples. From the ~40K SNPs available for C. gigas on the array, 23,388 were classified as good quality and retained for downstream analyses.

To assist with orientation of the GWAS results, a high density linkage map was constructed using Lep-map 3 (Rastas 2017) using the data from 23 oyster families assigned with Cervus (Kalinowski et al. 2007). Linkage groups were identified using the “SeparateChromosomes2” module using a LodLimit=60 and distortionLod=1. Data was then filtered to remove markers from families showing deviations expected Mendelian segregation (“dataTolerance=0.001”) and used with the OrderMarkers2 module to estimate the order of the markers on the linkage groups (LGs).

Genetic parameters for the trait were estimated using linear mixed models fitting animal as a random effect using ASReml 4 (Gilmour et al. 2014) fitting the following model:

\[ y = Xb + Zu + e \]  (I)

where \( y \) is the observed trait, \( b \) is the fixed effect of sex, \( u \) is the vector of additive genetic effects, \( e \) is the residual error and \( X \) and \( Z \) the corresponding incidence matrices for fixed effects and additive effects, respectively. The covariance structure for the genetic effect was calculated either using pedigree (\( A \)) or genomic (\( G \)) information (i.e. \( u \sim N\left(0, A \sigma^2\right) \) or \( N\left(0, G\sigma^2\right) \)), where \( G \) is the genomic kinship matrix and \( \sigma^2 \) is the genetic variance. The narrow sense heritability was estimated by the additive genetic variance and total phenotypic variance. The underlying liability scale was calculated according to Dempster & Lerner (1950).

GWAS was performed using the GenABEL package (Aulchenko et al. 2007) in R. Genotype data were filtered by the check.markers module keeping SNPs with a MAF>0.01, call rate >0.90 and Hardy-Weinberg Equilibrium < 1x10^-5 leaving 16,223 SNPs. Association analyses were run using the mmscore function fitting the first four principal components (PC). Additionally, GWAS analyses were re-run using the Efficient Mixed-Model Association eXpedited (EMMAX) software (Kang et al. 2010) and used as confirmation of the association. The genome-wide significance threshold was determined by Bonferroni correction at \( p < 0.05 \).

**Results and Discussion**

The heritability for survival based on pedigree was 0.101 ± 0.05 on the observed scale, and 0.254 on the underlying liability scale. Heritability estimated using the genomic relationship matrix was 0.078 ± 0.03 and 0.168 for the observed and liability scale, respectively. For viral load, estimated heritability based on pedigree and genomic matrices was 0.189 ± 0.084 and 0.127 ± 0.045, respectively. Although these estimates show significant
heritability for host resistance, they were lower than other recent OsHV-1 studies (Dégremont et al. 2015; Azéma et al. 2017; Camara et al. 2017), possibly related to the low mortality observed during the challenge (~25%), which could be linked to a higher OsHV-1 resistance of the sampled population compared to other batches of oyster spat exposed to the virus (data not published).

The linkage map contains 20,353 SNPs distributed on 10 LGs (in accordance with the C. gigas karyotype), with a length of 951 cM for the male map and 994 cM for the female map. The ~20K mapped SNPs correspond to 1,921 scaffolds and 149 contigs, according to the latest oyster genome assembly (GCA_000297895.1) (Zhang et al. 2012). The distribution of the markers on the LGs agrees with the linkage map described by (Hedgecock et al. 2015), who identified inconsistency between linkage group assignment and the physical map assignment, suggesting assembly error in within the oyster genome (Zhang et al. 2012) and covering ~87% of the genome. Likewise, we observed that approximately 38% (734 out 1,921) of the scaffolds with informative markers show evidence of errors in the assembly, due to assignment to at least two distinct LGs.

The GWAS results for survival trait identified two markers surpassing the genome-wide significant threshold, as shown in Figure 1. Of the ten markers showing the highest association after confirmation using the EMMAX software, four markers are linked to LG 6, but do not collocate on the same scaffold. The proportion of genetic variation explained by each of the top ten markers ranged between 0.019 and 0.047. The results for viral load showed two SNPs surpassing the genome-wide significance threshold from the GenABEL analysis, as shown in Figure 2, and the EMMAX analysis detected 10 markers surpassing the genome-wide significance threshold. The proportion of genetic variation explained by each of the top ten markers ranged between 0.0209 and 0.037. Several of the significant markers for both traits map to LG 6, either directly or by assessment of the nearest mapped marker in the case of unassigned SNPs, albeit there is little evidence for co-location of these SNPs within a specific region on that LG.

![Manhattan plot showing significant association to survival trait according to GenABEL](image.png)
Figure 1. Manhattan plot showing significant association to viral load trait according to GenABEL

Of the significant SNPs associated with the resistance traits, one was found to be significantly associated with both viral load and survival, and was located in the RAN Binding Protein 9-like which has recently been linked to the interferon gamma signalling pathway (Zhang et al. 2017). Another gene collocating with a significant SNP is Coronin (CORO1B), from a family of genes that have multi-faceted roles in immune response (Tokarz-Deptula et al. 2017). These and other genes may form the basis for downstream functional studies to assess their role in controlling variation in host response to viral infection in oysters.

Conclusion

This study is the first study to report a high density GWAS in Pacific oysters, and was enabled by the development of a SNP array (Gutierrez et al. 2017). Heritability of resistance to OsHV-1 in oysters was significant, but low to moderate in magnitude. The fact that this heritability was detected using both the pedigree and genomic relationship matrix implies that genomic prediction of resistance is likely to be effective. Using the genotype data, a high-density linkage map was constructed for C. gigas, and the GWAS identified numerous markers showing a genome-wide significant association with the traits, suggesting a polygenic nature but with evidence for significant QTL on LG 6, LG 8 and unassigned genome scaffolds. The GWAS identified putative QTL that need to be verified in independent populations, and we also need to evaluate the efficacy of genomic selection as a new tool to enhance host disease resistance.

References

Aquaculture. 469, 50-8.
Hemel Hempstead: VSN International Ltd. www.vsni.co.uk.