ABSTRACTS

SESSION 6: GENOMIC TECHNOLOGY

SINGLE NUCLEOTIDE POLYMORPHISM (SNP) DISCOVERY USING WHOLE-GENOME SEQUENCING OF HUNDREDS OF ANIMALS AND DEVELOPMENT OF A 50K SNP ARRAY FOR NILE TILAPIA

Yáñez J.M. [a,b,§], **Cáceres G.**[a], **Jedlicki, A.**[a] , **Maass A.**[c], **Díaz D.**[c], **Palma R.**[c], **Travisany D**[c], **Lawley C.**[d], **Lhorente J.P.**[b].

[a]*Faculty of Veterinary and Animal Sciences, University of Chile, Santiago, Chile*
[b]*Aquainnovo S.A., Puerto Montt, Chile*
[c]*Center for mathematical Modelling, University of Chile, Santiago, Chile*
[d] *Illumina, San Diego, CA, USA*

Abstract

Nile Tilapia (*Oreochromis niloticus*) is the second most farmed fish in the world and a sustainable source of protein for human consumption. Several genetic improvement programs are established for this species in the world and so far they are mainly based on conventional artificial selection using genealogical and phenotypic information to estimate the genetic merit of breeders and make selection decisions. The incorporation of genomic information, in which has been called genomic selection, can take genetic improvement of farmed tilapia to the next level. Genome-wide information can be exploited to efficiently incorporate traits that are difficult to measure in the breeding goal (i.e. carcass quality and disease resistance traits). A substantial number of single nucleotide polymorphisms (SNPs) are required to investigate phenotype–genotype associations and determine the genomic basis of economically important traits. In this study, we performed *de novo* SNP discovery in three different populations of farmed tilapias. A total of ~28 million non-redundant SNPs were identified across a set of 350 fish by Illumina (HiSeq 2500) whole-genome resequencing of individual samples. After filtering by Minor allele Frequency (MAF < 0.05), Call-rate (CR < 0.8) and Hardy-Weinberg equilibrium (bonferroni corrected p-value < 0.05) a total of 9,826,779 high quality SNPs were available. A total of ~1 million SNPs showed medium to high MAF values across all the populations analysed and they were positioned on the genome assembly O_niloticus_UMD1. A 50K Illumina BeadChip SNP panel is under development. This novel dense tilapia SNP panel will be very useful for the dissection of economically relevant traits, enhancing breeding programmes through genomic selection as well as supporting genetic studies in farmed populations Nile tilapia using high-resolution genome-wide information.

Keywords: next-generation sequencing, genome, genomic selection, Nile tilapia

[§]Corresponding author. Tel.: +56 2 2978 55 33

E-mail: jmyanez@uchile.cl

EXPERIENCES OF USING GENE TRANSCRIPTION DATA FOR eQTL ANALYSIS AND GEBV ESTIMATION

Anna K. Sonesson[a], Hooman Moghadam [ab], Matthew Baranski[ac], Theo H.E. Meuwissen[d]

[a] Nofima, Ås, 1431, Norway
[b] SalmoBreed, Bergen, 5035, Norway
[c] Marine Harvest, Bergen, 5835, Norway
[d] University of Life Sciences, Ås, 1432, Norway

Abstract

Gene transcription data is mainly used to find functional pathways for traits. Often the data sets are very small and specific treatments are tested. Here, we are presenting examples in Atlantic salmon, where SNPs derived from RNA sequencing data are used for expression Quantitative Trait Loci (eQTL) mapping and genomic breeding value estimation. In an eQTL analysis of Atlantic salmon from the SalmoBreed population, we used 362204 SNPs derived from transcriptomic data that had been analysed in 49 individuals with high and 49 individuals with low expression of the Δ6fad_b gene, which is an important gene in the fatty acid synthesis pathway. To correct for population structure we included a Genomic based relationship matrix based on the SNPs derived from the transcriptome data in our model. In addition, we fitted one SNP at a time to search for significant SNP effects. We found four significant eQTL, three were in the Δ6fad_b gene. These SNPs were verified in next generation fish to have large effects on the fatty acid composition. When a stringent minor allele frequency threshold was applied, important SNPs seemed to have been removed, resulting in that false negative are more detrimental than false in this type of analysis. In a GEBV analysis of a data set of 8000 Atlantic salmon from the Marine Harvest population challenged for Pancreas disease (binary trait) and genotyped with a 55k SNP chip, the accuracy of GEBVs was estimated using family-wise 10-fold cross-validation, i.e. entire families were either in the data or masked/predicted. A transcriptomic analysis was performed on 52 PD challenged fish from the Salmobreed population (26 dead and 26 survivors). Effects were estimated on 1316 SNPs, which were significantly differentially expressed, had significant allele specific expression, and which were part of the SNP chip. The accuracy of selection using the GBLUP method of the SNP chip data was 0.818. The accuracy of using the 1316 SNPs from the transcriptomic data only was 0.761. The accuracy of selection of using GBLUP plus a term with the most significant SNPs from the SNP chip and the transcriptome data fitted in a linear combination into clusters (PLS) was 0.924. The accuracy of selection using pedigree data only was 0.63. In the transcriptomic dataset, the 1316 most important SNPs derived from the transcriptomic data explained 37% of the genetic variance for PD resistance.

APIS: A NEW AUTO-ADAPTIVE PARENTAGE INFERENCE SOFTWARE TOLERANT TO MISSING PARENTS

Griot R.[a,b,c], Allal F.[c], Brard-Fudulea S.[a], Cariou S.[d], Haffray P.[a], Vandeputte M.[b,c,§]

[a] *Sysaaf, Syndicat des Sélectionneurs Avicoles et Aquacoles Français, Rennes 35000, France*
[b] *GABI, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas 78350, France*
[c] *MARBEC, Univ Montpellier, CNRS, Ifremer, IRD, Palavas-les-Flots 34250, France*
[d] *Ferme Marine de Douhet, La Brée les Bains 17840, France*

Abstract

In aquaculture breeding programs, parentage assignment based on genomic markers is necessary to reconstruct the pedigree of each fish. In this work, we have developed a likelihood-based assignment method which only requires parent and offspring genotypes to automatically set up the decision rules. The observed distributions of the log-likelihood of the first best and second-best offspring-sire-dam trios are compared, the second-best being used as a reference to detect offspring with missing parents. The user is asked for an acceptable error rate in the assignment, and the threshold is moved accordingly based on the observed distributions. We have compared this method with CERVUS (likelihood-based, with 90% and 95% confidence) and VITASSIGN (exclusion-based, with 1% or 5% mismatches allowed). In APIS, 1% and 5% accepted error rates were tested. Assignment was performed on 1068 real sea bass offspring derived from 14 dams and 39 sires. Two sets of randomly chosen SNP markers from the 57K Axiom_DlabCHIP were tested, one with 100 SNPs (exclusion power = 0.9999) and the other with 35 SNPs (exclusion power = 0.90). The impact of missing parents was tested by masking the genotype of 0, 4, or 20 sires. When the power was high (0.9999), all three software assigned >99% of the offspring without missing parents, but APIS had lower false assignment rates for fish with missing parents. When the power was low (0.90), it performed better than other software except when 20 sires were missing, where its global assignment rate was lower, but with a reliability >75%, while the reliability was 50% for VITASSIGN and <40% for CERVUS. Still, although the performance was better than that of other software, the observed error rate was higher than the expected value, thus some fine-tuning is required. The APIS parentage assignment software requires no parametrization other than the accepted rate of mis-assignment, and performs equally to or better than existing software, especially in cases with missing parents. Further investigations will be done about the impact of the size of the data set and the power of the marker set. The method will be also tested on other data sets.

Keywords: Parentage assignment, Single Nucleotide Polymorphism, Aquaculture

§ Corresponding author. Tel.: +33 4 67 13 04 07 ; Fax: +33 4 67 13 04 58
E-mail address: marc.vandeputte@inra.fr

CRUSTACEAN NUCLEAR LOCALIZATION SIGNALS HELP FACILITATING THE DELIVERY OF DNA INTO AUSTRALIAN RED-CLAW CRAYFISH CELLS

Chan D. H. Nguyen[a], Tomer Ventura[a] and Abigail Elizur[a§]

[a] *GeneCology Research Centre, Faculty of Science, Health, Education and Engineering, University of the Sunshine Coast, 4 Locked Bag, Maroochydore, Queensland, 4558, Australia*

Abstract

Aquaculture is the fastest growing food sector, in which crustacean aquaculture is estimated at 6,915,072 tonnes worldwide, which accounts for US$ 36.2 billion (FAO 2014). The rapid and continual increase in seafood demand requires excelling in aquaculture practice and adopting state-of-the-art biotechnology solutions to enhance productivity. Over the last few decades, the aquaculture industry has employed the delivery of active components such as recombinant hormones and either complete or partial viral capsules, to promote growth, desired features or actively prevent disease spread in the aquaculture product. Hundreds of different plasmids (DNA vectors) are being utilized as carriers to deliver foreign DNA into mammals and fish via either chemical or physical route. To date, there has not been any report of a viable method for reliably transferring foreign DNA into crustaceans. The design of the carrying vector that is adaptive to the host expression machinery is based on the well-established knowledge of vertebrate's transcription. However, this mechanism in crustaceans is less known, therefore constructing an adaptive plasmid accompanied by a proper transfection method is highly challenging. This study has utilized a bioinformatics approach to shortlist Nuclear Localisation Signals (NLSs) specific for crustaceans that can help deliver DNA into the nucleus via active transport through the nuclear complex pore. The NLS-plasmid complex was shown to enable protein expression in the Australian redclaw crayfish (*Cherax quadricarinatus*) primary cell and tissue cultures. These NLSs also increased the transfection efficiency in mammalian cells from 0.6% to 9% in suboptimal conditions. Hence, this integrated approach of bioinformatics and *in vitro* studies could represent the first milestone in developing a novel biotechnology to enable DNA delivery in the crustacean aquaculture industry.

Keywords: DNA delivery, nuclear localisation signal, NLS, Australian redclaw crayfish, Cherax quadricarinatus.
[§] Corresponding author. Tel: +61-7-54594813 ; E-mail address: aelizur@usc.edu.au

HYBRID ASSEMBLY OF THE BLACK TIGER SHRIMP GENOME (*PENAEUS MONODON*)

Juan D Montenegro[1,2§], Roger Huerlimann[1,3], Kirby Siemering[1,2], Lavinia Gordon[1,2], Matthew Tinning[1,2], Jafar S Jabbari[1,2], Jarrod L. Guppy[1,3], Greg J Coman[1,4], Nicholas M Wade[1,5], Melony J Sellars[1,5], Jeff A Cowley[1,5], Dallas Donovan[1,6], Mehar S Khatkar[1,7], Herman W Raadsma[1,7], Gregory E Maes[3,8,9], Kyall R Zenger[1,3], and Dean R Jerry[1,3]

[1] ARC Industrial Transformation Research Hub for Advanced Prawn Breeding
[2] Australian Genome Research Facility Ltd, The Walter and Eliza Hall Institute, 1G Royal Parade, Parkville VIC 3050, Australia
[3] Centre for Sustainable Tropical Fisheries and Aquaculture, College of Science and Engineering, James Cook University, Townsville, QLD 4811, Australia
[4] CSIRO Agriculture and Food, Integrated Sustainable Aquaculture Program, 144 North Street, Woorim, QLD 4507, Australia
[5] CSIRO Agriculture and Food, Integrated Sustainable Aquaculture Program, 306 Carmody Road, St Lucia, QLD 4067, Australia
[6] Seafarms Group Ltd, Level 11 225 St Georges Terrace, Perth, WA 6000, Australia
[7] Sydney School of Veterinary Science, Faculty of Science, The University of Sydney, NSW, Australia
[8] Laboratory of Biodiversity and Evolutionary Genomics, KU Leuven, Leuven, 3000, Belgium
[9] Center for Human Genetics, UZ Leuven- Genomics Core, KU Leuven, Leuven, 3000, Belgium

Abstract

The black tiger shrimp (*Penaeus monodon*) is a commercially valuable species with a natural distribution range from the coasts of the Arabic peninsula to the Pacific coasts of Asia and Australia. While it has decreased to ~13% of global farmed shrimp production over the past 2 decades, *P. monodon* remains the primary species farmed in Australia and other countries with restrictive live import policies and is regaining interest as an alternative to the now widely-farmed Pacific white shrimp (*Litopenaeus vannamei*). Pivotal to it re-emerging as a major farmed species are breeding programs to provide healthy seedstock resistant to the many pathogens that plague its culture. To assist in the ultimate aim of having an advanced breeding program for the species, we are constructing an annotated reference genome for *P. monodon*. We report here a draft genome assembled using an array of second and third generation sequencing technologies. The high repeat content in the species represented a considerable challenge insurmountable for traditional short read technologies and the assembly of linked reads from 10X Genomics Chromium libraries using the Supernova v 2.0 pipeline was not succesful. The initial use of single-molecule sequencing approaches (including the Pacific Biosciences Sequel and Oxford Nanopore Technologies MinION platforms) to generate long reads was hindered by the presence of inhibitors (most likely polysaccharides derived from chitin) in the DNA extracts. However, these difficulties were overcome through extensive optimization of the DNA extraction and purification methods. The contigs of the PacBio assembly were scaffolded using the 10X Chromium barcode data, and the original short reads were used to refine the assembly and to assess its quality and contiguity. Following this, genotype information from a large segregating family of crustaceans was used to produce a linkage map and contigs were anchored to putative linkage groups. Recombination frequencies between loci in the same linkage group will be used to estimate the nucleotide distance between the anchored scaffolds. Similarly, the linkage map will be used to orient the scaffolds within the pseudomolecule. The draft genome has been integrated with the recently-published transcriptome assembly.With these further refinements, the *P. monodon* genome assembly will provide a benchmark resource for shrimp researchers and breeders worldwide.

Keywords:

genome assembly, annotation, linkage map, single-molecule sequencing

§ Corresponding author. Tel.: +61 7 3346 9203
E-mail address: Juan.Montenegro@agrf.org.au