Analysis of gene expression and function involved with fat deposition in Yellowtail Kingfish, using RNA-seq data, NOFIMA, Norway

Paul Whatmore
PhD Candidate
University of the Sunshine Coast

AUSTRALIAN SEAFOOD COOPERATIVE RESEARCH CENTRE

2012/753

This project was conducted by the University of the Sunshine Coast in conjunction with Clean Seas Ltd

The Australian Seafood CRC is established and supported under the Australian Government's Cooperative Research Centres Program. Other investors in the CRC are the Fisheries Research and Development Corporation, Seafood CRC company members, and supporting participants.

Office Mark Oliphant Building, Laffer Drive, Bedford Park SA 5042
Postal Box 26, Mark Oliphant Building, Laffer Drive, Bedford Park SA 5042
Tollfree 1300 732 213 Phone 08 8201 7650 Facsimile 08 8201 7659
Website www.seafoodcrc.com ABN 51 126 074 048

Australian Government
Fisheries Research and Development Corporation

An Australian Government Initiative

ISBN: 978-1-925983-04-3

PROJECT NO: 2012/753

PRINCIPAL INVESTIGATOR: Paul Whatmore

ADDRESS:

University of the Sunshine Coast

90 Sippy Downs Drive
Sippy Downs
Queensland 4556

(PROJECT) OBJECTIVES OF RESEARCH TRAVEL GRANT/ INDUSTRY BURSARY

The primary objective was to receive training in analysing RNA-seq data, from experts in aquaculture bioinformatics, to enable me to analyse a transcriptome dataset for Yellowtail Kingfish which I generated as part of my PhD.

The analysis of RNA-seq data requires the use of specific software tools, access to a powerful computer server or server cluster, and in particular specialised knowledge. By working with NOFIMA staff, I had access to all of these, which guided me through the analysis of my RNA-seq dataset for Yellowtail Kingfish.

NON TECHNICAL SUMMARY:

Through the support and guidance of NOFIMA staff, I successfully constructed a full transcriptome using data from 30 Yellowtail Kingfish (YTK). A transcriptome represents the expressed genes for a particular tissue type, in this case muscle tissue. I also compared the individual transcriptomes from each of the 30 YTK to the full transcriptome, thereby enabling a comparison of gene expression patterns between individuals. We compared 15 animals from high fat families, and 15 from low fat families, and were able to identify over 800 genes involved in fatty acid metabolism, transport and deposition, as well as differential expression patterns for each of these genes.

OUTCOMES ACHIEVED TO DATE: The outcome of this project is knowledge of genes associated with flesh oil in cultured Yellowtail Kingfish and their relative influence of the flesh oil trait.

(PROJECT) OUTPUTS DEVELOPED AS RESULT OF TRAVEL GRANT/ INDUSTRY BURSARY:

Travel grant outputs from grant application:

1. Quantitative data of gene expression for genes associated with fat metabolism and deposition in YTK, based on analysis of RNA-seq data
   - **Achieved**.
2. Identification of thousands of single nucleotide polymorphisms (SNPs) throughout the YTK transcriptome
   - **Achieved**. The SNP database has been generated.
3. Identification of functional markers (functional SNPs, QTL associated SNPs or gene sequences) associated with fat metabolism and deposition in YTK
   - **Partly achieved**. We have identified several genes with divergent expression profiles between families and we have generated the SNP database. Further analysis will identify associations between SNP frequencies and expression data.
4. Scientific publications in international journals of the above
   - **In progress**. A manuscript including the above analysis is currently in preparation

ABOUT THE PROJECT/ACTIVITY

BACKGROUND AND NEED

Previously we have examined heritability of various commercially important traits in Yellowtail Kingfish (YTK). One of those traits was % fat in muscle tissue, which we found to be highly heritable. This provided a strong foundation to explore the genes involved in fat deposition and metabolism in YTK. To do so we sequenced the transcriptome of 30 individual YTK – 6 families and 5 members of each family that were identified as having high or low fat and a high estimated breeding value (EBV) calculated on muscle fat % data. EBV calculates the breeding value based on the performance of all members of a family.

Genomics and transcriptomics using next generation sequencing technology is rapidly becoming a dominant method of examining gene expression, novel genes, population structure, and performing genetic selection and stock improvement. It includes some of the most important and powerful molecular technologies in many biological disciplines and in many industries, including the aquaculture industry. One of these technologies is RNA-seq – deep sequencing of entire transcriptomes between tissues and individuals. RNA-Seq provides a far more precise measurement of levels of transcripts and their isoforms than other methods. As a geneticist, and due to its importance, it was vital for my professional development to learn the tools and techniques for analysing RNA-seq data.

Prior to my trip to Norway, we used Illumina RNA-seq to sequence the transcriptomes of the 30 high and low fat animals and families, which generated almost 60 billion sequenced bases. While producing such vast amounts of sequence data, RNA-seq is both a powerful method for examining genes and gene expression, but also requires specialised knowledge to analyse the data-set as well as powerful software and hardware tools to process the data.

From my discussions with NOFIMA staff, it was clear they had the knowledge, computer hardware and software at their disposal and were happy to train me in their use. Norway has one of the largest aquaculture industries in the world, and is a world leader in aquaculture biotechnology, with NOFIMA at the forefront of research in this field. As such I worked with, and was trained by, top researchers in RNA-seq analysis, which filled a knowledge gap. In turn the valuable knowledge and experience I gained is transferrable to USC and Australian researchers who are generating their own RNA-seq data-sets.

RESULTS

- Constructed a de-novo transcriptome for the YTK from almost 60 billion sequenced bases

- Identified a total of 333,257,460 transcripts
- 67% of transcripts were for novel or previously unknown genes
- 33% were identifiable from the NCBI BLAST database
- 862 genes were involved in fat metabolism, transport and deposition
- Majority of these genes were up-regulated in high fat families

Below is a table of a few of the more highly expressed genes involved in fat metabolism and deposition. TPM = transcripts per million.

Contig ID	TPM - High fat mean	Standard error	TPM - Low fat mean	Standard error	Protein
comp51223_c0_seq1	16.85	1.19	13.61	2.12	Carnitine O-acetyltransferase-like
comp66963_c0_seq1	13.48	0.85	10.43	0.90	Mitochondrial carnitine/acylcarnitine carrier protein-like
comp71564_c0_seq1	120.54	8.69	82.14	6.47	Muscle fatty acid binding protein
comp85976_c0_seq1	123.42	10.44	99.37	13.13	Fatty acid-binding protein, heart-like
comp67958_c0_seq1	26.62	3.33	9.22	1.36	Stearoyl CoA 9-desaturase
comp78184_c1_seq2	102.08	5.95	89.54	9.00	Dehydrogenase/reductase SDR family member 7C-A-like
comp78360_c0_seq1	82.25	3.71	62.39	4.23	Glutaryl-CoA dehydrogenase, mitochondrial-like
comp83060_c0_seq1	88.31	4.76	77.12	8.71	3-hydroxyacyl-CoA dehydratase 1-like

The examples above show a range of genes that have a variety of functions in fat metabolism, transport and deposition. These are typically members of larger functional gene groups. Carnitine transports long-chain fatty acids into the mitochondrial matrix, to be broken down for the citric acid cycle. Seven carnitine transport genes were found in the YTK transcriptome. Similarly, fatty acid binding proteins (FABP) are involved in fatty acid transport, but between inter and intramuscular tissues. Twenty FABPs were found. A total of 370 genes were identified that are involved with lipid transport. Stearoyl CoA 9-desaturase produces the fatty acid oleic acid from a stearic acid precursor. Nine fatty acid desaturases and their isoforms were found. Dehydrogenase/reductase SDR family member 7C-A-like, is an example of a known gene with only a predicted function; steroid hormone regulation. It has been identified in other studies to be highly expressed in adipose tissues but its involvement with fat metabolism is unknown. 3-hydroxyacyl-CoA dehydratase 1-like is responsible for the dehydration step in very long fatty acid synthesis (VLCFA). Twenty genes involved with VLCFA metabolism were found.

By identifying not only individual genes associated with fat metabolism, transport and deposition, but also larger gene families, we are identifying expression patterns between high and low fat YTK families.

INDUSTRY IMPACT

PROJECT OUTCOMES (THAT INITIATED CHANGE IN INDUSTRY)

SUMMARY OF CHANGE IN INDUSTRY
(What immediate changes might be expected for business/industry?)

It is expected that any industry change will occur within the next few YTK breeding cycles and thereafter, meaning 6 months +.

WHAT FUTURE AND ONGOING CHANGES ARE EXPECTED?
(What will be the impact?)

Stock improvement and selective breeding efficiency is vastly improved with the inclusion of genetic information. YTK selection programs are in thier early stages in Australia, and to be successful they need to be supported by good scientific processes. Key to this is selection of stock for specific traits. The traits need to be commercially important and respond well to selection. Flesh oil is a commercially valuable trait. For example, in the Japanese sashimi market kingfish is second most consumed after tuna and fatty-tissue kingfish command a premium price. In addition, we found muscle fat % to be strongly associated with condition, weight and length (thus also growth). With its high heritability and association with other economically important traits, selecting stock on the basis of % fat would yield rapid and effective improvement. The RNA-seq data collected for this project will allow fine-scale examination of specific genes associated with fat. This will allow targeted selection on the basis of the genes identified to have the largest impact on % fat. In addition, the dataset contains a vast number of single nucleotide polymorphisms (SNPs), which can be used to identify families, develop pedigrees and quantify traits. A variety of studies have shown that SNPs have lower error rates, are more reliable and accurate than microsatellites.

It is anticipated that this data will initially be used in industry by identifying specific fat-associated genes in individual animals and selecting broodstock on that basis. In the future it is likely other genes will be identified within this data-set, such as those associated with other commercial traits such as growth and weight. This would lead to a more powerful multi-genic selection approach. Finally, the development of a SNP chip would occur. As discussed, this would be far more accurate and powerful than the currently used microsatellite technology for developing pedigrees, monitoring inbreeding and quantifying trait heritability.

WHAT BARRIERS ARE THERE FOR CHANGES TO OCCUR?

CleanSeas Tuna, who supplied the animals for this project, are currently implementing a YTK genetic selection program, which has already been producing positive results, and the data collected in this project will further enhance this.

However the uptake of molecular technologies by the Australian aquaculture industry in general has been relatively slow and haphazard, particularly when compared to agriculture or aquaculture industries in other countries (e.g. Norwegian salmonid industry).

These technologies are demonstrably successful and effective. The Australian aquaculture industry needs to be forward thinking and have the will to accommodate a genetic approach to selection, rather than traditional (and vastly less efficient) breeding techniques.

Specifically with regards to this project, the industry needs to both have access to the data on fat-associated genes, and make the decision to utilise that data in their selection program.

IF NOT ALREADY HAPPENING, WHEN WILL THE CHANGES OCCUR?
(e.g. 2 businesses will adopt project findings and two more are expected to adopt findings within 12 months)

At the least, Clean Seas Tuna will likely initially adopt these project findings and incorporate them in their selective breeding program within the next 12-24 months. Once the research is published, it is likely that other researchers and companies will adopt the findings. Our results are applicable across most cultured fish species as most of the genes are present in all fish species and many genes are found even in invertebrates. Thus our results would be usable by a large range of aquaculture businesses.

WHAT IS THE LIKELIHOOD THAT THESE CHANGES WILL OCCUR?
(e.g. 50% chance that four businesses will adopt project findings)?

90% chance one business will adopt our findings within the next 12-24 months, and 50% that 2 or more businesses will adopt out findings in the next 3 years.

WHAT BARRIERS ARE THERE TO ADOPTION OF THESE CHANGES AND WHAT ACTION COULD BE TAKEN TO OVERCOME THESE?

(e.g. to adopt project findings will require group training/sharing equipment/invest additional capital etc.)

For the project findings to be incorporated into the current CleanSeas tuna YTK selection program, firstly the results needs finalised and disseminated to the company to gain support for the inclusion of such analysis in the genetic selection program. There is still further R&D required to establish a reliable set of SNPs associated with positive traits associated with fat content.

COMMUNICATION OF PROJECT/EXTENSION ACTIVITIES

WHAT IS THE OUTPUT THAT NEEDS TO BE COMMUNICATED?

Each of the four stated outputs needs to be communicated (see project outputs above), though to two distinct audiences. Scientific outputs, such as quantitative data of gene expression and functional markers need to be disseminated to the scientific community by publishing our findings (see output 4: Scientific publications in international journals). Other outputs, such as functional SNPs and other genetic information that can be used in a selection program, need to be shared with Industry.

WHO IS/ARE THE TARGET AUDIENCE/S?

The scientific community and the aquaculture industry, specifically companies which undertake finfish genetic selection programs.

WHAT ARE THE KEY MESSAGES?

For the scientific community:

- Patterns in differential expression for genes associated with lipid metabolism and transport
- Functional SNPs that affect expression for genes associated with lipid metabolism and transport

For the aquaculture industry

- Functional markers, including genes and SNPs, that can be used in a selective breeding program by identifying individual animals that will produce offspring with a higher fat content

WHAT IS THE CALL TO ACTION?

(What is it you want people to do once you communicate the key message to them – i.e. what change of behaviour or action do you want them to take?)

In the scientific community I would want other researchers to compare and contrast SNPs and gene expression patters for other cultured fish species.

In the aquaculture industry I would expect these findings to be incorporated into selection programs. At CleanSeas tuna a genetic selection program has been established and this information chould be used to enhance the program, specifically for breeding larger, fatter fish.

COMMUNICATION CHANNELS

(How can these messages be communicated and by who?):

Channel	*Who by*	*When*
Scientific publication in an international Journal	Paul Whatmore	June 2014

LESSONS LEARNED AND RECOMMENDED IMPROVEMENTS

WHAT IS YOUR FEEDBACK?

The CRC travel grant has enabled me to work with top people in aquaculture bioinformatics, and thus improve my knowledge and capabilities tremendously. Academic travel is vital for student development and the CRC travel grant program provides a wonderful opportunity. There is a small issue however – I was grateful for the funding I received, however a flat $5,000 for travel is problematic, as travel costs vary tremendously between countries. A research trip to Vietnam, for example, would cost a fraction of the cost of a trip to Norway. Norway is one of the most expensive countries in the world. But they also are world leaders in aquaculture biotechnology so was the logical destination for me. Possibly for future travel grants the payment should reflect the living costs for the specific destination, e.g. rather than a set $5,000 max, the CRC would provide funds for basic accommodation, travel and sundries food (all indexed by cost of living for the destination country).

FURTHER ACTION REQUIRED IN REGARDS TO COMMERCIALISATION?

(e.g. IP protection, licensing, sales, revenues etc)

Currently the data are still being analysed and hence not ready to be commercialised.

ACKNOWLEDGEMENTS

I would like to thank CleanSeas tuna for providing all the animals for this project, NOFIMA staff, in particular Matthew Barranski, for the training in RNA-seq analysis in Norway, and the Australian Seafood CRC for providing the funds and opportunity to greatly improve my scientific knowledge and abilities.