FISHERIES RESEARCH & DEVELOPMENT CORPORATION

FINAL REPORT (DEVELOPMENT AWARD)

AWARD CODE and TITLE

2009/315-32 2014 FRDC International Travel Award- Marianne Douglas

AWARD RECIPIENT: Marianne Douglas, Department of Primary Industries, Parks, Water and Environment (TAS)

HOST ORGANISATION: Hokkaido University, Hakodate, Japan

DATE: 20/03/2015

ACTIVITY UNDERTAKEN

Phylogenetic analysis training encompassed the creation of phylogenetic trees using 16S sequence analysis techniques and the creation of phylogenetic trees with the use of Multi-Locus Sequence Analysis (MLSA) based on sequences of several housekeeping genes.

To be able to apply the phylogenetic analysis techniques mentioned above, knowledge was gained in (i) the preparation of data for sequence analysis in BioEdit in order to create contig sequences, (ii) the use of LPSN (List of Prokaryotes Names with Standing in Nomenclature) and NCBI databases to collect all 16S sequence data of closely related species as well as how to choose the near-related outliers, (iii) construct a single database of sequences of near related taxa for multiple alignment in ClustalX, and finally (iv) to use database file for phylogenetic tree analysis in MEGA6.

For the technique of MLSA, knowledge was gained in (i) collecting house-keeping gene sequence data using NCBI nucleotide database, (ii) how to collate these house-keeping gene data files from 16S reference strains obtained in the phylogenetic 16S analysis, (iii) how to cut all sequences to the same length in MEGA to create alignment topologies and check for protein translation frames in Expasy, (iv) how to convert these files to a 'NEXUS' file in MEGA6, (v) how to concatenate the individual NEXUS files in SplitsTree4, and (vi) how to determine the robustness of the relationships between sequences in the resulting phylogenetic tree. A sequence similarity technique, DNA-DNA hybridisation assay, was also undertaken as part of this study.

OUTCOMES ACHIEVED TO DATE

Over the past 15 years, there have been sporadic, disease outbreaks in hatchery Atlantic salmon in Tasmania caused by an unidentified species of *Nocardia*, a bacterium associated with chronic infections. Sequence data was collected from representatives of the library of isolates. At Hokkaido University, phylogenetic analysis was undertaken using 16S and housekeeping gene data sets for the

Nocardia isolates using the newly mastered techniques. The results were very interesting with the Tasmanian strains forming their own cluster and closely related, but not necessarily identical, to *Nocardia jejuensis* (see Figure 1 in Appendixes); the evidence suggests that the Tasmanian isolates may represent a new species of *Nocardia*.

The knowledge gained during the study period in Japan was brought to immediate use at the Animal Health Laboratory. From a recent diagnostic case, a *Vibrio* species was isolated from diseased abalone. The organism could not be identified by conventional means. The 16S sequence was obtained for the *Vibrio* isolate and using the 16S data base and knowledge brought from Japan, identification through phylogenetic analysis and the construction of a phylogenetic tree, was achieved. Analysis of the data revealed that the isolate could be identified as *Vibrio lentus*, which was subsequently reconciled with the conventional identification method (see Figure 2 in Appendixes). The ability to identify new species or atypical isolates is an immediate practical example of how the newly gained knowledge could be used to identify aquatic pathogens from diagnostic cases.

Currently SOPs on the phylogenetic analysis technique are in the process of being written so that the procedure can be used at the Animal Health Laboratories for future diagnostic cases.

Acknowledgments

I am much indebted to Dr. Tomoo Sawabe of the Hokkaido University in Hakodate for hosting this study opportunity and providing me with the 16S *Virbrionaceae* and the *Vibrio harveyi* housekeeping genes sequences data sets for use in our laboratory, as well as helping me generate 16S and housekeeping genes sequences data sets for the *Nocardiaceae*. Both already will be of enormous use to our laboratory. I also acknowledge the time Dr. Sawabe has spent on teaching me the techniques described in 'Activities Undertaken'. I also would like to acknowledge two of Dr. Sawabe's PhD students, Nurhidayu binti Al-saari and Sayaka Mino for helping me gather sequence and reference data and pointing me in the right direction when I got stuck, and to Nurhidayu binti Al-saari for guiding me through the DNA-DNA hybridisation technique in the laboratory. Last but not least I would like to acknowledge the whole team of students at the Faculty of Fisheries and Science for making me feel welcome in Japan.

Background

The Animal Health Laboratory (AHL) of the Department of Primary Industries, Parks, Water & Environment plays a significant role in the diagnosis and management of infectious diseases throughout Tasmania's farming sector. The ability to rapidly and accurately identify emerging diseases is essential for AHL to provide this service. The Atlantic salmon industry is a rapidly expending industry in Tasmania and as such is particularly vulnerable to the emergence of new diseases. Failure of AHL to keep up to date with developing diagnostic and analytical methodologies may lead to delays or failures resulting in significant cost to the industry. The week long intensive training in Japan seeks to build capacity at AHL to respond to emerging diseases within the salmonid industry. The need for further development in the molecular testing facility was discussed with and was recognized by Dr. Jeremy Carson, whom is the principal investigator of a FRDC funded project to establish an Aquatic Animal Health Centre of Excellence at Mount Pleasant Laboratories. Furthermore, the aims for this project was endorsed by Dr. Steve Percival, Huon Aquaculture Company as

well as Dr. Adam Main of the Tasmanian Salmonid Growers' Association. Letters of support were provided when applying for funding.

Need

The Molecular Biology group at AHL provides diagnostic and disease investigation services for aquaculture industries in Tasmania. The increasing reliance on molecular techniques for the detection and identification of disease agents in aquatic animal health is undeniable. Due to the ever increasing number and more efficient molecular techniques becoming available and the dynamic nature of newly emerging pathogens, specialist training is increasingly important for the delivery of effective and useful services to aquaculture industries. Presently, species identification relies heavily on biochemical tests and cellular fatty acid analysis, or 16S rDNA gene sequencing. For well-defined microbial pathogens these approaches are cost-effective, however for unusual or newly emerged pathogens different techniques are required for identification. Multi-locus sequence analysis (MLSA) uses several stable housekeeping genes to assign a species designation, and is a technique that is easy to use, accurate, and has great discriminatory power (McTaggart *et al*., 2010). The MLSA skills acquired would enable us to more specifically and accurately perform species identification and determine genetic diversity among different isolates and build up a reference gene bank of sequences that can be used for identification in the future.

Objectives

1. Obtain training in advanced molecular biology techniques with an emphasis on DNA sequencing and phylogenetic analysis at a known international specialist laboratory.

2. Undertake training in bioinformatics programmes with an emphasis on Multilocus Sequence Analysis (MLSA) at a known international specialist laboratory.

3. The possibility of the generation of a gene bank of sequences in the future, which can be used for reference purposes when identifying unknown pathogens, which increases the speed of identification and therefore decreases potential losses to the aquaculture industry.

4. Strengthen the valuable network connection already established with Dr. Tomoo Sawabe, Hokkaido University, Japan, who is considered to be the internationally specialist for MLSA for the *Vibrionaceae*.

Methods

The training was undertaken at the Faculty of Fisheries, Hokkaido University in Hakodate, Japan. All training was performed using the University's computer system and specialist sequence analysis software as listed:

- BioEdit http://www.mbio.ncsu.edu/bioedit/bioedit.html
- ClustalX http://www.clustal.org/download/current/clustalx-2.1-macosx.dmg
- Mega6 http://www.megasoftware.net/mega.php
- Expasy http://web.expasy.org/translate
- SplitsTree4 http://www.splitstree.org

All the analytical software programs are open access may be downloaded and used without restriction.

Upon arrival, an initial conference was had with Dr. Sawabe, and his two PhD students, currently working on MLSA. Type and quality of already generated sequence data for the group of Atlantic salmon *Nocardia* isolates from Tasmania was assessed, and it was determined if new sequence data needed to be generated and how this might be undertaken.

Firstly, a *Nocardiaceae* 16S sequence data base was made using type strain information from the All-Species Living Tree project (http://arb-silva.de/projects/living-tree), which Dr. Sawabe provided (see attachments). The sequence data base was downloaded from the internet, using the Ribosomal II Data Project (http://rdp.cme.msu.edu), and data was checked and adjusted for completeness and accuracy, using LPSN (List of Prokaryotes Names with Standing in Nomenclature: (www.bacterio.net)). Once the reference database had been built the 16S sequence data of the Tasmanian *Nocardiaceae* isolates was added to the database. The sequence data was used for phylogenetic analysis and the construction of a phylogenetic tree, based on the 16S sequences of type strains within the family of *Nocardiaceae*, was generated using BioEdit, ClustalX, and MEGA6.

Remaining time was spent on obtaining the appropriate sequences of housekeeping genes within the family of *Nocardiaceae*. Sequence data was only generated for the same housekeeping genes that had been used for the Tasmanian *Nocardia* isolates, using the list of type strains generated the day before supplemented with data from the NCBI (National Centre for Biotechnology Information) nucleotide website (www.**ncbi**.nlm.nih.gov/nuccore).

The names of all isolates were simplified in preparation for MLSA and its graphical representation. On the remaining days of the study visit, all housekeeping gene data sets were checked for consistent information between the different housekeeping gene data sets and correct strain type (only type stains were used). Also, sequences for each housekeeping gene were cut to equal length within each house keeping gene sequence data set using MEGA6. Finally, the correct protein translation frame was determined using Expasy (http://web.expasy.org/translate/), and nucleotide numbers adjusted in order to obtain the correct protein reading frame. This was verified by translating the nucleotide sequence to a protein sequence and entering this information into BLAST (Basic Local Alignment Search Tool) protein search (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to check the identity of the protein coding gene. Finally, the generated data sets of the housekeeping genes were entered into SplitsTree4 and a phylogenetic tree generated using the sequences of the housekeeping genes (see Figure 1).

During this study the opportunity arose to perform the technique of DNA-DNA hybridisation. This two-day procedure measures the degree of genetic similarity between pools of DNA sequences, and is a necessary technique in order to confirm if a new species has been discovered by MLSA.

Results/Discussion

Key tools gained:

- Ability to use sophisticated sequence analysis software programs such as BioEdit, ClustalX, Mega6, Expasy and SplitsTree4.

- Knowledge on how to prepare and manipulate sequence data, and create or extend a gene bank of sequences data sets for immediate use or future reference for the bacterial groups *Vibrionaceae* and *Nocardiaceae*.

- Ability to perform phylogenetic analysis in order to create phylogenetic trees using 16S sequence data sets so as to determine genetic diversity among different isolates.

- Ability to perform phylogenetic analysis, using MLSA to more specifically and accurately perform species identification.

- Ability to perform DNA-DNA hybridisation.

- Strengthened the connection between AHL, DPIPWE and the Faculty of Fisheries, University of Hokkaido, Hakodate, Japan via Dr. Sawabe and his research group.

Benefits and Adoption

Tasmanian aquaculture industries (salmonids, abalone, oysters) will benefit from this project by AHL, as diagnostic service provider, being able to identify new or emerging pathogens efficiently and effectively. The analytical approaches which have been acquired, together with acquisition of gene sequence reference databases can be used for identifying unknown aquatic animal pathogens. An appropriate and current example would be the preliminary identification of Tasmanian Atlantic salmon isolates of *Nocardia* as being a likely new species related to *Nocardia jejuensis*. On returning from the study visit, the newly acquired analytical techniques were used to aid in the identification of the bacterial pathogen *Vibrio lentus*, previously unidentified from 16S gene sequencing alone.

This study has resulted in an added capability for rapid detection and identification of new or existing pathogens. The ability to accurately identify pathogens contributes directly to disease mitigation, management and control strategies leading to reduced cost of production for farms. In a broader sense the new analytical approach strengthens Tasmania's biosecurity

Specific training in DNA sequence analysis and its application to the identification of partly-identified or newly emerging aquatic animal pathogens adds to the strength of the Centre of Aquatic Animal Health and Vaccines that is being established at the Animal Health Laboratory through FRDC project 2013/051.

This project has strengthened the established relationship between staff at AHL and Dr. Sawabe and his team at Hokkaido University. Dr. Sawabe has offered to provide further assistance when the need arises by videoconference on Skype.

With the aid of technology transfer and written SOPs these new analytical techniques will be adopted and implemented as a general diagnostic tool for future disease investigations and diagnostic cases at AHL.

Further Development

The sequence analysis of the novel isolates of *Nocardia* from farmed Atlantic salmon has shown that the isolates may represent a new species. Multilocus sequence analysis is increasingly found to be a robust approach to defining species boundaries and is a good predictor of taxonomic structure. Formal evidence to propose a new species requires a polyphasic approach that must include a description of the phenotype along with an analysis of the genotype which must include the sequence of the 16S rRNA gene supplemented by MLSA analysis using housekeeping genes defined as predictive of the genus. While this approach has proved predictive, international rules for defining a new species also requires the use of whole genome DNA-DNA hybridisation measurements between the novel species and near related species as shown by phylogenetic analysis.

From the work undertaken with Dr Sawabe, there is good evidence that the novel isolates represent a new species with similarities to *N. jejuensis* but unlike the known fish pathogens *N. seriolae* or *N. salmonicida*. To confirm the status of the novel isolates DNA-DNA hybridisation needs to be undertaken. The general approach to the technique was given by Dr Sawabe's staff but will need to be undertaken in Australia with the novel isolates and related species. This work is important as it will help define the pathogen of farmed Atlantic salmon in Tasmania. Having established the nature of the species, it will be possible to develop molecular based techniques such as PCR for rapid and accurate identification. These essential tools can be used for disease investigation, monitoring and surveillance. Defining the species will also lay the foundation for any vaccine development work in the future.

Defining the novel species would build on the study program reported here. The work required would form the basis of a small project that would include the following activities:

- Generation of a complete base-pair 16S rRNA gene sequence data set for the Tasmanian Atlantic salmon *Nocardia* isolates.
- For reference purposes obtain the type strain of *Nocardia jejuensis* and other related species.
- Undertake DNA-DNA hybridisation of analysis of the novel *Nocardia* isolates with *N. jejuensis* and other related species as required.
- Phenotypic characterisation of the novel *Nocardia* isolates.
- Development of a PCR test for the specific identification of the novel *Nocardia* species.

The project would establish the taxonomic status of the new species and develop diagnostic tools that could be used for the identification and detection of the pathogen in Atlantic salmon.

References

McTaggart, L. R., Richardson, S. E., Witkowska, M. and Zhang, S. X. (2010) Phylogeny and Identification of *Nocardia* Species on the Basis of Multilocus Sequence Analysis. *Journal of Clinical Microbiology*, **48**: 4525–4533

Programs used:

http://www.mbio.ncsu.edu/bioedit/bioedit.html

http://www.clustal.org/download/current/clustalx-2.1-macosx.dmg

http://www.megasoftware.net/mega.php

http://web.expasy.org/translate

http://www.splitstree.org

Data bases used:

http://arb-silva.de/projects/living-tree

http://rdp.cme.msu.edu

www.bacterio.net

www.ncbi.nlm.nih.gov/nuccore

http://blast.ncbi.nlm.nih.gov/Blast.cgi

http://www.bacterio.net/

Appendices

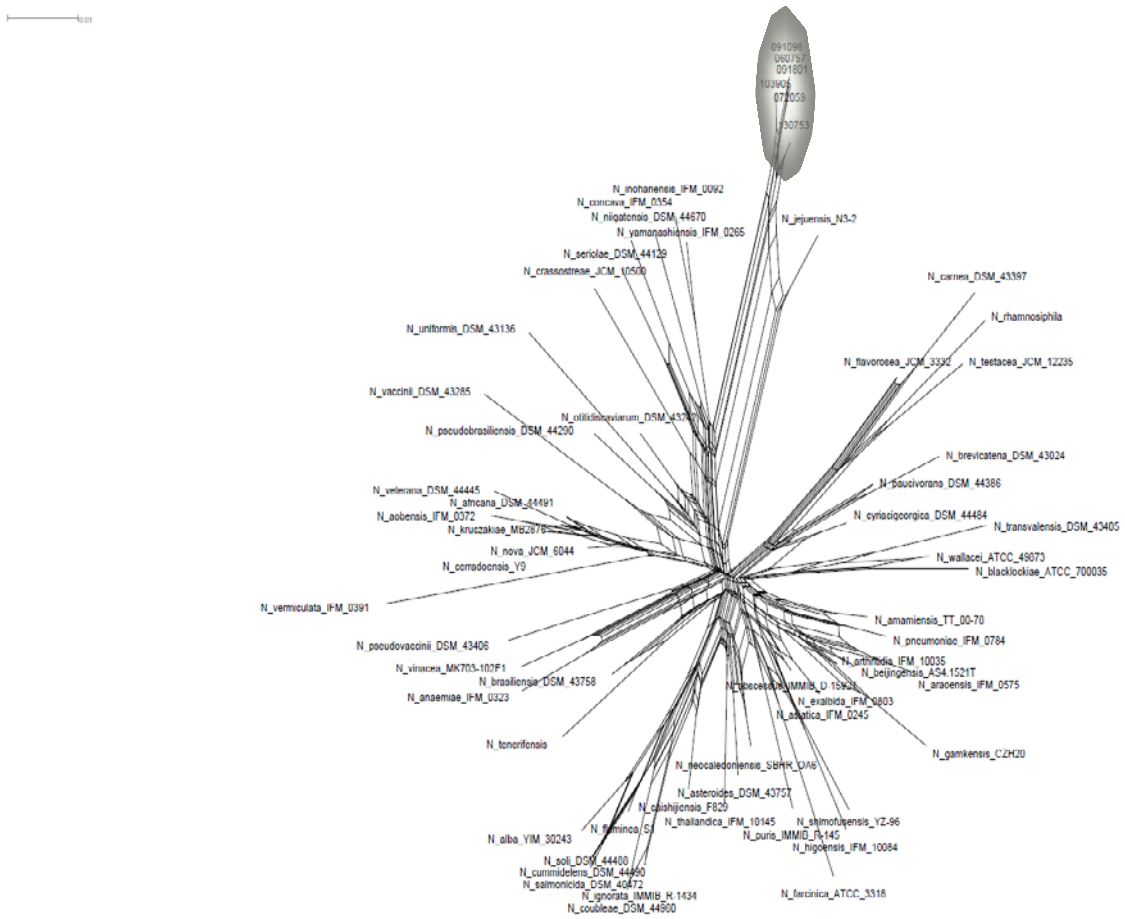

Figure 1: Phylogenetic tree based on MLSA of the housekeeping genes 16S, *gyrB* and *hsp65* for type strains of *Nocardiaceae* showing the position of the Tasmanian isolates within the family of *Nocardiaceae*. Bootstrap values are expressed as a percentage of 1000 replications

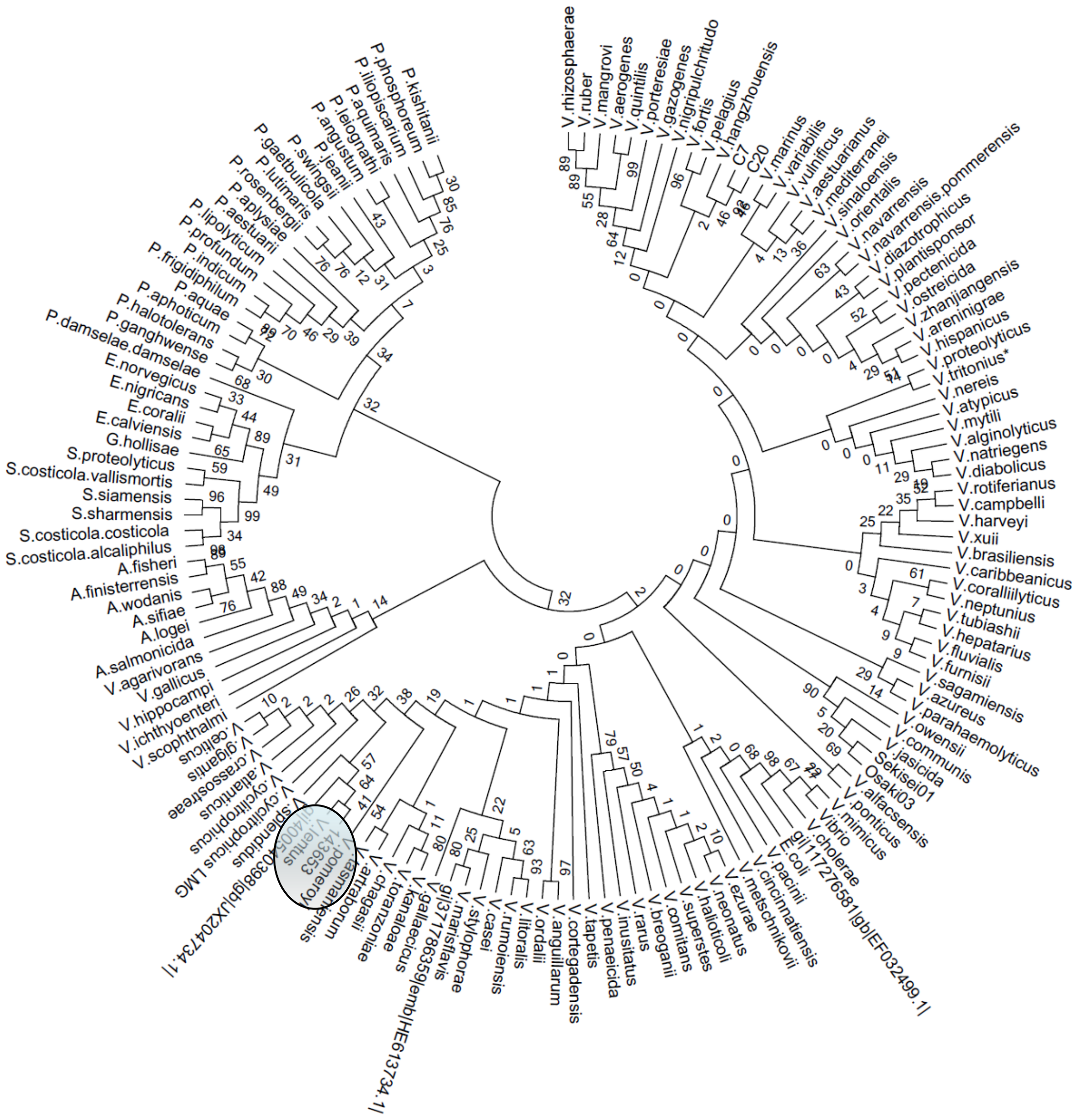

Figure 2: Maximum-likelihood phylogenetic tree, based on the partial 16S rRNA genes sequences of various *Vibrio* species, showing isolated14/3653 within the family of *Vibrionaceae*, and clustering with *V. lentus*. Bootstrap values are expressed as a percentage of 1000 replications