



Metagenomic Amplicon Sequencing as a Rapid and High-throughput Tool for Aquatic Biodiversity Surveys

PhD Thesis

Jennifer L.A. Shaw

Submitted February 2015



**Metagenomic Amplicon Sequencing as a Rapid
and High-throughput Tool for Aquatic
Biodiversity Surveys**

A Thesis Submitted By

Jennifer L. A. Shaw
B.Sc. (Hons) M.Sc. (Hons)

For The Degree Of
Doctor of Philosophy (Ph.D.)
February 2015

**Australian Centre for Ancient DNA (ACAD)
School of Biological Sciences**

**Faculty of Sciences
The University of Adelaide**

Supervisors:
Prof. Alan Cooper
Dr. Laurence J. Clarke
Dr. Laura S. Weyrich
Dr. Paul Monis (SA Water Corp)

Table of Contents

Summary	11
Declaration	13
Preface	15
Chapter 1. Introduction	19
1.1 Decline in aquatic system health and quality due to a global increase of human-mediated ecosystem stressors	21
1.1.1 Invasive species	22
1.1.2 Public health risks associated with the exploitation of food and water resources..	24
1.2 Biodiversity monitoring for effective management of aquatic ecosystem health and quality	28
1.2.1 Detecting Aquatic Vertebrates	29
1.2.2 Detecting invertebrates	30
1.2.3 Detecting microorganisms.....	31
1.2.4 Improving current monitoring methods	33
1.3 Metagenomic amplicon sequencing as a rapid and effective tool for biodiversity surveys	33
1.4 Thesis Overview	35
1.4.1 Chapter two: Using eDNA sequencing for aquatic biodiversity surveys: a beginner's guide	36
1.4.2 Chapter three: Ground-truthing environmental DNA: a comparison of eDNA and conventional fish survey methods in a complex river system	37
1.4.3 Chapter four: Genetic assessment of ballast and marine port sediment samples provides a useful tool to determine historical distribution of harmful microalgal taxa..	37
1.4.4 Chapter five: Using amplicon sequencing to characterize and monitor bacterial diversity in drinking water distribution systems	38
1.4.5 Chapter six: Assessing the impact of water treatment on bacterial biofilms in drinking water distribution systems using high-throughput DNA sequencing.....	39
1.4.6 Chapter seven: General discussion and conclusions	39
1.5 Significance of this thesis	40
1.6 References	41
Chapter 2. Using eDNA sequencing for aquatic biodiversity surveys: a beginner's guide	49
Statement of Authorship	50
Abstract	51
2.1 Introduction	52
2.2 An overview of eDNA sequencing for biodiversity surveys	55
Contamination considerations	58
2.3 Sample collection	59
2.4 Sample processing and DNA extraction	62
2.5 PCR amplification, primer design, and genetic marker choice	64
2.6 Next Generation Sequencing (NGS): primer construct and NGS platform choice	68
2.7 Bioinformatic Analysis	70
Bioinformatic processing tools	74
Visualisation and interpretation of eDNA data	75
Sequence read abundance and species richness	77
2.8 Concluding remarks	78
2.9 References:	79

Chapter 3. Ground-truthing environmental DNA: comparison of eDNA and conventional fish survey methods in a complex river system.	89
Statement of Authorship	90
Abstract	93
3.1. Introduction	94
3.2. Methods	96
3.2.1 Site information and sample collection	96
3.2.2 DNA extraction	98
3.2.3 PCR and DNA library preparation	99
3.2.4 Data processing	100
3.3. Results	101
3.3.1 Effect of sample media on eDNA species assemblage	101
3.3.2 Comparison of 12S and 16S genetic markers	103
3.3.3 Comparison of fyke net and eDNA sequencing surveys	105
3.3.4 Increased sample number improves detection of rare species	109
3.4. Discussion	112
3.4.1 Beware spurious identifications	112
3.4.2 Water eDNA better represents fish communities than sediment	113
3.4.3 Careful primer design and choice is critical	114
3.4.4 A higher number of replicate samples is needed to detect rare taxa	115
3.4.5 Future research suggestions	116
3.4.6 Conclusion	118
Acknowledgments	119
3.5. References	120
Supplementary materials for: Ground truthing eDNA: Comparison of eDNA and conventional fish survey methods in a complex river system	125
Chapter 4. Genetic assessment of ballast and marine port sediment samples provides a useful tool to determine the historical distribution of harmful microalgal taxa.	139
Statement of Authorship	140
Chapter 5. Using amplicon sequencing to characterise and monitor bacterial diversity in drinking water distribution systems.	173
Statement of Authorship	174
Supplementary Materials for: Using amplicon sequencing to characterize and monitor bacterial diversity in drinking water distribution systems.	189
Chapter 6. Assessing the impact of water treatment on bacterial biofilms in drinking water distribution systems using high throughput DNA sequencing	215
Statement of Authorship	216
Supplementary Information for: Assessing the impact of water treatment on bacterial biofilms in an experimental drinking water distribution system using high-throughput DNA sequencing.	227
Chapter 7. General Discussion	235
7.1 Thesis Summary	237
7.1.1 Addressing and overcoming current biomonitoring limitations	237
7.1.2 Validation of metagenomic techniques	242
7.2 Current limitations and required future research	245
7.2.1 DNA extraction causes bottlenecks in processing times	246
7.2.2 Spurious Reads Detected	248
7.2.3 Incomplete databases and lack of genetic reference genomes	251
7.2.4 Discrepancies between morphological- and genetic-based systematic classifications	252
7.2.4 Computational demand	253
7.2.5 Sequence abundance and taxon biomass	254

7.3 The Future Benefits of Metagenomic eDNA sequencing.....	255
7.3.1 Biomapping	255
7.3.2 Extending knowledge of microbial diversity and distribution	257
7.3.3 Using amplicon sequencing combined with other genetic tools to expand knowledge about anthropogenic impacts on biological communities	258
7.4 Concluding remarks.....	261
7.5 References.....	262
Author Curriculum Vitae	269

*For Mum, Dad, and Stephen.
(Your love and support made this possible)*
xxx

Summary

Healthy aquatic systems are essential for life on this planet, and provide a variety of goods and services for humans such as drinking water, food production, waste disposal, and climate regulation. Anthropogenic impacts such as over exploitation of aquatic resources, introductions of invasive species, pollution, and climate change pose great risks for the health and sustainability of these ecosystems. Because of these risks, extensive and detailed biological surveys are regularly required to monitor and manage aquatic ecosystem health. Traditional survey approaches, including morphological-based identification and counting of organisms, are time consuming, costly, and dependent upon highly skilled taxonomic experts. Recently developed molecular methods, where DNA mixtures present in environmental samples are sequenced and taxonomically identified using genomic markers, are rapid and cost-effective, and may substantially improve biological surveys and ultimately aquatic system management. In particular, metagenomic amplicon sequencing of environmental DNA (eDNA) can characterize hundreds to thousands of species within a single sample in a timely and cost effective manner, and allow hundreds of samples to be sequenced in a single reaction.

This PhD thesis aims to develop and refine eDNA amplicon sequencing approaches in order to examine current global problems in aquatic ecosystems. Specifically, this thesis includes a review of eDNA amplicon sequencing protocols and provides recommendations for sampling aquatic environments, laboratory procedures, and bioinformatics processes (Chapter 2). Following this, chapter three utilized eDNA sequencing to monitor invasive and threatened fish species in a sensitive and ecologically important river system, comparing results from fyke net- and eDNA-based surveys to analyze the accuracy and effectiveness of eDNA amplicon sequencing approaches (Chapter 3). In Chapter four, I

examine the introduction and distribution of harmful algal taxa in high-risk marine locations across the entire continent of Australia, characterising a variety of harmful algal species associated with international and domestic shipping activities to potentially manage and mitigate the spread of these species. The use of historical port sediment samples within this study provides valuable temporal information to establish a baseline for the biodiversity distribution of harmful algae, essential for several international conventions focused on limiting the transmission of these harmful microorganisms. In chapters five and six, eDNA sequencing was utilised to identify bacterial taxa within drinking water distribution systems (DWDSs). Chapter five focuses on comparing current commonly used compliance measures with metagenomic approaches while screening for key indicator and pathogenic species throughout two full-scale DWDSs. Chapter six utilized this novel sequencing approach to analyze the efficiency of different water treatment procedures, while focusing on limiting biofilm formation in DWDSs. Overall, the thesis develops and demonstrates the practical applications of metagenomic eDNA sequencing on three distinctive taxonomic groups: vertebrates, eukaryotic microbes, and bacteria. The methods developed throughout provide critical advances for environmental monitoring organisations, including governmental departments, shipping, fishing and aquaculture industries, and water quality corporations. Further, this thesis and the future use of these molecular approaches greatly extends the knowledge of both bacterial and eukaryotic microbial communities in natural and man-made aquatic environments, improving industry efficiency and mitigating public health crises.

Declaration

- This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by any other person, except where due reference has been made in the text.
- I give consent to this copy of my thesis when deposited in the University library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.
- The author acknowledges that copyright of published works contained within this thesis (as listed below) resides with the copyright holder(s) of those works.
- I also give my permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue, the Australian Digital Theses Program (ADTP) and also through web research engines, unless permission has been granted by the University to restrict access for a period of time.

Jennifer L. A. Shaw

Date 25th Feb 2015

Preface

This PhD thesis is a result of a three-year project conducted at the University of Adelaide, South Australia. Laboratory work was carried out in both the South Australia Water Corporation's microbiology laboratories in Victoria Square and dedicated low-contamination DNA facilities at the Australian Centre for Ancient DNA (ACAD) within the University of Adelaide. Both laboratories contain up to date appliances and technologies, which enabled this project to be carried out effectively with minimal contamination risks. All material enclosed in this document is my own work except where there is clear acknowledgement and reference to the work of others. This PhD study was funded by the Australian Research Council (LP0991985) and is part of a collaboration between ACAD and SA Water. SA Water has shown significant interest in the development of this technology due to a high cost currently associated with environmental and water quality monitoring, which they are aiming to reduce. This thesis also resulted in collaborations with other environmental organisations and universities; for example, the Department of Environment, Water and Natural Resources (DEWNR), the Murray-Darling Basin Authority, and the Institute for Marine and Antarctic Studies (IMARS) at the University of Tasmania (UTAS). The methods developed and discussed within this thesis will enable greater sampling effort and increase the amount of data available to decision makers, allowing more informed water management decisions at a time when water and food security, and general aquatic system health is critical.

*'Only when the last tree has died, the last river
been poisoned, and the last fish been caught, will
we realize we cannot eat money.'*

~ Cree Indian Proverb ~