Isolating microsatellite markers in the marine sponge *Cinachyrella alloclada* for use in community and population genetics studies

A thesis submitted to the University of Manchester for the degree of MPhil Environmental Biology in the Faculty of Life Sciences

Sarah Miriam Griffiths

2013

Contents

Word count – 11,118

List of figures and tables

Abstract

Molecular genetics is commonly used to address ecological questions and aid in ecosystem management and conservation. The level of genetic variation, and its distribution among populations, has ecologically important consequences, such as for associated community biodiversity and ecological resilience. Microsatellites are popular and useful molecular markers with many applications, but must be developed *a priori* for every species, which traditionally has been a costly and time-consuming process. Recently, next-generation sequencing approaches have been successfully used to isolate microsatellite markers for a lower cost and much more quickly than traditional methods offer. Here, we develop an existing method that uses Illumina paired-end sequencing, to isolate microsatellites for the marine sponge *Cinachyrella alloclada*. We use longer read lengths than previously used (2 x 250bp) and filter sequence data by quality in order to increase the number of amplifiable loci located and the chances of successful amplification. Out of 35 loci tested, 13 were successfully amplified in PCR reactions. These loci can go on to be used in population genetics and community genetics studies for *C. alloclada*, aiding understanding of its ecology. In addition, this method of microsatellite isolation offers a faster and cheaper way of obtaining molecular markers than traditional approaches, and uses quality filtering and longer read lengths to enhance the capture of amplifiable loci. This method may therefore be used for other species that would benefit from the availability of microsatellite markers, including other understudied marine sponges.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University's policy on Presentation of Theses

Acknowledgements

This project was only possible with the support of my supervisor, Professor Richard Preziosi, who provided invaluable assistance throughout the whole project. I also thank Nathan Truelove, who helped greatly with the project and provided much support with the sequencing and bioinformatics stages of the project especially. Thanks also to the Preziosi lab group for valuable feedback and assistance, and especially Rachael Antwis for the tea and moral support. I also thank the Walton lab for use of their lab space and equipment, and Dr. Petri Kampainnen for always being willing to share advice and ideas over a coffee. The expertise and knowledge of Andy Hayes and Stacey Holden at the Genomic Technologies Core Facility, and Peter Briggs and Ian Donaldson at the Bioinformatics Core Facility (both at the University of Manchester) have been an essential part of this project. Many thanks also to all of their colleagues for making such brilliant resources available to us all. Thanks are also due to those who assisted in fieldwork to collect samples. In Florida, Professor Mark Butler IV (Old Dominion University) and Dr. Donald Behringer (University of Flordia) and their teams collected the original sponge sample used for sequencing and further samples from Florida Bay used in microsatellite testing, for which I am very grateful. Finally, I must thank my long-suffering family and friends for their unconditional support.

Abbreviations

AFLP – amplified fragment length polymorphism

bp – base pair

DEFRA – Department for Environment, Food and Rural Affairs

MPA – marine protected area

mtDNA – mitochondrial DNA

NGS – next-generation sequencing

PAL – potentially amplifiable locus

Seq-to-SSR – developing microsatellites directly from sequencing data, with no prior enrichment

SNP – single nucleotide polymorphism

SSR – simple sequence repeat

Introduction

Marine ecosystems are both biologically and economically important, but are facing a plethora of problems which threaten their existence in their current forms. These threats include overfishing (Worm *et al.* 2009), climate change (Hoegh-Guldberg *et al.* 2007), ocean acidification (Cooley *et al.* 2006; Wittmann & Pörtner 2013) and pollution (Johnston & Roberts 2009). Halpern *et al.* (2008) indicate that human influences which drive ecological change have affected every marine ecosystem on Earth, with coral reefs receiving one of the highest impact factors. Coral reef ecosystems are a hotbed of biodiversity, containing more diversity per unit area than any other marine system (Knowlton *et al*. 2010). They are also of vital socioeconomic importance, providing food, protection from storms, tourism revenue and waste detoxification, amongst other goods and services (Moberg & Folke 1999). With continued degradation, it is feared that biodiversity will continue to decline, and vital ecosystem goods and services will be lost (Worm *et al.* 2006).

Molecular tools are increasingly being used to address ecological questions, increase our understanding of marine ecosystems and aid in their management and conservation (Féral 2002). Genetic variation in populations can have ecologically important effects, and is an important component of biodiversity (Hughes *et al.* 2008). When there is genetic variability for traits that affect fitness, the genetic diversity of a population as a whole can be important for its resilience and ability to adapt to changing conditions. This becomes especially important in the face of climate change, where the capacity for a population to evolve in response to elevated temperatures may be influenced by its levels of genetic variation (Hughes *et al.* 2003). In the seagrass *Zostera marina,* it has been shown that increasing genotypic diversity increases resilience of the ecosystem community by increasing resistance to grazing disturbance (Hughes & Stachowicz 2004) and temperature stress (Ehlers *et al.* 2008), and enhancing recovery of the population following disturbance (Reusch *et al.* 2005).

Intraspecific genetic diversity is also important beyond its effect on populations, as genes affect interactions with other species, and thus have the potential to show effects at the community level. In dominant host species, genetically-diverse populations have been shown to support more species-rich communities than genetically-poor ones (Johnson *et al.* 2006; Zytynska *et al.* 2011, 2012; Whitham *et al.* 2012). Although most work in this area has been carried out in terrestrial systems, Reusch *et al.* (2005) showed that higher levels of genotypic diversity in *Z. marina* supported a higher abundance of associated epifaunal invertebrates.

Due to the nature of the marine environment, many marine species have fragmented populations that are linked through movement of individuals, usually at the larval stage. Molecular tools can be used to show the levels of exchange, or connectivity, of individuals between spatially separate populations (Hellberg *et al.* 2002). The level of connectivity, or the amount of isolation a population experiences has important implications for the persistence of the population in the face of local environmental disturbance; well-connected populations can be repopulated by geographically separate populations, thus aiding their recovery (Botsford *et al.* 2009; Cowen & Sponaugle 2009; Bernhardt & Leslie 2013).

There are increasing calls to consider such genetic factors in the design of marine protected areas (MPAs) (Botsford *et al.* 2003; Palumbi 2003; Almany *et al.* 2009; Christie *et al.* 2010). A lack of genetic diversity within a marine reserve and low levels of connectivity with other populations render it vulnerable to disturbance, and diminishes its usefulness in supporting populations outside the reserve (Bell & Okamura 2005). By using connectivity information to build a network of MPAs, it may be possible to build resilience (Steneck *et al.* 2009). Restoration programmes may also benefit from considering molecular issues when selecting propagules. Higher levels of genetic diversity in *Z. marina* were shown to increase restoration success (Reynolds *et al.* 2012). Alternatively, genetically diverged populations may show local adaptation, requiring similar individuals to be selected for propagation to minimize any loss of fitness to the population based on mal-adapted organisms (Baums 2008). In order for successful management of marine resources, molecular techniques should be utilised where possible and appropriate. This will also give a more complete picture of marine biodiversity by showing levels of genetic diversity and its distribution among populations.

Marine sponges

Marine sponges (Phylum Porifera) are simple, sessile, filter-feeding animals found across the globe in a diverse range of ecosystems, from tropical reefs to polar regions, and shallow systems to the deep sea. In Caribbean coral reefs, sponges have a number of functional roles and are important components of the ecosystem, where they have a high biomass and species richness (Diaz & Rutzler 2001). Sponges have been shown to aid coral growth through their stabilising and consolidating effect on the substrate (Bell 2008). As filter feeders, they also have a significant impact on water quality and nutrient cycling in pelagic systems (Peterson *et al.* 2006; Fiore *et al.* 2013). They are also a food source for a number of reef-dwellers, including angelfish, parrotfish and hawksbill turtles (Wulff 1997, 2012). Numerous species provide a habitat within their body canals and cavities, supporting associated biodiversity (Frith 1976; Westinga & Hoetjes 1981; Gherardi *et al.* 2001; Padua *et al.* 2013). In addition, their body mass itself provides shelter and increases structural heterogeneity, providing habitats for other creatures such as fishes (Miller *et al.* 2012).

Due to their ecological interactions with other species and their effect on their environment, it is important to increase understanding of sponge ecology to ensure the sustainability of the ecosystems they are part of. This is especially important in areas where sponges are the dominant benthic life-form, where they can support communities through provision of nutrition and a three-dimensional habitat structure in a similar way to coral reefs (Schönberg & Fromont 2012). These 'sponge gardens' can also act as nursery grounds (Butler *et al.* 1995; Marliave *et al.* 2009), ensuring sustainability of populations in nearby ecosystems such as coral reefs. Furthermore, there are predictions that as coral cover decreases due to disturbance and climate change, in some areas phase-shifts may occur which turn coral reefs into 'sponge reefs' (Norström *et al.* 2009; Bell *et al.* 2013).

There have been some records of sponge communities suffering mass mortalities (Butler *et al.* 1995; Wulff 2006, 2013; Stevely *et al.* 2010). These have caused negative effects on both water quality (Peterson *et al.* 2006) and abundance in local populations of associated invertebrate species (Butler *et al.* 1995). In light of their importance to other organisms, and perhaps increasing

role as a dominant biogenic structure in reef habitats, understanding the ecology and dynamics of sponge populations becomes even more imperative. Molecular ecology may be an important tool in increasing this understanding. Molecular approaches could be used to better understand how connectivity and genetic diversity contribute to resilience and recovery of sponge populations in response to mortality events and population declines. In addition, molecular techniques could be used to show the degree to which genetic diversity in dominant sponge populations could contribute to associated biodiversity through community genetic effects. This increased understanding could contribute to management decisions, such as in restoration schemes and MPA design, and ultimately assist in conservation of sponge populations and the organisms they support.

Genetic markers

To study intraspecific genetic diversity and its distribution, ideally, genetic diversity would be measured across functionally significant loci to show ecologically and evolutionary relevant levels of diversity. However, identifying loci that are under selection, and determining the strength of that selection, is extremely challenging and is beyond the scope of most studies interested in genetic variation and its distribution. Population genetics models are therefore mostly based upon genetic markers. Simply defined by Sunnucks (2000) as 'heritable characters with multiple states at each character', molecular markers can be used to show genetic diversity in populations at a number of loci, which can then be used to address ecological questions. There are a number of different types of molecular markers, which all have their own strengths and weaknesses and usefulness depending on the study (see below), including the species being studied and the questions to be asked (Sunnucks 2000). In many cases, a molecular marker for use in intraspecies studies (rather than phylogeny or barcoding work) should be selectively neutral and be polymorphic among the population or species being studied to show variation. Neutrality of markers cannot be assumed (Van Oosterhout *et. al* 2004), and supposedly-neutral markers may be linked to traits which are under selection (Zytynska *et al.* 2012). These factors should be tested and their importance considered depending on the study in question. In addition, practical constraints and the economy of different markers are influential factors in any study.

Allozymes

In early studies, allozymes (polymorphic enzymatic proteins) were frequently used as population markers for sponges, including studies on the genetic diversity of populations (van Oppen 2002). Allozymes have many practical benefits, as they are relatively simple to use, cost-effective, give quick results, and can be used where there is no prior knowledge of the genome. They also tend to show high intraspecies variability in sponges (Uriz & Turon 2012). However, fresh or frozen (in liquid nitrogen) tissue is required which presents logistical problems in sample collection, electrophoretic gels can be difficult to analyse and compare between laboratories, and loci are not always neutral (van Oppen 2002; Uriz & Turon 2012). Due to these problems their use has been overtaken with genetic markers from the nuclear and mitochondrial genomes.

Mitochondrial DNA (mtDNA)

Sequences from the mitochondrial genome are often used as molecular markers to investigate phylogeny or population genetic structure. The cytochrome *c* oxidase subunit 1 (COI) has been

used to infer sponge phylogeny (Erpenbeck *et al.* 2007), and is currently being used in the Sponge Barcoding Project (Vargas *et al.* 2012). However, in sponges mtDNA shows an unusually slow rate of evolution and often has low variability within species (Duran *et al.* 2004b; Hoshino *et al.* 2008), and as a result its use in intraspecies work is generally discouraged (van Oppen 2002; Uriz & Turon 2012). Despite this, there are some instances where levels of intraspecific variability has been higher: DeBiasse *et al.* (2010) used the COI region to determine population structure in *Callyspongia vaginalis*, and Rua *et al.* (2011) found other mtDNA regions with higher nucleotide diversity than the COI region in a number of Demosponge species. Although this suggests that mtDNA variability depends on the species in question, and the region of the mitochondrial genome, time constraints may mean that other markers may be more reliable when extensive testing is not possible.

Amplified Fragment Length Polymorphism (AFLP)

AFLP (amplified fragment length polymorphism) markers are generated through enzyme restriction of genomic DNA followed by PCR amplification (Mueller & Wolfenbarger 1999). This technique holds many advantages; it is relatively easy and quick and produces hundreds of markers that are distributed through many DNA regions. However, AFLPs are most often used in bacteria, fungi and plant studies (Bensch & Åkesson 2005) and they are almost completely avoided in sponge work (Uriz & Turon 2012). Uriz & Turon (2012) advocate their use for sponge population genetics in combination with other markers. Difficulties in obtaining pure sponge DNA uncontaminated by bacterial or eukaryotic associate DNA mean that viable results are extremely difficult to obtain (van Oppen 2002). It is possible to obtain contaminant-free DNA – for example, Lopez *et al.* (2002) characterised AFLP markers in the sponge *Axinella corrugate* from DNA from a cell culture of the sponge cells – but these methods are time consuming and practically constraining, making markers with species-specific primers far easier to use when multiple samples are to be processed. In addition, AFLPs are dominant markers and therefore cannot distinguish between heterozygote and homozygote states, meaning that heterozygosity in a population cannot be measured (Mueller & Wolfenbarger 1999).

Nuclear DNA

Regions of nuclear DNA have been useful in sponge population genetic and phylogeographic studies. The *ITS* regions are popular and widely used, with PCR primers available for conserved regions which flank more variable regions (Uriz & Turon 2012). Although these markers have proved useful in studies on phylogeny and phylogeography (see van Oppen et al. 2002; Uriz & Turon 2012), there are some disadvantages to their use at the intraspecies population genetics level. These include low variability in the intraspecific level (Lopez *et al.* 2002; Bentlage & Wörheide 2007), although there is some promise for other nuclear regions which show higher levels of variability and so higher resolution (e.g., ALGII (Belinky *et al.* 2012), ATPSbeta-iII (Bentlage & Wörheide 2007)) however, these must be tested in other species to discover their utility across the Porifera.

Single Nucleotide Polymorphism (SNP)

SNPs are single-nucleotide substitutions that occur throughout the genome in both neutral regions and regions under selection. They are co-dominant, biallelic markers that have a range of applications, including in population genetics. However, their use is restricted by a lack of genomic data for non-model species, and questions surrounding the degree and type of selection loci are subject to and the level of variation among them (Helyar *et al.* 2011). In addition, analysis may be restricted due to the level of computer power and software capabilities required to process the large datasets needed in SNP analysis. To date, SNPs have not been used for the study of sponges, however in the future as technologies develop and more genomic information becomes available, they may be useful tools in sponge ecology (Uriz & Turon).

Microsatellites (SSR)

Microsatellites, or simple sequence repeats (SSR), are tandem repeats of up to 6 nucleotides in the nuclear genome. The number of repeat units commonly range from around 5 to 40, with the number of repeats (and therefore sequence length) corresponding to different alleles (Selkoe & Toonen 2006). Microsatellites are highly polymorphic and codominant (allowing heterozygotes and homozygotes to be distinguished), and most are selectivity neutral (which can and should be tested for during screening processes, see Selkoe & Toonen 2006). The characteristics make them well suited for use as a molecular marker and as a result, they are highly popular for population genetics work (Guichoux *et al.* 2011), and their use in sponge studies is advocated (Uriz & Turon 2012).

Microsatellites have been used in a variety of ways to study sponge populations. Population structure has been studied in multiple species at a range of spatial scales. Blanquer & Uriz (2010) detected genetic structure within populations, between populations in the same region, and between regions in *Scopalina lophyropoda,* using 7 microsatellite loci. Duran *et al.* (2004) also found genetic structure between and within sites at a large geographic scale in *Crambe crambe* using 6 microsatellites (this is in contrast to the low levels of structure found in the same species using mtDNA (Duran *et al.* 2004b)). Population structure at a small spatial scale has also been detected using microsatellites in *C.crambe* (Calderón *et al.* 2007) and *S. lophyropoda* (Blanquer *et al.* 2009). Microsatellites have also been used to study population genetic diversity of *Spongia officinalis* in the Mediterranean (Dailianis *et al.* 2011), to assess temporal as well as spatial structure in *Paraleucilla magna* populations in the northeast Iberean Peninsula (Guardiola *et al.* 2012) and to detect intraorganism genetic heterogeneity in *S. lophyropoda* (Blanquer & Uriz 2011). This range of studies shows the utility of microsatellites and their effectiveness in a range of species.

Microsatellite development

Traditionally, microsatellite development involves screening partial genomic libraries using repeat-containing probes. In brief, genomic DNA is fragmented, size-selected for small fragments, which are then ligated onto plasmids. Recombinant clones are then screened for repeat sequences using probes, and primers developed for PCR amplification of microsatellite loci (see Zane *et al*. (2002) for a detailed and comprehensive description of these traditional methods). To combat the problem of low numbers of microsatellites being obtained, enrichment techniques are commonly carried out

to increase the number of particular types of microsatellite motifs in libraries (Glenn & Schable 2005). Traditional methods have been successfully used to develop microsatellites in a range of species (Gardner 1999; Duran *et al.* 2002; Diniz *et al.* 2004; Lee *et al.* 2004), however, there are many disadvantages to this type of method. The process is costly and time-consuming, which may prevent some groups from utilising this type of marker. This type of method also gives a relatively low number of markers, which can sometimes be insufficient to address certain questions, such as genetic bottlenecking (Zalapa *et al.* 2012; Hoban *et al*. 2013).

Developments in sequencing technology have transformed the field of microsatellite isolation. Next-generation sequencing (NGS) allows millions of sequences to be read in parallel for a much cheaper per-base cost than traditional Sanger methods, and far more quickly. Programmes such as msatcommander (e.g., Sharma *et al*. 2012; Xu *et al*. 2012), QDD (Geismar & Nowak 2012; Wainwright *et al.* 2012), Tandem Repeat Finder (e.g., Nanninga et al. 2012) and PAL_FINDER (Castoe *et al.* 2012a; Stoutamore *et al.* 2012) are then used to find specified repeat motifs within the reads, with options available to design primers for suitable flanking regions. Roche's 454 Life Sciences (hereafter referred to as '454') use a pyrosequencing method (see Zapala *et al*. 2012 for a good overview of the sequencing process) to generate reads of 300-800bp in length, and a total of 350-720Mb of data per run. These reads are long enough to show a sufficient number of base pairs surrounding repeat motifs for effective primers to be designed for, allowing PCR amplification. Although libraries can be enriched before sequencing (Santana et al. 2009; Horreo et al. 2013; Gonzalez & Zardoya 2013), many studies do not carry out any prior enrichment, and successfully identify thousands of microsatellite loci (Abdelkrim *et al.* 2009; Gardner *et al.* 2011; Castoe *et al.* 2012b; Nanninga *et al.* 2012; Xu *et al.* 2012). Dubbed 'Seq-to-SSR' (Castoe *et al.* 2012a), this direct approach reduces the number of steps, and ultimately time and money, needed to isolate microsatellites. Due to the relatively lower cost and time required to find a large number of potential microsatellite loci, 454 has been referred to as the 'method of choice' for microsatellite development (Schoebel *et al.* 2013).

Illumina sequencing, which uses a 'sequencing-by-synthesis' method, has been utilised less for microsatellite development than 454. This is due to its shorter read lengths, which allow less scope for identifying repeat sequences plus adequate flanking regions. However, with the advancement of Illumina sequencing technology, it is now possible to identify microsatellites and flanking regions by sequencing from both ends (paired-end sequencing) of longer length reads. Castoe *et al.* (2012a) used Illumina paired end sequencing to successfully locate thousands of repeat motifs with adequate flanking regions (potentially amplifiable loci or PALs) in two birds and one snake species. Paired end sequencing was carried out on the GAIIx platform to obtain read lengths of 2 x 114bp (birds) and 2 x 116bp (snake). The programme PAL_FINDER was then used to search for microsatellite repeat motifs, PALs, and design suitable primers to amplify them in PCR reactions. The lowest number of PALs found in 5 million reads was 72,125 for a bird (Clark's nutcracker *Centrocercus minimus*), and the highest number was 174,370 for the snake (Burmese python *Python molurus bivittatus*).

Similar to 454, using Illumina sequencing is faster and gives a higher number of markers than traditional methods. However, Illumina offers the advantage over 454 of a much lower relative cost: $0.12- $0.39 per Mb read Illumina sequencing; compared to $8- $16 per Mb for 454 (Zalapa *et al.*

2012). This makes it an attractive and more economic option for microsatellite development. This relatively new approach has now been used successfully in a number of studies (Love *et al.* 2012; Nunziata *et al.* 2012; O'Bryhim *et al.* 2012, 2013; Stoutamore *et al.* 2012).

So far, studies using Illumina sequencing for microsatellite development have used the GAIIx platform or HiSeq, which offers maximum read lengths of 2 x 150pb. Illumina's latest sequencing platform, the MiSeq offers the option of yet longer reads (2 x 300bp). This offers the possibility of finding more PALs, which also gives a greater degree of freedom in selecting the best or most desired type of loci (see below). It may also allow more species to be included in a single Illumina run, reducing the per-species cost for microsatellite development. In addition, a 2 x 250bp run on the MiSeq takes approximately 41 hours (Illumina 2013), compared to 14 days on the GAIIx or 8.5 days on the HiSeq for 2 x 150bp runs, thus delivering results faster which may be important when there are time constraints on a project.

Optimising Seq-to-SSR methods

Although hundreds of potential markers can be found using NGS Seq-to-SSR approaches, the practicalities of the screening process means that the number of primers actually developed is far less. Zalapa *et al.* (2012) estimate that it is only possible to screen an average of 1% of possibly amplifiable loci found using NGS methods. Loci must be tested for their amplification success in PCR reactions and levels of polymorphism to discover their suitability as a genetic marker, a process which involves considerable costs and lab effort (Peery *et al.* 2013). In addition, success rates for the number of amplifiable, polymorphic loci found out of the loci screened vary widely among studies (see table 1). Although very high success rates have been experienced in some studies (Alfadala *et al.* 2012; Arcangeli *et al.* 2012, (95 and 100% of loci screened found to be amplifiable and polymorphic respectivly)), it is more commonly less than 60% (see table 1).

It may be possible to maximise this success rate by careful selection of loci and primer pairs. Firstly, longer repeat units generally show more polymorphism in populations, and many studies advocate picking >8 repeat units to yield greater allelic variability (Guichoux *et al.* 2011; Xu *et al.* 2012). Primers must also be carefully designed to maximise the chances of amplification success; for example by using primer design software such as primer3, which is often incorporated into microsatellite-locating software. It may be advisable to plan the PCR reagents and conditions that are to be used, and design primers that will work optimally for that particular reaction. For example, Qiagen recommend specific parameters in primer length, annealing temperature and GC content for use in its Type-it® Microsatellite PCR Kit, which is optimised for multiplex microsatellite PCR (Qiagen 2009).

It is also possible that sequencing errors, where bases are miscalled, could lead to unsuccessful amplification as primers may be unable to bind to flanking regions. Towards the end of reads, sequence quality tends to drop, however many studies do not carry out any filtering or trimming on their reads to remove low-quality areas before searching for microsatellites. This increases the risk of primers being designed for areas where bases have been miscalled, and subsequent PCR failure. By ensuring only high quality bases are included in reads, success rates in obtaining amplifiable loci could be increased.

Table 1: Success rates in obtaining amplifiable, polymorphic microsatellite loci from the loci screened in studies using NGS approaches to microsatellite development.

Studies included in this table found using the following search criteria in the ISI Web of Knowledge: 'Illumina microsatellite' and '454 microsatellites'. NB this is not intended as an exhaustive list of studies using NGS approaches to microsatellite development, but is intended as random selection to illustrate varied success rates in obtaining useful markers during the screening process.

Species	Phylum	Number of loci tested	Number of amplifiable, polymorphic loci	% of successful loci out of loci tested	Reference
Leptodea leptodon	Mollusca	48	16	33.3%	(O'Bryhim *et al.* 2012)
Rhinichthys osculus spp	Chordata	48	23	47.9%	(Nunziata *et al.* 2012)
Prosopium williamsoni	Chordata	48	22	45.8%	(O'Bryhim *et al.* 2013)
Paralithodes platypus	Arthropoda	48	23	47.9%	(Stoutamore *et al.* 2012)
Baetis alpinus	Arthropoda	12	5	41.6%	(Schoebel *et al.* 2013)
Stethophyma grossum	Arthropoda	50	10	20%	(Schoebel *et al.* 2013)
Fredericella sultana	Bryozoa	30	16	53.3%	(Schoebel *et al.* 2013)
Tlacuatzin canescens	Chordata	24	24	100%	(Arcangeli *et al.* 2012)
Batagur trivittata	Chordata	48	30	62.5%	(Love *et al.* 2012)
Mimulus ringens	Angiosperms	96	42	43.75%	(Nunziata & Karron 2012)
Lagopus muta	Chordata	22	9	40.9%	(Schoebel *et al.* 2013)
Eliomys quercinus	Chordata	12	7	58.3%	(Schoebel *et al.* 2013)
Amphiprion bicinctus	Chordata	100	40	40%	(Nanninga *et al.* 2012)
Silurus asotus	Chordata	70	47	67.1%	(Xu *et al.* 2012)
Acomys dimidiatus	Chordata	20	19	95%	(Alfadala *et al.* 2012)
Sebastes schlegelii	Chordata	30	17	56.6%	(Yasuike *et al.* 2013)

Aims and objectives

This study aims to develop a new set of microsatellite markers in the marine sponge *Cinachyrella alloclada*, which does not currently have any developed. *Cinachyrella alloclada* (Uliczka, 1929) is a Demosponge in the family Tetillidea. It has a wide distribution, found throughout the Caribbean, and is a common species in coral reef habitats. *C. alloclada* has been noted for its resistance to environmental disturbance (pers. comm M. Butler IV), making it an interesting candidate for genetic study. To do this we use a Seq-to-SSR approach, using the newer and previously unused Illumina MiSeq platform to yield 2 x 250bp reads. We also filter and trim the sequences to remove low-quality bases from reads, thus reducing the chance of errors in the flanking regions that could cause mis-priming and subsequent PCR failure. We test 35 PALs using PCR to identify microsatellites that can be used for population genetic and community genetics studies with *C. alloclada*, which will lead to a better understanding of this species ecology and conservation.

Methods and Results

i) Sample collection and DNA extraction

Ethics statement – Sponge samples were obtained from collaborators from the University of Florida and Old Dominion University, USA, and were collected using non-lethal methods. The samples were imported into the UK with permission from DEFRA under general license IMP/GEN/2010/01.

1.5cm^3 tissue samples from eight *Cinacyrella alloclada* individuals were collected by SCUBA diving, from Florida Bay, Florida, USA, and preserved on ice until processing later in the day. The samples were cut up into smaller pieces (~0.4cm^3) to aid preservation and to search for contaminating macroinvertebrates, and were then stored in 15ml 100% ethanol. The ethanol was changed after one day, and the samples were stored on dry ice during transport to Manchester, where the ethanol was changed again and they were stored at 4°C. DNA was extracted using the Qiagen® DNeasy DNA extraction kit (spin column method) following the manufacturers protocol. DNA was eluted in 200µl elution buffer and stored at 4°C until further use. DNA concentration and quality was tested using a NanoDrop 2000 spectrophotometer (Thermo Scientific) and a Quibit® fluorometer (Life Technologies), and by running 5ul of DNA solution on a 1.8% agarose gel. The DNA concentration was approximately 20ng/µl, and the bands on the gel were not fragmented or smeared, indicating high quality and non-degraded DNA.

ii) Library prep and sequencing

50ng of DNA from one individual was used to create a paired-end library using the Illumina Nextera® DNA Sample Preparation Kit according to the manufacturer's protocol, with adaptors used for identification purposes. Paired-end sequencing was carried out in half a flow cell lane of the Illumina MiSeq platform at the Genetics Core Facility at the University of Manchester, yielding a total of 3,481,868 (2 x 1,740,934) 250bp length paired end reads. The left and right raw reads were separated into two files, and converted to the FASTQ format.

iii) Quality filtering

Files containing the raw sequencing reads were imported into Galaxy-Golem, a customised local installation of the Pennsylvania State University's Galaxy, run by the University of Manchester's FLS Bioinformatics Core Facility. Galaxy is an open-access web-based bioinformatics tool which allows researchers to run a number of analyses on large data-sets, including NGS data. Quality of the raw reads was checked using the *FASTQC: ReadQC* tool (v0.51) (incorporated into Galaxy from the external programme FASTQC, developed by the Babraham Bioinformatics group). This tool checks the quality of raw data arising from high-throughput sequencing, and outputs information about the reads including lengths, quality scores and GC content. Raw reads were run through TRIMMOMATIC v0.30 (Loshe et al 2012), a quality-trimming tool for Illumina single- and paired-end reads. The 'sliding window' function was used, which removes reads where the average quality within a specified window falls below a specified quality threshold. A window size of 4bp were used, with a minimum mean quality threshold Phred score of 20 (99% base call accuracy). Minimum read length was set to 50, and the leading and trailing functions (which remove bases

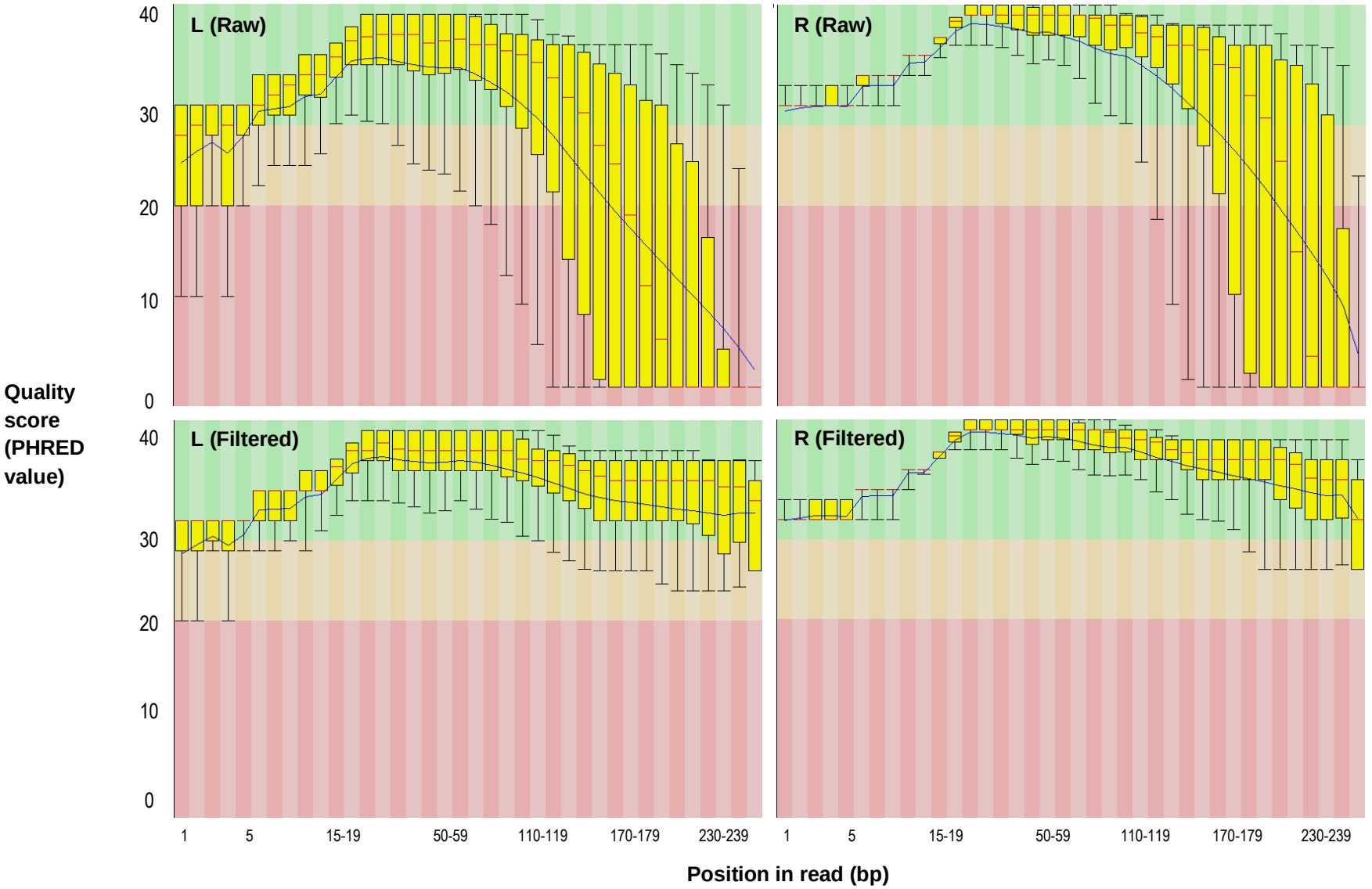

Figure 1: Comparison of per base sequence quality scores of the left (L) and right (R) reads in the raw reads and filtered reads.

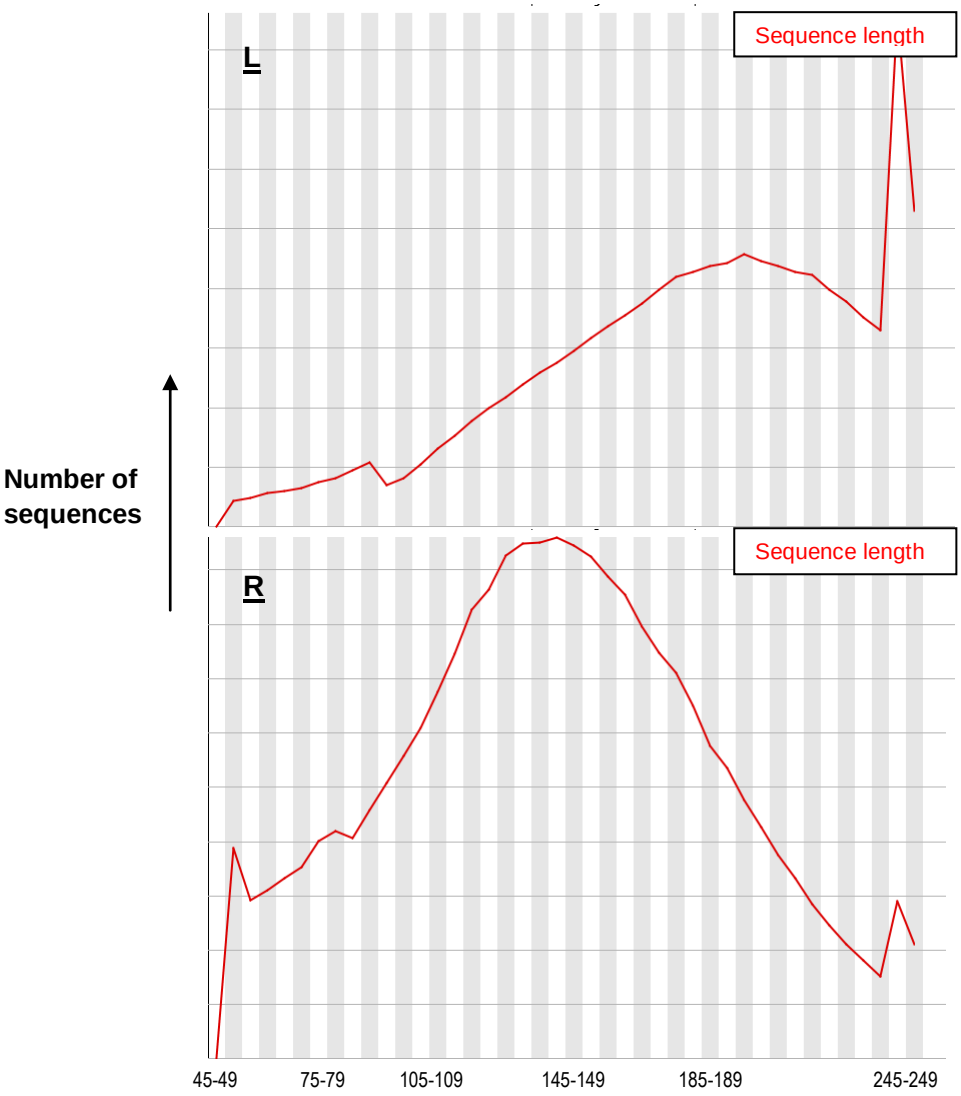

Figure 2: Sequence length distributions over all sequences in the left hand (L) and right hand (R) reads after quality trimming by TRIMMOMATIC

below a specified quality score at the start and end of reads respectively) were both set to 3. In total, 1,137,519 pairs survived, and the resulting output files containing the surviving left and right reads were then quality checked again with the *FASTQC: READQC* tool in Galaxy Golem (see figure 1). Sequence lengths ranged between 50 and 250bp for both the left and right hand reads, however, the sequence length distribution differed between the two (see figure 2), with a higher proportion of the right hand reads reaching longer sequence lengths of 240-250bp, compared to around 150bp for right hand reads.

iv) Microsatellite identification and primer design

Microsatellites with sufficient flanking regions (PALs) were identified using the programme PAL_FINDER v0.02 (Castoe *et al*. 2012a), which was incorporated into Galaxy-Golem for ease of data storage and computer processing power purposes. PAL_FINDER includes an installation of Primer3 v2.0.0 (Rozen et. al. 2000; Koressaar & Remm et al 2007), which designs primers for the microsatellite loci located by the programme. The configuration file was adjusted to screen for di (2-mer)- tri (3-mer)- tetra (4-mer)- penta (5-mer)- and hexa (6-mer) -nucleotide repeats with a minimum of 8 repeat units. Eight repeat units was chosen as a minimum as longer repeat stretches tend to be more polymorphic in populations (Castoe *et al.* 2012a; Xu *et al.* 2012).

Primer design parameters were set as follows: minimum melting temperature (*Tm*) 60°C; optimum *Tm* 68°C; maximum difference in *Tm* between primers in pair 2°C; length 20bp- 30bp; GC content 40-60%, and all other design settings were left to default. These parameters were chosen to work optimally with a specialised microsatellite PCR kit for multiplexing (Type-it® Microsatellite PCR kit (Qiagen)), in order to enhance amplification success in primer screening and future genotyping (Qiagen 2009). To test the effects of quality filtering of reads on the number and type of microsatellites found, the PAL_FINDER programme was run first with the raw, unfiltered reads, then again with the reads filtered by TRIMMOMATIC (see table 2).

Table 2: *Number of microsatellite loci [and number of PALs] found by PAL_FINDER in the raw reads vs. the quality filtered reads.*

Repeat type	Number of microsatellite loci [number of PALs]	
	Raw reads	**Filtered reads**
2-mer	75606 [30337]	22544 [6617]
3-mer	9082 [4618]	4715 [1883]
4-mer	6092 [2359]	2818 [680]
5-mer	416 [36]	206 [8]
6-mer	1399 [108]	662 [19]

The PALs and primer pairs obtained from the filtered file were further examined to select appropriate loci for screening. To ensure that the primers selected only amplify one locus, Castoe *et al*. (2012a) recommend selecting loci where both the forward and the reverse primer only occur once in the entire data set. After this filtering process, 1472 loci remained. Tri- and tetra- nucleotide repeats were chosen above di-nucleotide repeats, as these are more easily scored in genotyping (Castoe *et al.* 2012a). In addition, perfect repeat motifs were selected, because in compound and imperfect repeats a given length variant could correspond to a number of different alleles, whereas perfect repeats follow the stepwise mutation model more faithfully, which is assumed for many population genetic analyses (Guichoux *et al.* 2011). Based on these criteria, 35 loci (24 tri-nucleotide and 16 tetra-nucleotide) were chosen for screening.

vi) PCR amplification of selected loci

To check for amplification success, primers were tested on 8 individuals from Florida Bay, USA. Unlabelled oligonucleotide primers for were purchased at concentrations of 100µM from Sigma. Singleplex PCR reactions were carried out using the Type-it® Microsatellite PCR kit (Qiagen) in a reaction volume of 5µl, with a thermal cycle of: 95°C for 5 minutes, 28 x (95°C for 30 seconds, 60°C for 90 seconds, 72°C for 30 seconds) and 60°C for 30 minutes. 3µl of PCR product solution was checked on a 1.8% agarose gel against Hyperladder IV size standard (Bioline) to determine if PCR amplification was successful, and the approximate size range of fragments. A locus was classed as successfully amplified if at least 7 of the 8 samples tested showed one or two clear

bands on the gel (see figure 3a). This was observed for 13 of the 35 primer pairs tested (see table 3 for the repeat motifs and primer sequences). 'Unsuccessful' excluded loci were those which showed PCR product in none or few of the samples tested, or showed more than two bands for a single sample (indicative of non-specific priming) (see figure 3b).

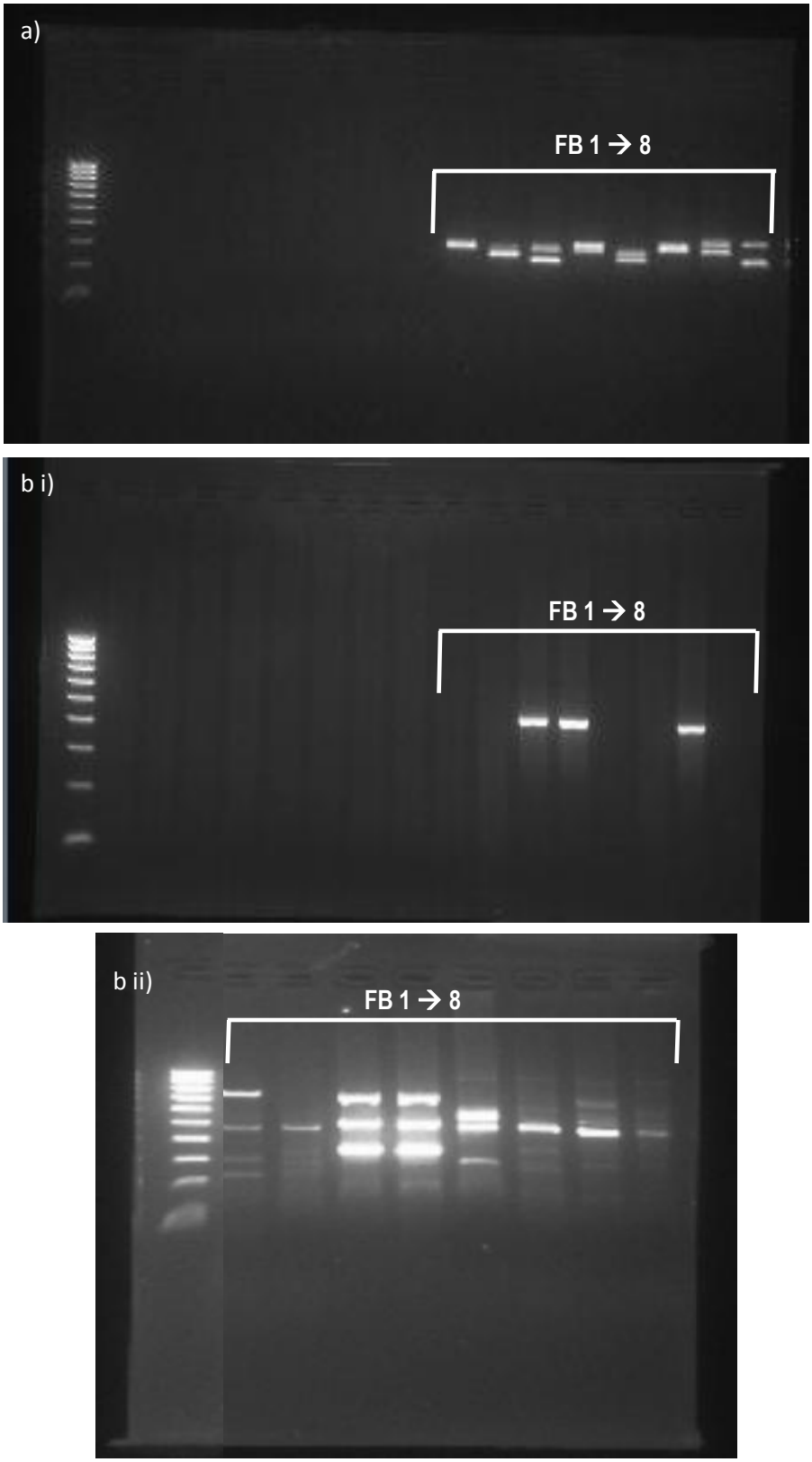

Figure 3: Agarose gel electrophoresis of PCR products for eight Florida Bay C. alloclada individuals (FB 1 – 8). Size standard = Hyperladder IV (Bioline)
Typical results for a) *a 'successful' locus (Call13)*
b) *'unsuccessful' loci due to* i) *Amplification failure (Call11)*
ii) *Non-specific priming (Call10)*

Table 3: *Primer sequences, repeat motifs and estimated allele sizes for successfully amplified microsatellite loci*

Locus	Repeat motif	Forward primer sequence	Reverse primer sequence	Estimated allele size range (bp)
Call2	TTC	CACTAGACATCATGGCAGGAAATGACC	CACGTATGTGGGCAGCTCGTTAGC	250-300
Call3	TTC	TAACAAAACCCCTTTGATCATCGCC	AAAGGTTTGGCAGACCCCATGTAGG	300-550
Call5	TTC	CACTTCAGCAAGTTTTGATGTAGGCTGG	TCCATAGGATTATACCATGACCAGAGAGCC	250-450
Call12	TTC	TCACCTGTATGTTGTCATAAAGTGGC	TCCACAAGTAAGGGTTTCAAAACAGG	450-900
Call13	TCTG	GAGGACAAGAGTATGGGGTCATGTGG	ATGGTTGTTTTGAACATGTGATGCC	100-250
Call14	TCTG	TCCACTTGTGTGTGTCTATGTCTGCC	CTGATGGAGGCCAAACACTTCCC	200-300
Call16	TTC	TGTTCTACTCTACTTTGCAGAATGTGCTCG	CCATCACAGGACTACCATCACAATG	250-450
Call21	TTC	TCACCTGTATGTTGTCATAAAGTGGC	TCCACAAGTAAGGGTTTCAAAACAGG	150-350
Call24	TCTG	GAGGACAAGAGTATGGGGTCATGTGG	ATGGTTGTTTTGAACATGTGATGCC	150-300
Call26	ATAC	CCATTTTGACCACTTCAAATTCTGTGC	ATCAGGCGTTGTTGAATCACTGAGC	500-550
Call29	TCTG	TCCACTTGTGTGTGTCTATGTCTGCC	CTGATGGAGGCCAAACACTTCCC	150-250
Call34	TGCG	GGAATATGTAGATACCAGCTAAACTGCTGC	TATGATTCCAAATCCATGACACAGG	150-250
Call36	TCTG	TGGTTCACTTCCTCTCATTTTGATGG	CTTCCCAAGGGAGCAAGTTTACAGG	150-400

Discussion

In this study, Illumina paired-end sequencing was successfully used to develop thirteen microsatellite markers for the marine sponge *Cinachyrella alloclada* that can be amplified using PCR. After quality filtering and trimming, 9207 PALs (with primers that adhered to the specified parameters) were located by the PAL_FINDER software. Out of the 35 that were tested, 13 were successfully amplified in PCR. These results add to the growing number of studies that use Illumina sequencing for the development of microsatellite markers (Arcangeli *et al.* 2012; Nunziata & Karron 2012; O'Bryhim *et al.* 2012, 2013; Sharma *et al.* 2012; Stoutamore *et al.* 2012), and supports claims that it is a rapid and relatively inexpensive way of yielding hundreds of potential markers.

This method has many advantages over traditional methods. A 2 x 250bp run on the Illumina MiSeq platform takes under 48 hours, and running TRIMMOMATIC and PAL_FINDER in Galaxy takes under 48 hours. Theoretically, primer testing could begin within a week of the initial sequencing, making this far quicker than the multiple steps used in traditional methods. The Illumina method is also cost effective: at the University of Manchester Genomic Technologies Core Facility, the cost per flow cell lane is £1500 for the MiSeq platform. Here, half a lane was used, costing the equivalent of £750 for the sequencing, and the bioinformatics programmes were free. Although primer testing has additional costs, this method is still very competitive compared to traditional methods carried out by commercial companies. In addition, the sequencing approach yields thousands of potential markers, giving a large amount of choice for researchers to select the type of microsatellite suitable for their needs.

The approach used in this study was based on the methods of Castoe *et al.* (2012a), however, here reads were filtered and trimmed to ensure all bases had a Phred score of above 20 (meaning 99% base call accuracy). This greatly reduced the number of reads used and the resulting number of microsatellite loci and PALs located by PAL_FINDER, with the number of microsatellite loci reduced by 61,650, and the number of PALs by 28,259. This means that without processing, a large number of PALs have been detected in reads which have a probability of error of over 1% (i.e., 1 bp in 100 miscalled), and figure 1 shows that the quality at the end of the raw reads held quality scores as low as 0. The risk in detecting microsatellites in lower-quality reads is that bases in priming sites may be miscalled, which could lead to primers not binding and unsuccessful PCR. As testing markers requires time and money, it is desirable to reduce the likelihood of loci unsuccessfully amplifying by carrying out filtering processes. It is impossible to tell from this study if filtering processes actually had a discernible effect on the success rate in testing loci. However, with so many loci available even post-filtering, there seems no disadvantage to doing so.

This study also differed from Castoe *et al.* (2012a) in the sequencer and read lengths used: here, sequencing was carried out on the MiSeq platform instead of the GAIIx platform; and 2 x 250bp reads were used instead of the 2 x 116bp/ 2 x 114bp reads in the original study. Longer read lengths are advantageous as it increases the likelihood of finding microsatellites with adequate flanking regions with which to design primers – 454 has been used more for microsatellite development than Illumina for this reason. Although quality trimming meant that the reads used

were actually between 50bp and 250bp, figure 2 shows that in the left hand read file, the highest number of reads fell at around 240bp, and the highest number of reads were around 150bp in the right hand file. This still provides longer read lengths overall and has the added advantage of all bases being above a quality threshold. As quality tends to drop at the end of reads, it is likely that reads of 116bp/ 114bp may have had low quality at the end, however, trimming here may have made the reads too short to provide adequate information. Although this cannot be tested in the present study, as a sequencing run was not simultaneously carried out to yield shorter reads, it is widely accepted that longer read lengths are better for microsatellite development. Here, longer read lengths can be used without the associated costs of 454 sequencing, giving less of a trade off between read length and cost.

To further determine the utility of these markers for assessing population genetic diversity, polymorphism levels should be checked in a number of individuals, which will show the variation present at each locus. It is possible to observe polymorphism in some of the markers tested here, where two bands per sample (i.e., a heterozygote) can be seen on a gel (e.g., see figure 3), or different size bands are present. However, by using fluorescently labelled primers, alleles can be scored and the precise length of the microsatellites can be determined, giving a more accurate and useful measure of polymorphism levels. For example, when allele lengths may only differ by a few base pairs, a standard small gel will not show any detectable difference in the position of the bands. Loci should also be checked for null alleles, linkage disequilibrium and neutrality to ensure that they abide by the assumptions with which microsatellite analyses are conducted (Selkoe & Toonen 2006). By carrying out such screening processes, the resultant set of markers can be applied to address a range of ecological questions, such as determining spatial and temporal population genetic structure, and the effects of population genetic diversity on associated communities and population resilience.

Although it is agreed that the study of sponge ecology would be enhanced by the availability of microsatellite markers, the cost and time constraints of traditional methods have acted as a significant barrier (Uriz & Turon 2012). This study has shown that Illumina sequencing is an effective method for developing markers in a more rapid and cost effective manner, as well as allowing a greater choice in the type of microsatellites to be developed. We have modified an existing method (Castoe *et al.* 2012a) to enhance it by increasing read length and quality with the aim of increasing the number of microsatellites that can be found and the chances of amplification success. Moreover, we successfully developed 13 microsatellites for *Cinachyrella alloclada,* which previously had no genetic markers developed.

References

Abdelkrim, J., Robertson, B., Stanton, J.-A. & Gemmell, N. (2009). Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *BioTechniques*, 46, 185–92.

Alfadala, S., Dawson, D.A., Horsburgh, G.J., Behnke, J.M., Bajer, A., Mohallal, E.M.E., *et al.* (2012). Large-scale isolation of Eastern spiny mouse Acomys dimidiatus microsatellite loci through GS-FLX 454 titanium sequencing. *Conservation Genetics Resources*, 5, 519–524.

Almany, G.R., Connolly, S.R., Heath, D.D., Hogan, J.D., Jones, G.P., McCook, L.J., *et al.* (2009). Connectivity, biodiversity conservation and the design of marine reserve networks for coral reefs. *Coral Reefs*, 28, 339–351.

Arcangeli, J., Cervantes, F.A., Lance, S.L., Salazar, M.I. & Ortega, J. (2012). Twenty-four microsatellite markers for the gray mouse opossum (*Tlacuatzin canescens*): development from Illumina paired-end sequences. *Conservation Genetics Resources*, 5, 367–370.

Baums, I.B. (2008). A restoration genetics guide for coral reef conservation. *Molecular Ecology*, 17, 2796–2811.

Belinky, F., Szitenberg, A., Goldfarb, I., Feldstein, T., Wörheide, G., Ilan, M., *et al.* (2012). ALG11-a new variable DNA marker for sponge phylogeny: comparison of phylogenetic performances with the 18S rDNA and the COI gene. *Molecular Phylogenetics and Evolution*, 63, 702–13.

Bell, J.J. (2008). The functional roles of marine sponges. *Estuarine, Coastal and Shelf Science*, 79, 341–353.

Bell, J.J., Davy, S.K., Jones, T., Taylor, M.W. & Webster, N.S. (2013). Could some coral reefs become sponge reefs as our climate changes? *Global Change Biology*, 19, 2613–24.

Bell, J.J. & Okamura, B. (2005). Low genetic diversity in a marine nature reserve: re-evaluating diversity criteria in reserve design. *Proceedings of The Royal Society B: Biological Sciences*, 272, 1067–74.

Bensch, S. & Åkesson, M. (2005). Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology*, 14, 2899–2914.

Bentlage, B. & Wörheide, G. (2007). Low genetic structuring among *Pericharax heteroraphis* (Porifera: Calcarea) populations from the Great Barrier Reef (Australia), revealed by analysis of nrDNA and nuclear intron sequences. *Coral Reefs*, 26, 807–816.

Bernhardt, J.R. & Leslie, H.M. (2013). Resilience to climate change in coastal marine ecosystems. *Annual Review of Marine Science*, 5, 371–92.

Blanquer, A. & Uriz, M. (2011). “Living together apart”: The hidden genetic diversity of sponge populations. *Molecular Biology Evolution*, 28, 2435–2438.

Blanquer, A., Uriz, M. & Caujapé-Castells, J. (2009). Small-scale spatial genetic structure in Scopalina lophyropoda, an encrusting sponge with philopatric larval dispersal and frequent fission and fusion events. *Marine Ecology Progress Series*, 380, 95–102.

Blanquer, A. & Uriz, M.J. (2010). Population genetics at three spatial scales of a rare sponge living in fragmented habitats. *BMC Evolutionary Biology*, 10.

Botsford, L.W., Micheli, F. & Hastings, A. (2003). Principles for the design of marine reserves. *Ecological Applications*, 13, 25–31.

Botsford, L.W., White, J.W., Coffroth, M.-A., Paris, C.B., Planes, S., Shearer, T.L., *et al.* (2009). Connectivity and resilience of coral reef metapopulations in marine protected areas: matching empirical efforts to predictive needs. *Coral Reefs*, 28, 327–337.

Butler, M., Hunt, J., Herrnkind, W., Childress, M., Bertelsen, R., Sharp, W., *et al.* (1995). Cascading disturbances in Florida Bay, USA: cyanobacteria blooms, sponge mortality, and implications for juvenile spiny lobsters *Panulirus argu*s. *Marine Ecology Progress Series*, 129, 119–125.

Calderón, I., Ortega, N., Duran, S., Becerro, M., Pascual, M. & Turon, X. (2007). Finding the relevant scale : clonality and genetic structure in a marine invertebrate (*Crambe crambe*, Porifera). *Molecular Ecology*, 16, 1799–1810.

Castoe, T., Poole, A., de Koning, A., Jones, K.L., Tomback, D.F., Oyler-McCance, S.J. *et al.* (2012a). Rapid microsatellite identification from Illumina paired-end genomic sequencing in two birds and a snake. *PLoS ONE*, 7.

Castoe, T.A., Streicher, J.W., Meik, J.M., Ingrasci, M.J., Poole, A.W., de Koning, A.P.J., *et al.* (2012b). Thousands of microsatellite loci from the venomous coralsnake *Micrurus fulvius* and variability of select loci across populations and related species. *Molecular Ecology Resources*, 12, 1105–13.

Christie, M.R., Tissot, B.N., Albins, M. A, Beets, J.P., Jia, Y., Ortiz, D.M., *et al.* (2010). Larval connectivity in an effective network of marine protected areas. *PloS ONE* 5, e15715.

Cooley, S.R., Kite-Powell, H.L. & Doney, S.C. (2006). Ocean acidification's potential to alter global marine ecosystem services. *Oceanography*, 22, 172–181.

Cowen, R. & Sponaugle, S. (2009). Larval dispersal and marine population connectivity. *Annual Review of Marine Science*, 1, 443–466.

Dailianis, T., Tsigenopoulos, C.S., Dounas, C. & Voultsiadou, E. (2011). Genetic diversity of the imperilled bath sponge *Spongia officinalis* Linnaeus, 1759 across the Mediterranean Sea: patterns of population differentiation and implications for taxonomy and conservation. *Molecular Ecology*, 20, 3757–72.

DeBiasse, M.B., Richards, V.P. & Shivji, M.S. (2010). Genetic assessment of connectivity in the common reef sponge, *Callyspongia vaginalis* (Demospongiae: Haplosclerida) reveals high population structure along the Florida reef tract. *Coral Reefs*, 29, 47–55.

Diaz, M. & Rutzler, K. (2001). Sponges: an essential component of Caribbean coral reefs. *Bulletin of Marine Science*, 69, 535–546.

Diniz, F.M., Maclean, N., Paterson, I.G. & Bentzen, P. (2004). Polymorphic tetranucleotide microsatellite markers in the Caribbean spiny lobster, *Panulirus argus*. *Molecular Ecology Notes*, 4, 327–329.

Duran, S., Pascual, M., Estoup, A. & Turon, X. (2002). Polymorphic microsatellite loci in the sponge *Crambe crambe* (Porifera: Poecilosclerida) and their variation in two distant populations. *Molecular Ecology Notes*, 2, 478–480.

Duran, S., Pascual, M., Estoup, A. & Turon, X. (2004a). Strong population structure in the marine sponge Crambe crambe (Poecilosclerida) as revealed by microsatellite markers. *Molecular Ecology*, 13, 511–522.

Duran, S., Pascual, M. & Turon, X. (2004b). Low levels of genetic variation in mtDNA sequences over the western Mediterranean and Atlantic range of the sponge *Crambe crambe* (Poecilosclerida). *Marine Biology*, 144, 31–35.

Ehlers, A., Worm, B. & Reusch, T.B.H. (2008). Importance of genetic diversity in eelgrass Zostera marina for its resilience to global warming. *Marine Ecology Progress Series*, 355, 1–7.

Erpenbeck, D., Duran, S., Rützler, K., Paul, V., Hooper, J.N. a. & Wörheide, G. (2007). Towards a DNA taxonomy of Caribbean demosponges: a gene tree reconstructed from partial mitochondrial CO1 gene sequences supports previous rDNA phylogenies and provides a new perspective on the systematics of Demospongiae. *Journal of the Marine Biological Association of the UK*, 87.

Féral, J. (2002). How useful are the genetic markers in attempts to understand and manage marine biodiversity? *Journal of Experimental Marine Biology and Ecology*, 268, 121–145.

Fiore, C.L., Baker, D.M. & Lesser, M.P. (2013). Nitrogen biogeochemistry in the Caribbean sponge, *Xestospongia muta*: A Source or Sink of Dissolved Inorganic Nitrogen? *PloS ONE*, 8, e72961.

Frith, D. (1976). Animals associated with sponges at North Hayling, Hampshire. *Zoological Journal of the Linnean Society*, 58, 353–362.

Gardner, M.G. (1999). Brief communication. Isolation of microsatellite loci from a social lizard, Egernia stokesii, using a modified enrichment procedure. *Journal of Heredity*, 90, 301–304.

Gardner, M.G., Fitch, A.J., Bertozzi, T. & Lowe, A.J. (2011). Rise of the machines--recommendations for ecologists when using next generation sequencing for microsatellite development. *Molecular Ecology Resources*, 11, 1093–101.

Geismar, J. & Nowak, C. (2012). Isolation and characterisation of new microsatellite markers for the stonefly Brachyptera braueri comparing a traditional approach with high throughput 454 sequencing. *Conservation Genetics Resources*, 5, 413–416.

Gherardi, M., Giangrande, A., & Corriero, G. (2001). Epibiontic and endobiontic polychaetes of *Geodia cydonium* (Porifera , Demospongiae) from the Mediterranean Sea, *Hydrobiologica,* 443, 87–101.

Glenn, T.C. & Schable, N.A. (2005). Isolating microsatellite DNA loci. *Methods in Enzymology*, 395, 202–22.

Gonzalez, E.G. & Zardoya, R. (2013). Microsatellite DNA capture from enriched libraries. *Methods in Molecular Biology*, 1006, 67–87.

Guardiola, M., Frotscher, J. & Uriz, J. (2012). Genetic structure and differentiation at a short-time scale of the introduced calcarean sponge *Paraleucilla magna* to the western Mediterranean. *Hydrobiologia*, 687, 71–84.

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., *et al.* (2011). Current trends in microsatellite genotyping. *Molecular Ecology Resources*, 11, 591–611.

Halpern, B.S., Walbridge, S., Selkoe, K.A., Kappel, C. V, Micheli, F., D'Agrosa, C., *et al.* (2008). A global map of human impact on marine ecosystems. *Science*, 319, 948–952.

Hellberg, M., Burton, R., Neigel, J. & Palumbi, S.R. (2002). Genetic assessment of connectivity among marine populations. *Bulletin of Marine* , 70, 273–290.

Helyar, S.J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M.I., Ogden, R., Limborg, M.T. *et al.* (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources*, 11 (Suppl. 1), 123-136.

Hoban, S.M., Gaggiotti, O.E. & Bertorelle, G. (2013). The number of markers and samples needed for detecting bottlenecks under realistic scenarios, with and without recovery: a simulation-based study. *Molecular Ecology*, 22, 3444–3450.

Hoegh-Guldberg, O., Mumby, P.J., Hooten, A.J., Steneck, R.S., Greenfield, P., Gomez, E., *et al.* (2007). Coral reefs under rapid climate change and ocean acidification. *Science*, 318, 1737–42.

Horreo, J.L., Alonso, J.C., Palacín, C. & Milá, B. (2013). Identification of polymorphic microsatellite loci for the endangered great bustard (*Otis tarda*) by high-throughput sequencing. *Conservation Genetics Resources*, 5, 549–551.

Hoshino, S., Saito, D.S. & Fujita, T. (2008). Contrasting genetic structure of two Pacific Hymeniacidon species. *Hydrobiologia*, 603, 313–326.

Hughes, A., Inouye, B., Johnson, M.T.J., Underwood, N. & Vellend, M. (2008). Ecological consequences of genetic diversity. *Ecology Letters*, 11, 609–623.

Hughes, A.R. & Stachowicz, J.J. (2004). Genetic diversity enhances the resistance of a seagrass ecosystem to disturbance. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 8998–9002.

Hughes, T.P., Baird, a H., Bellwood, D.R., Card, M., Connolly, S.R., Folke, C., *et al.* (2003). Climate change, human impacts, and the resilience of coral reefs. *Science*, 301, 929–33.

Illumina (2013) MiSeq benchtop sequencer performance specifications [online]. Available at <http://www.illumina.com/systems/miseq/performance_specifications.ilmn> [Accessed 4/8/2013].

Johnson, M., Lajeunesse, M. & Agrawal, A. (2006). Additive and interactive effects of plant genotypic diversity on arthropod communities and plant fitness. *Ecology Letters*, 9, 24–34.

Johnston, E.L. & Roberts, D.A. (2009). Contaminants reduce the richness and evenness of marine communities: a review and meta-analysis. *Environmental Pollution,* 157, 1745–52.

Knowlton, N, Brainard RE, Fisher R, Moews M, Plaisance L & Caley MJ (2010) Coral reef biodiversity. In: A McIntyre (ed) *Life in the World's Oceans: Diversity, Distribution, and Abundance.* Blackwell Publishing Ltd. Ch.4.

Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics* (23), 1289-91.

Lee, S.L., Tani, N., Ng, K.K.S. & Tsumura, Y. (2004). Isolation and characterization of 20 microsatellite loci for an important tropical tree *Shorea leprosula* (Dipterocarpaceae) and their applicability to S. parvifolia. *Molecular Ecology Notes*, 4, 222–225.

Lopez, J. V, Peterson, C.L., Willoughby, R., Wright, A.E., Enright, E., Zoladz, S., *et al.* (2002). Characterization of genetic markers for in vitro cell line identification of the marine sponge *Axinella corrugata*. *The Journal of Heredity*, 93, 27–36.

Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, Usadel B & Robi NA (2012) A user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res* (40) (Web Server issue), W622-7.

Love, C.N., Hagen, C., Horne, B.D., Jones, K.L. & Lance, S.L. (2012). Development and characterization of thirty novel microsatellite markers for the critically endangered Myanmar Roofed Turtle, *Batagur trivittata*, and cross-amplification in the Painted River Terrapin, *B. borneoensis*, and the Southern River Terrapin, *B. affins*, using paired-end Illumina shotgun sequencing. *Conservation Genetics Resources*, 5, 383–387.

Marliave, J., Conway, K., Gibbs, D., Lamb, A. & Gibbs, C. (2009). Biodiversity and rockfish recruitment in sponge gardens and bioherms of southern British Columbia, Canada. *Marine Biology*, 156, 2247–2254.

Miller, R.J., Hocevar, J., Stone, R.P. & Fedorov, D. V. (2012). Structure-forming corals and sponges and their use as fish habitat in Bering Sea submarine canyons. *PloS ONE*, 7, e33885.

Moberg, F. & Folke, C. (1999). Ecological goods and services of coral reef ecosystems. *Ecological Economics*, 29, 215–233.

Mueller, U.G. & Wolfenbarger, L.L. (1999). AFLP genotyping and fingerprinting. *Trends in Ecology & Evolution*, 14, 389–394.

Nanninga, G., Mughal, M., Saenz-Agudelo, P., Bayer, T. & Berumen, M. (2012). Development of 35 novel microsatellite markers for the two-band anemonefish *Amphiprion bicinctus*. *Conservation Genetics Resources*, 5, 515–518.

Norström, A., Nyström, M., Lokrantz, J. & Folke, C. (2009). Alternative states on coral reefs: beyond coral-macroalgal phase shifts. *Marine Ecology Progress Series*, 376, 295–306.

Nunziata, S., Karron, J., Mitchell, R.J., Lance, S.L., Jones, K.L. & Trapnell D.W. (2012). Characterization of 42 polymorphic microsatellite loci in *Mimulus ringens* (Phrymaceae) using Illumina sequencing. *American Journal of Botany*.

Nunziata, S.O., Lance, S.L., Jones, K.L., Nerkowski, S.A. & Metcalf, A.E. (2012). Development and characterization of twenty-three microsatellite markers for the freshwater minnow Santa Ana speckled dace (*Rhinichthys osculus spp*., Cyprinidae) using paired-end Illumina shotgun sequencing. *Conservation Genetics Resources*, 5, 145–148.

O'Bryhim, J., Chong, J.P., Lance, S.L., Jones, K.L. & Roe, K.J. (2012). Development and characterization of sixteen microsatellite markers for the federally endangered species: *Leptodea leptodon* (Bivalvia: Unionidae) using paired-end Illumina shotgun sequencing. *Conservation Genetics Resources*, 4, 787–789.

O'Bryhim, J., Somers, C., Lance, S.L, Yau, M., Boreham, D.R., Jones, K.L. & Taylor, E.B. (2012). Twenty-two novel microsatellite markers for the mountain whitefish, *Prosopium williamsoni* and cross-amplification in the round whitefish, *P. cylindraceum*, using paired. *Conservation Genetics Resources.*

Van Oppen, M. (2002). Nuclear markers in evolutionary and population genetic studies of scleractinian corals and sponges. *Proceedings of the Ninth International Coral Reef Symposium*, 1, 131–138.

Padua, A., Lanna, E. & Klautau, M. (2013). Macrofauna inhabiting the sponge *Paraleucilla magna* (Porifera: Calcarea) in Rio de Janeiro, Brazil. *Journal of the Marine Biological Association of the UK*, 93, 889–898.

Palumbi, S. (2003). Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications*, 13, 146–158.

Peery, M.Z., Reid, B.N., Kirby, R., Stoelting, R., Doucet-Bëer, E., Robinson, S., *et al.* (2013). More precisely biased: increasing the number of markers is not a silver bullet in genetic bottleneck testing. *Molecular Ecology*, 22, 3451–3457.

Peterson, B., Chester, C., Jochem, F. & Fourqurean, J.W. (2006). Potential role of sponge communities in controlling phytoplankton blooms in Florida Bay. *Marine Ecology Progress Series*, 328, 93–103.

Qiagen (2009) Type-it Microsatellite PCR Handbook [pdf] Available at <http://www.qiagen.com/Products/Catalog/Assay-Technologies/End-Point-PCR-and-RT-PCR-Reagents/Type-it-Microsatellite-PCR-Kit#resources> [Accessed 10/11/12].

Reusch, T.B.H., Ehlers, A., Hämmerli, A. & Worm, B. (2005). Ecosystem recovery after climatic extremes enhanced by genotypic diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 2826–31.

Reynolds, L., McGlathery, K. & Waycott, M. (2012). Genetic diversity enhances restoration success by augmenting ecosystem services. *PloS ONE*, 7, 1–7.

Rua, C.P.J., Zilberberg, C. & Solé-Cava, A.M. (2011). New polymorphic mitochondrial markers for sponge phylogeography. *Journal of the Marine Biological Association of the United Kingdom*, 91, 1015–1022.

Santana, Q., Coetzee, M., Steenkamp, E., Mlonyeni, O., Hammond, G., Wingfield, M., *et al.* (2009). Microsatellite discovery by deep sequencing of enriched genomic libraries. *BioTechniques*, 46, 217–23.

Schoebel, C.N., Brodbeck, S., Buehler, D., Cornejo, C., Gajurel, J., Hartikainen, H., *et al.* (2013). Lessons learned from microsatellite development for nonmodel organisms using 454 pyrosequencing. *Journal of Evolutionary Biology*, 26, 600–11.

Schönberg, C. & Fromont, J. (2012). Sponge gardens of Ningaloo Reef (Carnarvon Shelf, Western Australia) are biodiversity hotspots. *Hydrobiologia*, 687, 143–161.

Selkoe, K.A. & Toonen, R.J. (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, 9, 615–29.

Sharma, R., Goossens, B., Kun-Rodrigues, C., Teixeira, T., Othman, N., Boone, J.Q., *et al.* (2012). Two different high throughput sequencing approaches identify thousands of de novo genomic markers for the genetically depleted Bornean elephant. *PloS ONE*, 7, e49533.

Steneck, R.S., Paris, C.B., Arnold, S.N., Ablan-Lagman, M.C., Alcala, a. C., Butler, M.J., *et al.* (2009). Thinking and managing outside the box: coalescing connectivity networks to build region-wide resilience in coral reef ecosystems. *Coral Reefs*, 28, 367–378.

Stevely, J.M., Sweat, D.E., Bert, T.M., Sim-smith, C. & Kelly, M. (2010). Sponge Mortality at Marathon and Long Key , Florida : Patterns of Species Response and Population Recovery. *Proceedings of the Gulf and Caribbean Fisheries Institute,* 63, 384-400.

Stoutamore, J.L., Love, C.N., Lance, S.L., Jones, K.L. & Tallmon, D. (2012). Development of polymorphic microsatellite markers for blue king crab (Paralithodes platypus). *Conservation Genetics Resources*, 4, 897–899.

Sunnucks, P. (2000). Efficient genetic markers for population biology. *Trends in Ecology & Evolution*, 15, 199–203.

Uriz, M.J. & Turon, X. (2012). Sponge ecology in the molecular era. *Advances in Marine Biology*, 61, 345-410.

van Oosterhaut, C., van Heuven, M. & Brakefield, P. (2004) On the neutrality of genetic markers: pedegree analysis of genetic variation in fragmented populations. *Molecular Ecology*, 13, 1025-1034.

Vargas, S., Schuster, A., Sacher, K., Büttner, G., Schätzle, S., Läuchli, B., *et al.* (2012). Barcoding sponges: an overview based on comprehensive sampling. *PloS ONE*, 7, e39345.

Wainwright, B.J., Arlyza, I.S. & Karl, S.A. (2012). Isolation and characterization of twenty-six, polymorphic microsatellite loci for the tropical sea urchin, *Colobocentrotus atratus*. *Conservation Genetics Resources*, 5, 379–381.

Wainwright, B.J., Arlyza, I.S. & Karl, S.A. (2013). Isolation and characterization of twenty-three polymorphic microsatellite loci for Linckia laevigata. *Conservation Genetics Resources*, 5, 157–159.

Westinga, E. & Hoetjes, P.C. (1981). The intrasponge fauna of *Spheciospongia vesparia* (Porifera, Demospongiae) at Curagao and Bonaire, *Marine Biology*, 150, 139–150.

Whitham, T.G., Gehring, C. a, Lamit, L.J., Wojtowicz, T., Evans, L.M., Keith, A.R., *et al.* (2012). Community specificity: life and afterlife effects of genes. *Trends in Plant Science*, 17, 271–81.

Wittmann, A.C. & Pörtner, H.-O. (2013). Sensitivities of extant animal taxa to ocean acidification. *Nature Climate Change*, doi:10.1038/nclimate1982.

Worm, B., Barbier, E.B., Beaumont, N., Duffy, J.E., Folke, C., Halpern, B.S., *et al.* (2006). Impacts of biodiversity loss on ocean ecosystem services. *Science*, 314, 787–90.

Worm, B., Hilborn, R., Baum, J.K., Branch, T. A , Collie, J.S., Costello, C., *et al.* (2009). Rebuilding global fisheries. *Science*, 325, 578–85.

Wulff, J. (2006). Rapid diversity and abundance decline in a Caribbean coral reef sponge community. *Biological Conservation*, 127, 167–176.

Wulff, J. (2012). Ecological Interactions and the distribution, abundance, and diversity of sponges. *Advances in Marine Biology*, 61, 273–344.

Wulff, J. (2013). Integrative and comparative biology recovery of sponges after extreme mortality events : morphological and taxonomic patterns in regeneration versus recruitment, *Integrative and Comparitive Biology,* 1–12.

Wulff, J.L. (1997). Parrotfish predation on cryptic sponges of Caribbean coral reefs. *Marine Biology*, 129, 41–52.

Xu, D., Zhang, Y. & Peng, Z. (2012). High-throughput microsatellite marker development in Amur catfish (*Silurus asotus*) using next-generation sequencing. *Conservation Genetics Resources*, 5, 487–490.

Yasuike, M., Noda, T., Fujinami, Y. & Sekino, M. (2013). Tri-, tetra- and pentanucleotide-repeat microsatellite markers for the Schlegel's black rockfish *Sebastes schlegelii*: the potential for reconstructing parentages. *Conservation Genetics Resources*, 5, 577–581.

Zalapa, J.E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E., *et al.* (2012). Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *American Journal of Botany*, 99, 193–208.

Zane, L., Bargelloni, L. & Patarnello, T. (2002). Strategies for microsatellite isolation: a review. *Molecular Ecology*, 11, 1–16.

Zytynska, S.E., Fay, M.F., Penney, D. & Preziosi, R.F. (2011). Genetic variation in a tropical tree species influences the associated epiphytic plant and invertebrate communities in a complex forest ecosystem. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 366, 1329–36.

Zytynska, S.E., Khudr, M.S., Harris, E. & Preziosi, R.F. (2012). Genetic effects of tank-forming bromeliads on the associated invertebrate community in a tropical forest ecosystem. *Oecologia*, 170, 467–75.