Identification, Generation and Validation of Endogenous *Symbiodinium microadriaticum P*romoters and Promoter Regions to Drive Exogenous Gene Expression

Romain Theo Yann Landolfini

School of Bioscience, University of Kent

Thesis for the MSc in Biotechnology and Bioengineering

2017/2018

Supervisors: Mark Smales & James Budge

Abstract

Tropical reef ecosystems worldwide are suffering due to anthropogenic CO_2 which is leading to global climate change and rising sea surface temperatures. This damage is caused by a breakdown of the symbiosis between the coral and the endosymbiotic dinoflagellate algae (*Symbiodinium*) living inside them. This results from the photoinhibition of photosynthesis through damage to photosystem II which results in the production of reactive oxygen species, destructive to both the *Symbiodinium* and the coral cell. Studies have shown that there is a thermally tolerant phenotype that exists within clades of *symbiodinium* which has led researchers to believe that genetically engineering the endosymbionts as a conservation strategy for corals across the globe may be a legitimate avenue to test. In this study we create a custom expression construct for *Symbiodinium* made up of a pcDNA™ 3.1/Hygro(+) vector, an mCherry gene with a V5-His tag and putative endogenous *Symbiodinium microadriaticum* promoters. We were successful in creating the majority of the construct however, sequencing results showed that our putative promoter regions did not match the sequences we expected to be isolated. Although these are not the results we expected they do not impede the application of a custom expression construct for *Symbiodinium* transformations.

Introduction

Coral reefs all around the word have suffered major declines, principally believed to be due to anthropogenic sources (Glynn, 1996; Wilkinson, 2004; Hughes T. P. et al., 2003; Bruno J. F. et al., 2007; De'ath G. et al., 2012). A growing body of evidence suggests that global climate change is the most impactful threat to the world's reef ecosystems and the populations that depend on them (Glynn P. W., 1984; Baker A. C. et al., 2008; Spalding M. D. et al., 2015; Heron S. F. et al., 2016). Recent global warming trends have resulted in rising sea surface temperatures (SST) which trigger the bleaching of the corals living in the reef (Heron S. F. et al., 2016; Hughes T. P. et al., 2017). Coral bleaching refers to the loss of pigmentation from the corals as a consequence of the photoinhibition of photosynthesis and photodamage to the D1 protein in photosystem II (PSII) of the endosymbiotic dinoflagellate resulting in the production of reactive oxygen species (ROS) by transport chain electrons (Hoegh-Guldberg O., 1999; Warner M. E., 1999; Lesser M. P. et al., 2004; Lesser M. P., 2006; DeSalvo M. K. et al., 2008). This disrupts the symbiosis between corals and their endosymbiotic dinoflagellate algae (*Symbiodinium spp.*). This is particularly destructive to the corals as they owe their success to the microalgae that populate their tissues; these provide around 90% of the coral's nutritional requirements in the form of photosynthetically derived metabolites (Muscatine L. et al., 1977; Trench R. K., 1979; Muscatin L., 1990). The loss of the *Symbiodinium* during extended periods of time often leads to coral death (Glynn P. W., 1984; Goreau T. et al., 2000; Berkelmans R. et al., 2004; Eakin C. M. et al., 2016, Hughes T. P. et al., 2017). It is therefore believed that maintaining or promoting a healthy symbiosis may be the key to safeguarding the corals during the upcoming and current unprecedented rates of global warming (Edenhofer O. et al., 2014).

Studies have shown that there exist several major evolutionary lineages or *Symbiodinium* clades (Coffroth M. A. et al., 2005; Pochon X. et al., 2006; Pochon X. et al., 2010). It has also been discovered that genetic variations among these clades can influence bleaching susceptibility (Berkelmans R. et al., 2006; DeSalvo M. K. et al., 2010; Yuyama I. et al., 2012; Levin R. A. et al., 2016). Namely, clade D has been shown to display a more thermally tolerant phenotype than other clades (Glynn P. W. et al., 2001; Rowan R., 2004; Berkelmans R., 2006). This has led some to investigate the possibility that certain coral-symbiont pairs (or holobionts) may provide ecological advantages in elevated SSTs as well as the possibility of developing coral resistance through assisted evolution as a means of coral conservation (Baker A. C., 2001; van Oppen M. J. et al., 2015).

A more direct approach of genetically engineering has been contemplated and attempted on multiple occasions, however only one case of successful genetic transformation has been documented (ten Lohuis M. R. et al., 1998; Ortiz-Matamoros M. F. et al., 2015a; Ortiz-Matamoros M. F. et al., 2015b, Chen J. E. et al., 2017; Levin R. A. et al., 2017). In this study we attempt to make use of recent development in genomic and transcriptomic datasets on *Symbiodinium microadriaticum* (*S. microadriaticum*) to develop a custom expression construct that might facilitate future transformation attempts (Bayer T. et al., 2012; Baumgarten S. et al., 2013; Shoguchu E. et al., 2013; Quigley K. M. et al., 2014; Aranda M. et al., 2016; Levin R. A. et al., 2016; Parkinson J. E. et al., 2016).

In this study, we build a custom expression construct using selectable marker genes mCherry with the addition of a V5-His tag, and AmpR alongside putative endogenous *S. microadriaticum* promoters isolated by polymerase chain reaction (PCR). This construct offers a template for generating an expression construct made completely of custom expression constructs (Levin R. A. et al., 2017). This construct will hopefully be the first step towards generating a successful method of engineering *Symbiodinium* in the scope of increasing coral tolerance to bleaching.

Methods

Putative promoter region selection

Promoter regions were selected based on transcriptomics studies which defined a set of *S. microadriaticum* genes as those that are highly expressed in the genome. These genes encode for the large neutral amino acid transporter small subunit 4, the retinal-specific ATP-binding cassette transporter, peridinin Chlorophyll mRNA, caroteno chlorophyll-c-binding protein, a major basic nuclear protein, a gene encoding for a hypothetical protein, potentially a dinoflagellate viral nucleoprotein (Table 1) (Baumgarten S. et al., 2013; Levin R. A et al., 2016; Parkinson J. E. et al., 2016; Levin R. A. et al., 2017). We firstly found the sequences for these genes using the European Nucleotide Archive. We then used nucleotide blast to match these sequences to the *S. microadriaticum* genome (strain CCMP2467) to identify the location of the genes and subsequently our regions of interest for putative promoter elements. We then selected a pair of 1kb and 500bp sequences upstream of the (ATG) start codon for each gene. For the gene encoding for the large neutral amino acid transporter small subunit 4 we took three 1kb and 500bp regions. This was due to a high frequency of introns and multiple ATG start codons in the sequence. The first 500bp and 1kb pair was taken directly upstream of the first start codon, the second directly upstream of the second and the third directly upstream of the third. For the 1kb region upstream of the gene encoding for a caroteno Chlorophyll-c-binding protein, we took a slightly longer sequence (1,488bp) in order to avoid designing our forward primer to amplify directly within a section of DNA consisting of highly repetitive motifs of adenine and thymine.

Primer design

For the amplification of promoter elements in the *S. microadriaticum* genome, primers were designed to amplify 500bp or 1kb sequences upstream of the start codons of our genes of interest, with the addition of target sequences for our chosen restriction enzymes. The restriction enzymes were selected based on the restriction sites directly upstream and downstream of the CMV promoter on the pcDNA™ 3.1/Hygro(+) vector, the sequence for which we downloaded from SnapGene®. We then used NEBcutter® to view the list of non-cutter enzymes to assure our selected restriction enzymes would not cut inside our sequences. We cross referenced this list with the enzyme pair that would cut out the CMV promoter to make our restriction enzyme selection and chose MluI or MfeI for our forward primers and NheI for our reverse primers. Our forward primer sequences were designed with a TAT overhang followed by the restriction site (ACGCGT for MluI and CAATTG for MfeI) and the start of our promoter region sequence. These were then put into the eurofins Genomics oligo analysis tool and nucleotides were removed from the 3' end until the melting temperature is 63 ± 2°C. Reverse primers were designed with an ATA overhang followed by the restriction site (GCTAGC). The reverse complement was then taken and processed using the eurofins genomics oligo analysis tool. The mCherry primers were designed with restriction sites for BamHI on the forward primer (GGATCC) and SacII on the reverse primer (CCGCGG). The reverse primer did not amplify the stop codon on the mCherry gene as we want the V5-His tag to be expressed. Once the mCherry gene was inserted into the pcDNA™ 3.1/V5-His-TOPO® vectors, the mCherry and V5-His tag were amplified using primers containing restriction sites NheI (GCTAGC) on the forward primer and NotI (GCGGCCGC) on the reverse primer (Table 1). Primers for sequencing were selected 100bp upstream of the MluI restriction site and 100bp downstream of the NotI restriction site in order to successfully describe the promoter-mCherry-V5-His sequence.

Table 1: list of the primers used during PCR amplification, screening and sequencing. The oligonucleotide name refers to the name given to the primers and the shortened name of the genes (we will use these shortened names in the rest of the paper). The letters are acronyms for the gene product, the "p" stands for the primers amplifying a promoter region, letters directly after the "p" designate a different region upstream of the gene and the following numbers and letter signify the number of hundred base pairs amplified and the nature of the primer (forward or reverse). The sequence is the actual nucleotide sequence of the primers, Tm [C°] is the melting temperature of the sequences, the nomenclature is the complete description of the primer and the restriction enzyme is the enzyme that will target the restriction site on the primer.

Oligonucleotide Name	**SEQUENCE**	**Tm [C°]**	**NCBI code**	**Nomenclature**	**Restriction enzyme**
lnats4_pA5F	TATACGCG TTTGGCGT TTCGGGA	62.4	CCMP2467	Large neutral amino acid transporter small subunit 4 promoter region A, 500bp forward primer	MluI
lnats4_pA10F	TATACGCG TCCGGAAT CGTTCTAG	62.7	CCMP2467	Large neutral amino acid transporter small subunit 4 promoter region A, 1000bp forward primer	MluI
lnats4_pAR	ATAGCTAG CCTCACCC TCAAAGG	62.4	CCMP2467	Large neutral amino acid transporter small subunit 4 promoter region A, reverse primer	NheI
lnats4_pB5F	TATACGCG TCTTCTTAT CGCTAGGA CT	63.4	CCMP2467	Large neutral amino acid transporter small subunit 4 promoter region B, 500bp forward primer	MluI
lnats4_pB10F	TAT ACGCGT CGCTCCTT TTAAGGGT	63	CCMP2467	Large neutral amino acid transporter small subunit 4 promoter region B, 1000bp forward primer	MluI
lnats4_pBR	ATA GCTAGC GTGCACGA TTTCACAC	63	CCMP2467	Large neutral amino acid transporter small subunit 4 promoter region B, reverse primer	NheI
lnats4_pC5F	TATACGCG TTGCTACA TGTGGGGT TTA	63.4	CCMP2467	Large neutral amino acid transporter small subunit 4 promoter region C, 500bp forward primer	MluI
lnats4_pC10F	TATACGCG TAGAATGA AGGGTCCC T	63	CCMP2467	Large neutral amino acid transporter small subunit 4 promoter region C, 1000bp forward primer	MluI
lnats4_pCR	ATAGCTAG CGGCCGCA ATCTACAC	64.4	CCMP2467	Large neutral amino acid transporter small subunit 4 promoter region C, reverse primer	NheI

rsabct_p5F	TATACGCG TGCGTGGA GGCC	63.5	CCMP2467	Retinal-specific ATP-binding cassette transporter promoter region, 500bp forward primer	MluI
rsabct_p10F	TATACGCG TGCTTTGT GTAAGCCC A	63	CCMP2467	Retinal-specific ATP-binding cassette transporter promoter region, 1000bp forward primer	MluI
rsabct_pR	ATAGCTAG CCAATCTA GGCTCCAC	62.7	CCMP2467	Retinal-specific ATP-binding cassette transporter promoter region, reverse primer	NheI
pcm_p5F	TATACGCG TAAAACCC GAAACCCC A	63	2951	Peridinin Chlorophyll mRNA promoter, 500bp forward primer	MluI
pcm_p10F	TATACGCG TTGGAGCT TGAGGTTC A	63	2951	Peridinin Chlorophyll mRNA promoter region, 1000bp forward primer	MluI
pcm_pR	ATAGCTAG CGGCGATA GGAACGG	64.2	2951	Peridinin Chlorophyll mRNA promoter region, reverse primer	NheI
ccbp_p5F	TAT ACGCGT AGTATCCT TTCAAGGT GAG	63.7	CCMP2467	Caroteno chlorophyll-a-c-binding protein promoter region, 500bp forward primer	MluI
ccbp_p10F	TATACGCG TGAAGTGG TGAGTCG	62.4	CCMP2467	Caroteno chlorophyll-a-c-binding protein promoter region, 1000bp forward primer	MluI
ccbp_pR	ATAGCTAG CGTTAGTG CCCGGAAT	62.7	CCMP2467	Caroteno chlorophyll-a-c-binding protein promoter region, reverse primer	NheI
mbnp2_p5F	TATCAATT GCCAAAAC GTGGGCAC AC	63.2	CCMP2467	Major basic nuclear protein 2 promoter region, 500bp forward primer	MfeI
mbnp2_p10F	TATCAATT GCCGGCCT TTCGTCCT	62.7	CCMP2467	Major basic nuclear protein 2 promoter region, 1000bp forward primer	MfeI
mbnp2_pR	ATAGCTAG CCGCGAGG AAAAGCC	64.2	CCMP2467	Major basic nuclear protein 2 promoter region, reverse primer	NheI
hp_p5F	TATACGCG TGGCCACC GTGG	63.5	CCMP2467	Hypothetical protein promoter region, 500bp forward primer	MluI
hp_p10F	TATACGCG TATCTGGG GTTTACGG	62.7	CCMP2467	Hypothetical protein promoter region, 1000bp forward primer	MluI
hp_pR	ATAGCTAG CGGATGAG ACACGGTG	64.4	CCMP2467	Hypothetical protein promoter region, reverse primer	NheI

mCh_F	TATGGATC CATGGTGA GCAAGGG	62.4	CCMP2467	mCherry forward primer	BamHI
mCh_R	ATACCGCG G CTTGTACA GCTCG	64	CCMP2467	mCherry reverse primer	SacII
mChv5his_F	TATGCTAG CATGGTGA GCAAGGG	62.4	CCMP2467	mCherry v5his tag forward primer	NheI
mChv5his_R	ATACTTAA GTCAATGG TGATGGTG ATGATGA	62.9	CCMP2467	mCherry v5his tag reverse primer	AflII
test_F	CTGAGTAG TGCGCGAG CAAA	57.8	Not available	sequencing forward primer	Not available
test_R	TGGCACCT TCCAGGGT CAAG	59.8	Not available	sequencing reverse primer	Not available

Competent *E. coli* cells

We prepared the competent cells by growing 2 overnight starter cultures in 5ml LB medium (autoclaved 10g tryptone, 5g yeast extract, 10g NaCl) from single colonies from a plate streaked out with our desired *E. coli* cell line. We then inoculated two flasks containing 50ml SOB medium (autoclaved 20g tryptone, 5g of yeast extract, 0.5g of NaCl, 10ml of 0.25M KCl and 5ml of 2M $MgCl_2$) with 1ml of the starter cultures. These were incubated at 37°C, 200rpm until OD_{600} was between 0.4 – 0.6 (OD_{600} was at 0.427 and 0.430). Cultures were then transferred to pre-chilled flasks and centrifuged at 4°C for 15 minutes at 3500rpm. Pellets were then resuspended in 10ml ice-cold 100mM$CaCl_2$ and were incubated on ice for 30 minutes. Cells were then harvested at 3500rpm for 15 minutes at 4°C before being resuspended in 2ml of ice-cold 100mM $CaCl_2$. The two samples were pooled together and 1ml of 80% glycerol was added. This was dispensed as 50µl aliquots into chilled cryotubes that were flash frozen on dry ice before storage at -80°C.

E. coli transformation

E. coli cells were transformed with plasmid DNA by adding the DNA to the 50µl aliquots of competent cells and keeping these on ice for 30 minutes. In the case of the pcDNA™ 3.1/Hygro(+) and the pcDNA™ 3.1/V5-His-TOPO® vectors, 0.5µl were added to the aliquots, whereas for the ligation mixes, the full 10µl mixes were added. These were then heat-shocked at 42°C for 75 seconds and were then immediately placed on ice for 2 minutes. 450µl were then added to the samples before being placed in a shaking incubator at 37°C for 1 hour. In the case of the pcDNA™ 3.1/Hygro(+) and the pcDNA™ 3.1/V5-His-TOPO® vectors, 50µl was plated out onto ampicillin resistant plates whereas 200µl were plated out when using the ligation mixes. The ampicillin resistant plates were made by autoclaving a 1L bottle containing: 5g of tryptone, 2.5g of yeast

extract, 5g of NaCl, 7.5g of agar and 500ml of milliQ water. Once autoclaved and cool enough, 500µl of 50mg/ml ampicillin was added and stirred in right before plating.

PCR amplification

DNA sequences were amplified by Phusion PCR. For promoter region amplification we performed 50µl reactions. These were made up of 10µl of 5x HF buffer, 1µl of dNTPs (10µM), 2µl of each primer (which were diluted 1 in 10 from the stock concentration), 0.56µl of genomic *S. microadriaticum* DNA (100ng) and 0.5µl of Phusion polymerase with the rest of the volume made up with nuclease free water. These were amplified using the following cycles: initial denaturing was for 30 seconds at 98°C followed by 30 cycles of a 10 second denaturing step at 38°C, a 30 second annealing step at 62°C and a 30 second elongation step, both at 72°C. The final elongation step was 7 minutes long at 72°C. The mCherry plasmid was amplified in 100µl reactions and mCherry-V5-His in 50µl reactions. Both were made using 20ng of DNA.

PCR screening

DNA sequences from transformed *E. coli* cells were screened by GoTaq PCR. Colonies were picked from their respective plates, streaked out onto a grid plate and resuspended in 5µl of nuclease free water in a PCR tube. The rest of the 20µl reaction was made up of 4µl of GoTaq x5 buffer, 1.6µl of 25mM $MgCl_2$, 0.4µl 10mM dNTPs, 0.8µl of each primer and 0.1µl GoTaq polymerase. The PCR programme was as follows: 5 minutes at 94° C followed by 30 cycles of 30 seconds at 94°C, 30 seconds at 50°C and 1 minute at 72°C. This is followed by 7 minutes at 72°C and a 20°C hold. We used T7 and BGHR primers to screen the insertion of mChery into the pcDNA™ 3.1/V5-His-TOPO® vector and we used the same primer sets used to amplify the promoter regions to screen the insertion of promoter regions into the pcDNA™ 3.1/Hygro(+) vector. 5µl of the PCR reactions were then analysed using agarose gel electrophoresis.

DNA purification

DNA samples produced by digestions were purified using the Wizard®SV Gel and PCR Clean-Up System kit from Promega and following the Quick Protocol.

Miniprep plasmid DNA purification

For plasmids produced by ligation and transformed into *E. coli* cells we used the QIAprep® Spin Miniprep Kit and followed the Quick-StartProtocol to isolate and purify the plasmid.

Vector and backbone digestion

When digesting our backbones and inserts, we aimed to use 3000ng of backbone and 2000ng of insert in a 20µl reaction. When our DNA concentrations were not high enough to achieve the desired amounts, we added as much of the DNA as possible into a 20µl reaction. Reactions were incubated

at 37°C at 200rpm for a duration of time in accordance with manufacturer instructions. For two single digests, the DNA was purified in between digestions before being analysed and separated using agarose gel electrophoresis before a last purification step. For double digests, samples were analysed and separated using agarose gel electrophoresis immediately after the incubation period and subsequently purified. We started by inserting the mCherry gene into the pcDNA™ 3.1/V5-His-TOPO® vector in between the BamHI and SacII restriction sites (Figure 1a) and inserting our selected promoter regions into the pcDNA™ 3.1/Hygro(+) vector in between the MluI and NheI restriction sites (Figure 1b). The mCherry-V5-His sequence, once in the pcDNA™ 3.1/V5-His-TOPO® vector, was then digested and inserted into the promoter-Hygro construct using NheI and NotI.

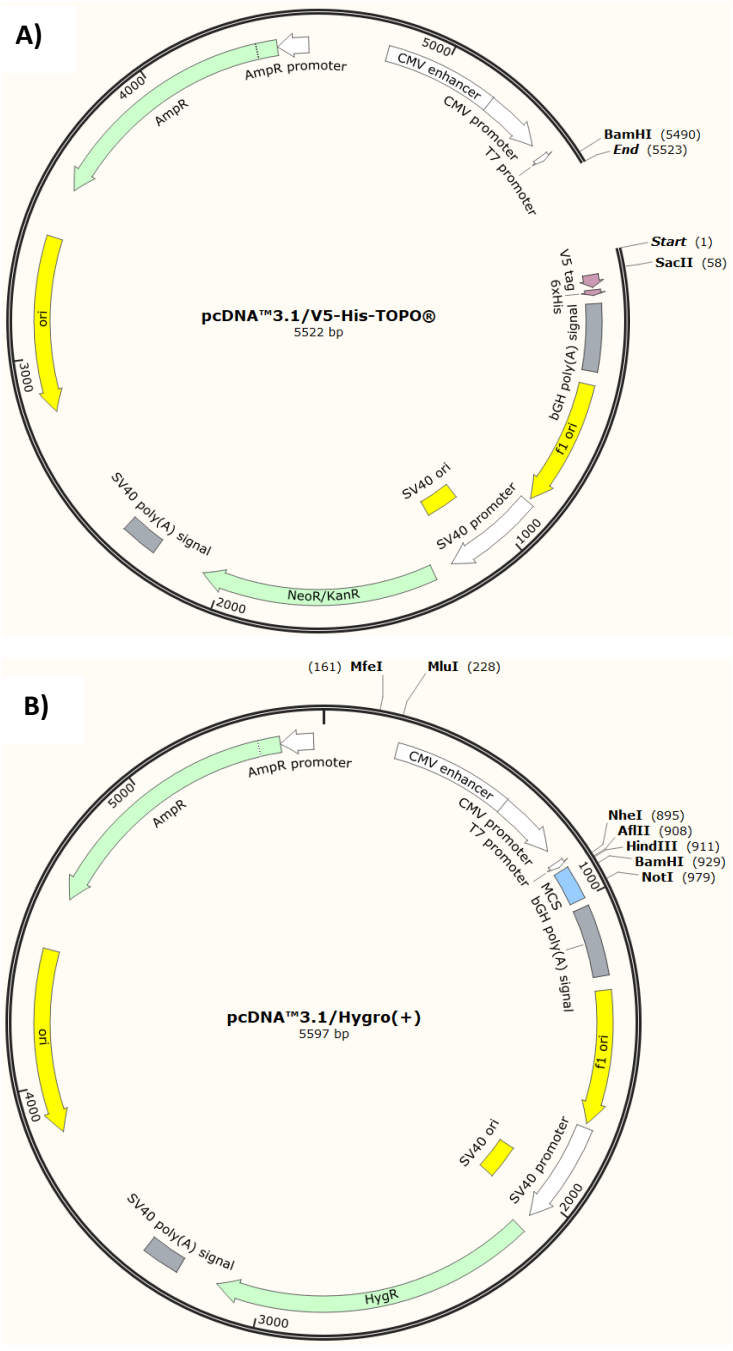


Figure 1: SnapGene® map of our vector templates and restriction sites used for digestions. **A)** Linear map of the pcDNA™ 3.1/V5-His-TOPO® vector showing BamHI and SacII restriction sites used during the ligation of the mCherry gene. **B)** Map of the pcDNA™ 3.1/Hygro(+) vector and the restriction sites used for the ligation of promoter regions (MluI and NheI) and mCherry-V5-His (NheI and NotI).

Ligation

Ligation mixes were made up of 2µl of the vector backbone, 6µl of the DNA insert, 1µl of the ligase buffer and 1µl of the T4 ligase enzyme. For blank ligation mixes, the DNA insert was replaced with nuclease free water. These were left to incubate overnight in cold storage before being transformed into *E. coli* cells.

Sequencing

Samples for the sequencing to GENEWIZ® and were sequenced by Sanger Sequencing. Forward and reverse primers were designed as 20bp oligoneucleotides 100bp upstream of the MluI restriction sites and 100bp downstream of the NotI restriction site (Table 1). The same primers were used for each sample. The sequences (supplementary data 1) were aligned using the MultiAlin tool. Reverse compliments of the reverse sequences were taken before alignment. Theoretical sequences were designed for alignment (Supplementary data 2).

Results

Generating the mCherry-V5-His oligonucleotide:

The fabrication of our custom expression construct was performed by firstly inserting the mCherry gene into the pcDNA™ 3.1/V5-His-TOPO® plasmid. We did this by transforming the pcDNA™ 3.1/V5-His-TOPO® plasmid into *E. coli* cells and performing a miniprep on single colonies to bulk up our backbone. We also amplified the mCherry gene insert using PCR. Once we had both our backbone and insert, we digested both using the BamHI and SacII restriction enzymes and performed a ligation followed by a transformation into *E. coli* cells which were plated on ampicillin resistant plates. We took a total of 10 single colonies that grew on the plate including 1 transformant containing the blank ligation and tested the mCherry-V5-His construct using PCR screening and ran these on agarose gel (Figure 2A). This revealed successful constructs which we amplified by PCR to produce mCherry-V5-His oligonucleotides. The primers used allowed the vectors to be digested by NheI (restriction site on the forward primer) and NotI (restriction site on the reverse primer). We tested these purified constructs again using agarose gel electrophoresis (Figure 2B). These results revealed bands at around 800bp, suggesting the successful amplification of the mCherry-V5-His oligonucleotide.

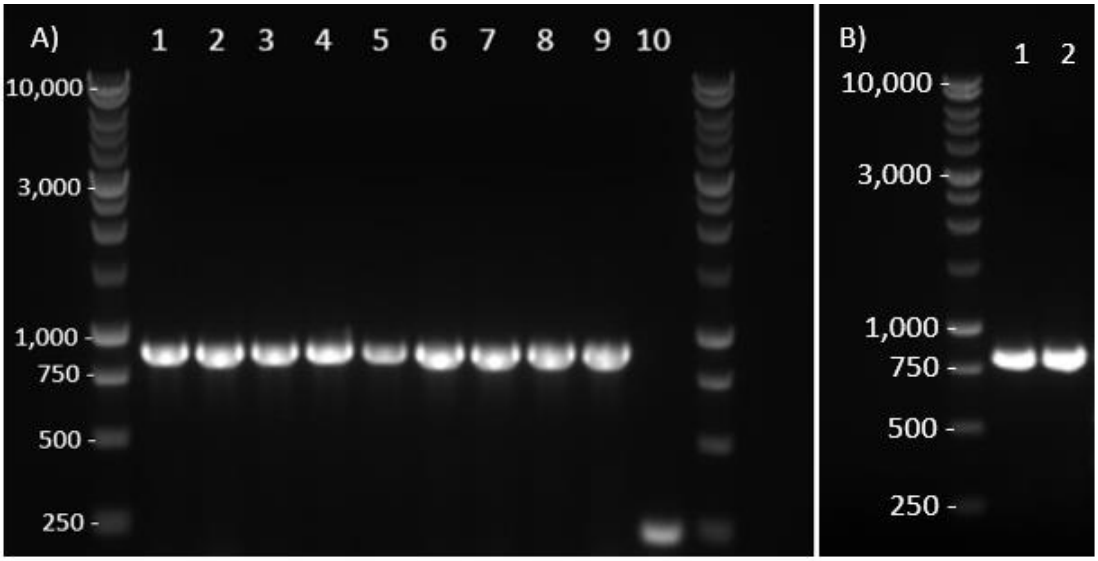


Figure 2: Agarose gel electrophoresis of PCR products from PCR screening. **A)** Samples [1] through [9] are vectors formed from the ligation of mCherry and the pcDNA™ 3.1/V5-His-TOPO® backbone. The bands observed for these are at approximately 875bp, which matches the expected size of 792bp for the mCherry gene and 114bp for regions between the gene and restriction sites (total of 906bp). Sample [10] is a blank ligation. **B)** Samples [1] and [2] are mCherry-V5-His vectors containing NheI and NotI restriction sites.

Generating the promoter-Hygro construct

We started this process by isolating the promoter regions from *S. microadriaticum* DNA using PCR amplification. We then verified the PCR products using agarose gel electrophoresis to verify the successful amplification of the sequences (Figure 3a). This revealed that we successfully isolated 14 out of our 16 putative promoter regions. We then proceeded to digest and ligate the promoter regions harbouring a MluI and NheI restriction site into the pcDNA™ 3.1/Hygro(+) vector and tested the constructs using PCR screening (Figure 3b-c). We screened three colonies for each of our constructs and made starter cultures and a subsequent miniprep for three successful samples (500bp lnats4_pA, 1kb ccbp_p and 1kb hp_p).

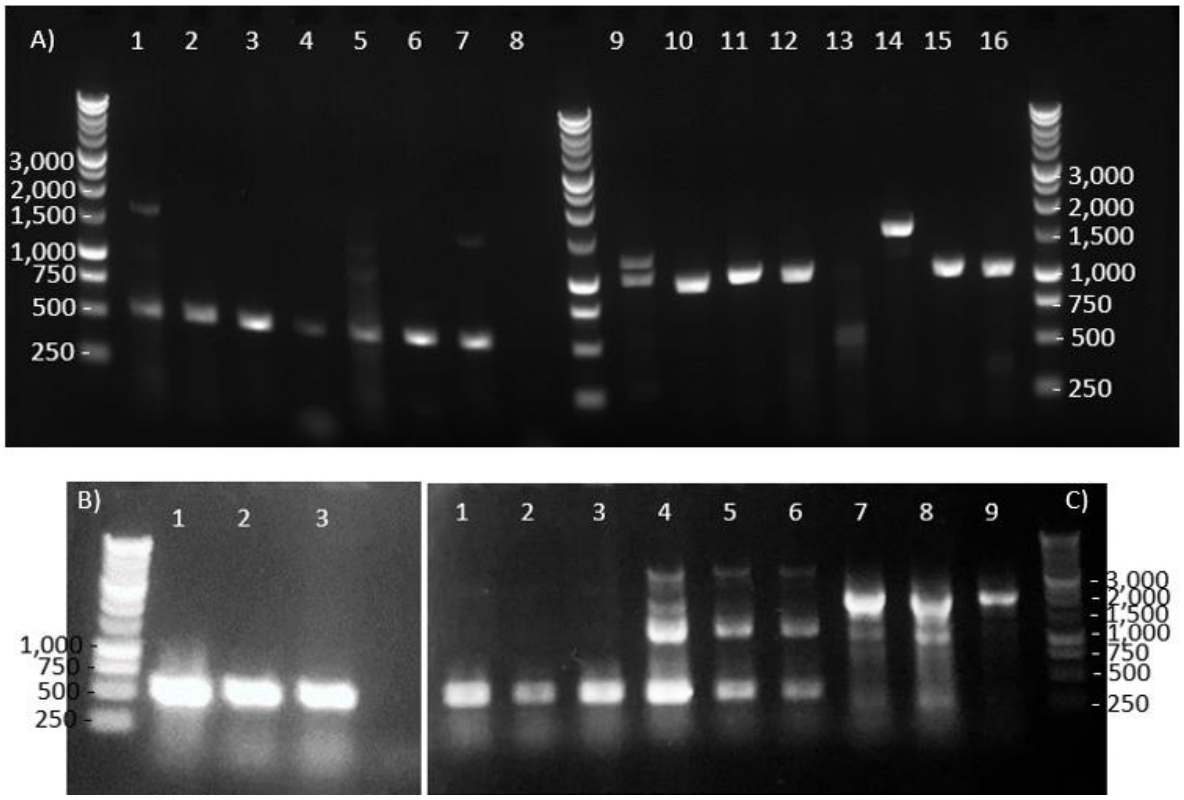


Figure 3: Agarose gel electrophoresis of PCR products for **A)** the amplification of promoter regions by PCR using the primer sets in Table 1. Samples [8] and [13] were unsuccessfully amplified. [1] 500bp lnats4_pA. [2] 500bp lnats4_pB. [3] 500bp lnats4_pC. [4] 500bp rsacpt_p. [5] 500bp pcm_p. [6] 500bp ccbp_p. [7] 500bp mbnp2_p. [8] 500bp hp_p. [9] 1kb lnats4_pA. [10] 1kb lnats4_pB. [11] 1kb lnats4_pC. [12] 1kb rsacpt_p. [13] 1kb pcm_p. [14] 1kb ccbp_p. [15] 1kb mbnp2_p. [16] 1kb hp_p. **B)** the screening of three repeat ligation of the 500bp promoter regions of the lnats4_pA gene and the pcDNA^{TM} 3.1/Hygro(+) vector and subsequent transformation into *E. coli* cells. Sample [1] was taken to make a starter culture and subsequent miniprep. **C)** the screening of the ligation of the 500bp promoter regions of the ccbp_p gene [1-3] and the 1kb promoter region of the ccbp_p [4-6] and hp_p genes [7-9]. Samples [4] and [8] were taken to make starter cultures for a miniprep.

Generating the promoter-mCherry-V5-His-Hygro construct

To create our final vector construct, we digested the mCherry-V5-His insert containing NheI and NotI restriction sites alongside the promoter-Hygro constructs using these two enzymes. This was followed by a ligation between the promoter-Hygro backbone and the mCherry-V5-His insert and finally a transformation into *E. coli* cells that were grown on ampicillin plates. After incubation, we took three cells for each of our ligation (excluding blanks) and miniprepped them before digesting using NheI and NotI and analysis using agarose gel electrophoresis (Figure 4). These results suggest that the ligation of mCherry-V5-His into the 1kb ccbp_p-Hygro constructs was unsuccessful as the 1kb band is not the expected size of the ccbp_p promoter and mCherry-V5-His. We would expect a successful insertion here to produce a band at 2,300bp (around 1,500bp for the promoter and around 800bp for mCherry-V5-His). We expected the successful insertion of mCherry-V5-His into the 1kb hp_p-Hygro construct to result in a 1800bp and therefore believe this construct to be successful.

The band for the insertion of mCherry-V5-His into the 500bp lnats4_pA-Hygro construct is at around 1300bp, the size we expected for the construct.

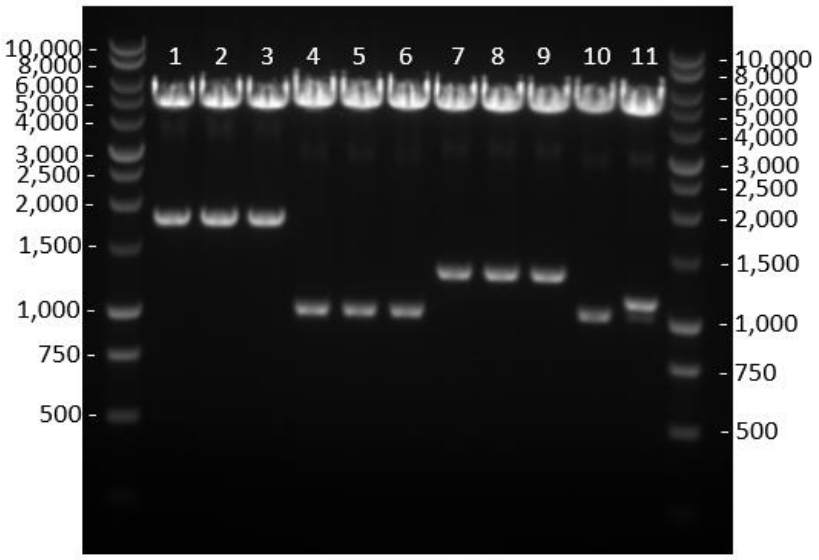


Figure 4: Agarose gel electrophoresis of our three promoter-mCherry-V5-His constructs. [1] – [3] are the constructs containing the 500bp lnats4_pA promoter, [4] – [6] contain the 1kb ccbp promoter and [7] – [9] contain the 1kb hp_p promoter. [10] and [11] are blank ligates containing the 1kb ccbp and the 1kb hp_p promoter respectfully (the blank containing the 500bp lnats4_pA promoter did not grow on the ampicillin plates).

Sequencing data

We sent the two samples we thought to be successful for sequencing. The forward sequencing for the 500bp lnats4_pA-mCherry-V5-His-Hygro construct failed due to early termination. Aligning the 500bp lnats4_pA-mCherry-V5-His theoretical sequence to the reverse sequence produced by sequencing for the 500bp lnats4_pA-mCherry-V5-His-hygro construct showed a strong match in the region of the mCherry-V5-His insert. This match continues 75bp into the 500bp lnats4_pA promoter however the sequence retrieved from sequencing did not continue further into the promoter (Supplementary figure 1).

Aligning the theoretical 1kb_hp_p_mCherry-V5-His sequence to the reverse sequence for the 1kb_hp_p_mCherry-V5-His construct showed a strong match in the region of the mCherry-V5-His insert (Supplementary figure 2). However, matches on the promoter sequence were poor and suggested that the amplified sequence was not the expected algal promoter.

Discussion

In this study, we discuss the process leading to the creation of a custom expression construct comprising of an endogenous *S. microadriaticum* putative promoter region, an mCherry gene and a V5-His tag all sequentially incorporated into a pcDNA™ 3.1/Hygro(+) vector. This was done by amplifying an mCherry gene, encased between BamHI and SacII restriction sites and ligating it into a pcDNA™ 3.1/V5-His-TOPO® vector. The mCherry-V5-His sequence was then amplified by PCR, encased by NheI and NotI restriction sites. The microalgal promoter regions were then amplified by PCR, encased by either MfeI and NheI or MluI and NheI restriction sites. These were then ligated into the pcDNA™ 3.1/V5-His-TOPO® vector before insertion of the mCherry-V5-His sequence. This lead to the creation of two supposed successful constructs: the first incorporating the 500bp large neutral amino acid transporter small subunit 4 putative promoter region, followed by the mCherry gene and V5-His tag. The second incorporating the 1kb putative promoter region of a hypothetical protein, supposedly a dinoflagellate viral nucleoprotein, followed by the mCherry gene and the V5-His tag. These final constructs were sent to be sequenced for verification and results showed the successful incorporation of mCherry and the V5-His tags. However, sequencing data showed that the promoter regions amplified did not match the expected sequences.

Given that our amplified sequences for both promoter regions matched the expected sizes, we believe that the reason for the disjunction between the sequencing data and the expected sequence may be due to the unusual biological features possessed by dinoflagellates. These include some of the largest reported nuclear genomes, unique promoter architecture, the presence of constantly condensed liquid-crystalline chromosomes, prevalent gene tandem arrays as well as the presence of multi-copy genes (Shoguchi E et al., 2013; Zhang H. et al., 2013; Lin S. et al., 2015; Aranda M. et al., 2016; Murray S. A. et al., 2016; Levin R. A. et al., 2017), features whichhave resulted in difficulties with transforming *Symbiodinium* in the past (Chen J. E. et al., 2017). Although these are more likely to result in difficulties with transformation into *Symbiodinium*, it may be possible that the highly repetitive nature of the *Symbiodinium* genome may have resulted in the amplification of an unexpected sequence. It is also possible that the harvesting conditions of the *S. microadriaticum* cells may have allowed for mutations or variations between the sequenced genome and our sample. However, these cells were supplied by Manuel Aranda from the King Abdul University of Science and Technology (KAUST) and although we do not have specific information on the exact process or origin of the cells, the process was presumably not unlike those applied in Aranda M. et al., 2016.

Although these results suggest that our experiment has failed in successfully isolating promoter regions from *S. microadriaticum*, it has not disproven the effectiveness of a custom expression construct. We believe that there still is potential in creating a custom expression construct made up of endogenous *S. microadriaticum* components such as the one described in Levin R. A. et al., 2017. Furthermore, the current trends in bioengineering strategies being applied to corals and their symbionts suggest a growing interest and genuine consideration for the application of bioengineering as a conservation strategy for tropical coral ecosystems (Levin R. A. et al., 2017; Cleves P. A. et al., 2018). The next step would be to reattempt creating an expression construct specifically designed to be tested in *Symbiodinium* cells and transform and modify its expression in *Symbiodinium* cells.

References

Aranda, M., Li, Y., Liew, Y.J., Baumgarten, S., Simakov, O., Wilson, M.C., Piel, J., Ashoor, H., Bougouffa, S., Bajic, V.B. and Ryu, T., 2016. Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. *Scientific Reports*, *6*, p.39734.Baker, A. C. 2001 Reef corals bleach to survive change. Nature 411, 765–766.

Baker, A.C., Glynn, P.W. and Riegl, B., 2008. Climate change and coral reef bleaching: An ecological assessment of long-term impacts, recovery trends and future outlook. *Estuarine, coastal and shelf science*, *80*(4), pp.435-471.

Baumgarten, S., Bayer, T., Aranda, M., Liew, Y.J., Carr, A., Micklem, G. and Voolstra, C.R., 2013. Integrating microRNA and mRNA expression profiling in Symbiodinium microadriaticum, a dinoflagellate symbiont of reef-building corals. *BMC genomics*, *14*(1), p.704.

Bayer, T., Aranda, M., Sunagawa, S., Yum, L.K., DeSalvo, M.K., Lindquist, E., Coffroth, M.A., Voolstra, C.R. and Medina, M., 2012. Symbiodinium transcriptomes: genome insights into the dinoflagellate symbionts of reef-building corals. *PloS one*, *7*(4), p.e35269.

Berkelmans, R., De'ath, G., Kininmonth, S. & Skirving, W. J. A comparison of the 1998 and 2002 coral bleaching events on the Great Barrier Reef: spatial correlation, patterns, and predictions. Coral Reefs 23, 74-83 (2004).

Berkelmans, R., De'ath, G., Kininmonth, S. and Skirving, W.J., 2004. A comparison of the 1998 and 2002 coral bleaching events on the Great Barrier Reef: spatial correlation, patterns, and predictions. *Coral reefs*, *23*(1), pp.74-83.

Bruno, J.F. and Selig, E.R., 2007. Regional decline of coral cover in the Indo-Pacific: timing, extent, and subregional comparisons. *PLoS one*, *2*(8), p.e711.

Chen, J.E., Cui, G. and Lastra, M.A., 2017. The genetic intractability of Symbiodinium microadriaticum to standard algal transformation methods. *bioRxiv*, p.140616.

Cleves, P.A., Strader, M.E., Bay, L.K., Pringle, J.R. and Matz, M.V., 2018. CRISPR/Cas9-mediated genome editing in a reef-building coral. Proceedings of the National Academy of Sciences, 115(20), pp.5235-5240.

Coffroth, M.A., 2005. Genetic diversity of symbiotic dinoflagellates in the genus Symbiodinium. *Protist*, *156*, pp.19-34.

De'ath, G., Fabricius, K.E., Sweatman, H. and Puotinen, M., 2012. The 27–year decline of coral cover on the Great Barrier Reef and its causes. *Proceedings of the National Academy of Sciences*, p.201208909.

DeSalvo, M.K., Voolstra, C.R., Sunagawa, S., Schwarz, J.A., Stillman, J.H., Coffroth, M.A., Szmant, A.M. and Medina, M., 2008. Differential gene expression during thermal stress and bleaching in the Caribbean coral Montastraea faveolata. *Molecular ecology*, *17*(17), pp.3952-3971.

DeSalvo, M.K., Sunagawa, S., Fisher, P.L., Voolstra, C.R., IGLESIAS-PRIETO, R. and Medina, M., 2010. Coral host transcriptomic states are correlated with Symbiodinium genotypes. *Molecular ecology*, *19*(6), pp.1174-1186.

Eakin, C.M., Liu, G., Gomez, A.M., De La Cour, J.L., Heron, S.F., Skirving, W.J., Geiger, E.F., Tirak, K.V. and Strong, A.E., 2016. Global coral bleaching 2014–2017: status and an appeal for observations. *Reef Encounter*, *31*(1), pp.20-26.

Glynn, P.W., 1996. Coral reef bleaching: facts, hypotheses and implications. *Global change biology*, *2*(6), pp.495-509.

Glynn, P.W., 1984. Widespread coral mortality and the 1982–83 El Niño warming event. *Environmental Conservation*, *11*(2), pp.133-146.

Glynn, P.W., Maté, J.L., Baker, A.C. and Calderón, M.O., 2001. Coral bleaching and mortality in Panama and Ecuador during the 1997–1998 El Niño–Southern Oscillation event: spatial/temporal patterns and comparisons with the 1982–1983 event. *Bulletin of Marine Science*, *69*(1), pp.79-109.

Goreau, T., McClanahan, T., Hayes, R. and Strong, A.L., 2000. Conservation of coral reefs after the 1998 global bleaching event. *Conservation Biology*, *14*(1), pp.5-15.

Heron, S.F., Maynard, J.A., Van Hooidonk, R. and Eakin, C.M., 2016. Warming trends and bleaching stress of the world's coral reefs 1985–2012. *Scientific reports*, *6*, p.38402.

Hoegh-Guldberg, O., 1999. Climate change, coral bleaching and the future of the world's coral reefs. *Marine and freshwater research*, *50*(8), pp.839-866.

Hughes, T.P., Baird, A.H., Bellwood, D.R., Card, M., Connolly, S.R., Folke, C., Grosberg, R., Hoegh-Guldberg, O., Jackson, J.B., Kleypas, J. and Lough, J.M., 2003. Climate change, human impacts, and the resilience of coral reefs. *science*, *301*(5635), pp.929-933.

Hughes, T.P., Kerry, J.T., Álvarez-Noriega, M., Álvarez-Romero, J.G., Anderson, K.D., Baird, A.H., Babcock, R.C., Beger, M., Bellwood, D.R., Berkelmans, R. and Bridge, T.C., 2017. Global warming and recurrent mass bleaching of corals. *Nature*, *543*(7645), p.373.

Lesser, M.P., 2006. Oxidative stress in marine environments: biochemistry and physiological ecology. *Annu. Rev. Physiol.*, *68*, pp.253-278.

Lesser, M.P. and Farrell, J.H., 2004. Exposure to solar radiation increases damage to both host tissues and algal symbionts of corals during thermal stress. *Coral reefs*, *23*(3), pp.367-377.

Levin, R.A., Beltran, V.H., Hill, R., Kjelleberg, S., McDougald, D., Steinberg, P.D. and van Oppen, M.J., 2016. Sex, scavengers, and chaperones: transcriptome secrets of divergent Symbiodinium thermal tolerances. *Molecular biology and evolution*, *33*(9), pp.2201-2215.

Levin, R.A., Voolstra, C.R., Agrawal, S., Steinberg, P.D., Suggett, D.J. and van Oppen, M.J., 2017. Engineering strategies to decode and enhance the genomes of coral symbionts. *Frontiers in microbiology*, *8*, p.1220.

Lin, S., Cheng, S., Song, B., Zhong, X., Lin, X., Li, W., Li, L., Zhang, Y., Zhang, H., Ji, Z. and Cai, M., 2015. The Symbiodinium kawagutii genome illuminates dinoflagellate gene expression and coral symbiosis. *Science*, *350*(6261), pp.691-694.

Muscatine, L., 1990. The role of symbiotic algae in carbon and energy flux in reef corals. *Ecosystems of the world. Coral Reefs*.

Murray, S.A., Suggett, D.J., Doblin, M.A., Kohli, G.S., Seymour, J.R., Fabris, M. and Ralph, P.J., 2016. Unravelling the functional genetics of dinoflagellates: a review of approaches and opportunities. *Perspect. Phycol*, *3*(1), pp.37-52.

Muscatine, L.E.O.N.A.R.D. and Porter, J.W., 1977. Reef corals: mutualistic symbioses adapted to nutrient-poor environments. *Bioscience*, *27*(7), pp.454-460.

Ortiz-Matamoros, M.F., Islas-Flores, T., Voigt, B., Menzel, D., Baluška, F. and Villanueva, M.A., 2015. Heterologous DNA uptake in cultured Symbiodinium spp. Aided by Agrobacterium tumefaciens. *PloS one*, *10*(7), p.e0132693.

Ortiz-Matamoros, M.F., Villanueva, M.A. and Islas-Flores, T., 2015. Transient transformation of cultured photosynthetic dinoflagellates (Symbiodinium spp.) with plant-targeted vectors. *Ciencias Marinas*, *41*(1), pp.21-32.

Parkinson, J.E., Baumgarten, S., Michell, C.T., Baums, I.B., LaJeunesse, T.C. and Voolstra, C.R., 2016. Gene expression variation resolves species and individual strains among coral-associated dinoflagellates within the genus Symbiodinium. *Genome Biology and Evolution*, *8*(3), pp.665-680.

Pochon, X., Montoya-Burgos, J.I., Stadelmann, B. and Pawlowski, J., 2006. Molecular phylogeny, evolutionary rates, and divergence timing of the symbiotic dinoflagellate genus Symbiodinium. *Molecular phylogenetics and evolution*, *38*(1), pp.20-30.

Pochon, X. and Gates, R.D., 2010. A new Symbiodinium clade (Dinophyceae) from soritid foraminifera in Hawai'i. *Molecular phylogenetics and evolution*, *56*(1), pp.492-497.

Quigley, K.M., Davies, S.W., Kenkel, C.D., Willis, B.L., Matz, M.V. and Bay, L.K., 2014. Deep-sequencing method for quantifying background abundances of Symbiodinium types: exploring the rare Symbiodinium biosphere in reef-building corals. *PloS one*, *9*(4), p.e94297.

Rowan, R., 2004. Coral bleaching: thermal adaptation in reef coral symbionts. *Nature*, *430*(7001), p.742.

Shoguchi, E., Shinzato, C., Kawashima, T., Gyoja, F., Mungpakdee, S., Koyanagi, R., Takeuchi, T., Hisata, K., Tanaka, M., Fujiwara, M. and Hamada, M., 2013. Draft assembly of the Symbiodinium minutum nuclear genome reveals dinoflagellate gene structure. *Current biology*, *23*(15), pp.1399-1408.

Spalding, M.D. and Brown, B.E., 2015. Warm-water coral reefs and climate change. Science, 350(6262), pp.769-771.

Trench, R.K., 1979. The cell biology of plant-animal symbiosis. *Annual Review of Plant Physiology*, *30*(1), pp.485-531.

van Oppen, M.J., Oliver, J.K., Putnam, H.M. and Gates, R.D., 2015. Building coral reef resilience through assisted evolution. *Proceedings of the National Academy of Sciences*, *112*(8), pp.2307-2313.

Warner, M.E., Fitt, W.K. and Schmidt, G.W., 1999. Damage to photosystem II in symbiotic dinoflagellates: a determinant of coral bleaching. *Proceedings of the National Academy of Sciences*, *96*(14), pp.8007-8012.

Wilkinson, C.C., 2004. *Status of coral reefs of the world: 2004*. Australian Institute of Marine Science (AIMS).

Yuyama, I., Harii, S. and Hidaka, M., 2012. Algal symbiont type affects gene expression in juveniles of the coral Acropora tenuis exposed to thermal stress. *Marine environmental research*, *76*, pp.41-47.

Zhang, H., Zhuang, Y., Gill, J. and Lin, S., 2013. Proof that dinoflagellate spliced leader (DinoSL) is a useful hook for fishing dinoflagellate transcripts from mixed microbial samples: Symbiodinium kawagutii as a case study. *Protist*, *164*(4), pp.510-527.

Supplementary data and figures

Supplementary data 1: Sequences recovered from sequencing. These samples were processed by removing nonsense nucleotides from both the 3' and 5' ends. The reverse compliments of the reverse sequences were taken instead of the raw data.

1kb hp_p-mCherry-V5-His forward sequence:

GCTTGACCGANAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAG
ATATACGCGTGCTTTGTGTAAGCCCAGGGCAGCTCAGATAAGCCTCAAGGCGGATGAGAAAGAGAAGTCGAT
TTTAGTTGGAAGTTCGTTCGGTAGGCCTTAGGACAGCTGAGATCAAGTTCAGAAGCGATGTGAGCCACGTCTG
ATGCTTTTCTTTTCGGTGTTAGGACTGAGGACTTTGAAAGCATATCGGGGCCAAGGGGATTTGAGCACGGACA
TTGTCAGGGTTCGGGCGTTAAGGAATCTTAGGGGTTAGGGTTTAAGACCTAGGTTCCGGGCTTACGTGCGCG
CGGGGGTACAACAGCTTACGACGAGACGTTCATTGAATCTGCCCCAACTCTTGCTCCAAGCTGGTGGTGCGTC
GCAACGTTTAGCGGGGACGGTGGTCTCGCCCAAAAGGGGAGTGTTCTTGCGTGCTTCGCAGGACTTCATGCA
GATTGGGTTTCCAATAGTATTAGCTCTAGGCAAAGAAGCAAACGTGACTAACACCGTTGTTTCTGCACGCACG
CAGCGTGGAGGCCAACTCGCTGCGTATTCAAAGGAAGACTCCAAGCCCCGACACCATCAAACCTGAGTCCCGT
GAAGCCCTTCGCCATCCCGACGGGACCTTATAACAAAGCCCTACAGAGCCAAAGTTTACATCCCCGAAGTTTC
AACCCCCCAGAAGAAACCCCAGAATGAAAGAACTCAGGCCAGAAAGCAACGTACGAAACCTCCAAGCAACCC
TGTACTCCTACACTAACGTGATCCGCAAACCCCAAAGTAACATATGCTTCAAAAGATTGCCTGAGGCGTGCCCC
CTGGCGAAGTGTTTTGGGCCAGGGGCCCTAGCGACGCCGTGTGGTCGGAGAAGAATCGGGGATACATGGGC
TTTATATCGGAATTGATAATCATGCTGGTCTGTGCGAATGTTTAGGTCTCTATTGAGTCGTAGGTTTGATGAGT
GTTTACAATAGGGCGTTATTGAGTTTTCTATGAACAGGGTGGGGGTTTAGCGGTGGGAGCCTAGATTGGCTA
GCATGGTGAGCAAGGGNGNAGGAGGATAACATGGNCCATCATCAAGGAGTTCATGNNCTTCAAGGTGCACA
TGGGAGGGCTCCGTGAACGGCCACAAATTCGAGAATCAAGGGCCAAGGGCAAGGGCCGGCCCNTNCCANG
GGCACCCNAAACCGCCAAGCTGAANGGNNACCAAGGGGGGGCCCCCTGGCCCTTTGCCNGGGGAAATCCTG
GTCCCCTCAATTNANNGGTAAGGGTTCCAAGGGCCTAAGGGAAANNCCCCCGCCGAAAATTCCCCAAAAACT
TGAAANNTGTCNTTCCCCCAGAGGGTTTTAAGGGGGGAAGGGGGGGGGAAAATTTTCCGGAACCGNG

1kb hp_p-mCherry-V5-His reverse sequence:

GCTTGACCGANAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAG
ATATACGCGTGCTTTGTGTAAGCCCAGGGCAGCTCAGATAAGCCTCAAGGCGGATGAGAAAGAGAAGTCGAT
TTTAGTTGGAAGTTCGTTCGGTAGGCCTTAGGACAGCTGAGATCAAGTTCAGAAGCGATGTGAGCCACGTCTG
ATGCTTTTCTTTTCGGTGTTAGGACTGAGGACTTTGAAAGCATATCGGGGCCAAGGGGATTTGAGCACGGACA
TTGTCAGGGTTCGGGCGTTAAGGAATCTTAGGGGTTAGGGTTTAAGACCTAGGTTCCGGGCTTACGTGCGCG
CGGGGGTACAACAGCTTACGACGAGACGTTCATTGAATCTGCCCCAACTCTTGCTCCAAGCTGGTGGTGCGTC
GCAACGTTTAGCGGGGACGGTGGTCTCGCCCAAAAGGGGAGTGTTCTTGCGTGCTTCGCAGGACTTCATGCA
GATTGGGTTTCCAATAGTATTAGCTCTAGGCAAAGAAGCAAACGTGACTAACACCGTTGTTTCTGCACGCACG

CAGCGTGGAGGCCAACTCGCTGCGTATTCAAAGGAAGACTCCAAGCCCCGACACCATCAAACCTGAGTCCCGT
GAAGCCCTTCGCCATCCCGACGGGACCTTATAACAAAGCCCTACAGAGCCAAAGTTTACATCCCCGAAGTTTC
AACCCCCCAGAAGAAACCCCAGAATGAAAGAACTCAGGCCAGAAAGCAACGTACGAAACCTCCAAGCAACCC
TGTACTCCTACACTAACGTGATCCGCAAACCCCAAAGTAACATATGCTTCAAAAGATTGCCTGAGGCGTGCCCC
CTGGCGAAGTGTTTTGGGCCAGGGGCCCTAGCGACGCCGTGTGGTCGGAGAAGAATCGGGGATACATGGGC
TTTATATCGGAATTGATAATCATGCTGGTCTGTGCGAATGTTTAGGTCTCTATTGAGTCGTAGGTTTGATGAGT
GTTTACAATAGGGCGTTATTGAGTTTTCTATGAACAGGGTGGGGGTTTAGCGGTGGGAGCCTAGATTGGCTA
GCATGGTGAGCAAGGGNGNAGGAGGATAACATGGNCCATCATCAAGGAGTTCATGNNCTTCAAGGTGCACA
TGGGAGGGCTCCGTGAACGGCCACAAATTCGAGAATCAAGGGCCAAGGGCAAGGGCCGGCCCNTNCCANG
GGCACCCNAAACCGCCAAGCTGAANGGNNACCAAGGGGGGGCCCCCTGGCCCTTTGCCNGGGGAAATCCTG
GTCCCCTCAATTNANNGGTAAGGGTTCCAAGGGCCTAAGGGAAANNCCCCCGCCGAAAATTCCCCAAAAACT
TGAAANNTGTCNTTCCCCCAGAGGGTTTTAAGGGGGGAAGGGGGGGGGAAAATTTTCCGGAACCGNG

500bp_lnats4_pA-mCherry-V5-His reverse sequence

CCNTAAACCCTAACCCTAAAACCCCTAAATTCTAATCCCTAAACCCCGCAAGCCCGGCAGCCTTTGAGGGTGA
GGCTAGCATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGC
ACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGG
CACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAG
TTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGA
GGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCT
GCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCA
GAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGA
TCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAG
AAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGAAATCACCTCCCACAACGAGGACTACA
CCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGCCG
CGGTTCGAAGGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTACGCGTACCGGTCATCATCACCATCA
CCATTGAGCGGCCGCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTGC
CAGCCATCTGTGT

Supplementary data 2: Theoretical sequences of the constructs used for alignment with sequencing data. These sequences were designed using the NCBI putative *S. microadriaticum* promoter sequences (in black), the sequences for the restriction enzymes used for digestions (red), the mCherry sequence (yellow), the V5-His tag sequence (green) and 100bp upstream and downstream of the MluI and NheI restriction sites (respectively) on the pcDNA™ 3.1/Hygro(+) vector (blue).

Sequence for 1kb_hp_p-mCherry-v5-His construct

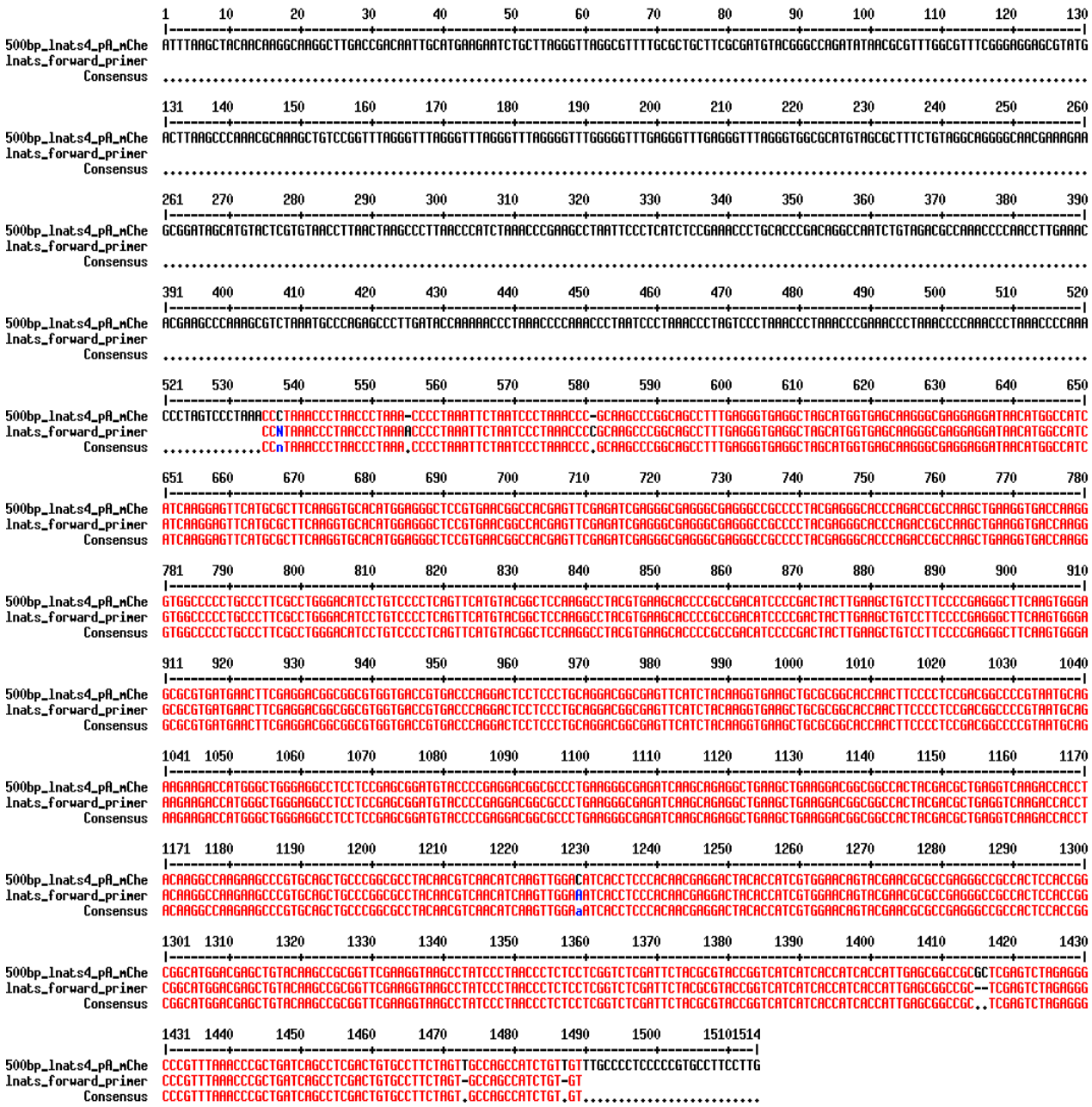
ATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGC
TGCTTCGCGATGTACGGGCCAGATATAACGCGTATCTGGGGTTTACGGGAGGGTTCAGGGCCTAGATTGTGG
GGCATACACGCTTAGAGGATTTGGATGTAGGTGTCGTGGTGGGAATTCGCAGGATGAGTGCGCTTTAGAAGC
CTCTCGACTTTTGGGCAGGGTCCTTGGGGCCTCGGATCTGTAAGGCAGCTTTCCTGCTTGAGCCTCCATCCATC
CGCGCTTCTCTCCATTCAGACAAGCCAACTTCGACCCGAACCCAACCGCACGAGTCCCCATGCCACCTGATGCT
TCTGAAGACCTGCACGCTCGAATCAATGATGATTCGGGTCAATGCTCCGCGGCTTTTCCAAAGCCCTCCTACAC
CTTCAAACTTTCCAACGGTTCGAAGAAGCTTGCGAACCGGATCCCTGAAGCCTACAGAACCCCAAAGCCTCAT

CATCTCCTAAAGCCCGAAAGCCCCAATCAAATCCTCGCCTGGTTTCCTCCGAGGTTTTGCCCGCGGCACGTGTC
ATCTCCCGCAATGCCGGCTTGGCCACCGTGGCCACGCCCGCTCGCATCCCTGAGTGCTGGTGCAGAGCTAGTT
GCCGAACTAGTAAGAGTCAGTCCAGGTCCTGGCATGCTCAGAACTTAGACATGCAGTGAGCAAGTCACCGAA
ACTTTAATCTTGCTTGGTTCATAGAGCATGGTCGCTGCTGGTCGCTTCAAAATTCCCGATGCTGCACTCCCACG
AATGTGCGTGAGAGACTATAGACTTTGGACATCTTTGGAATGATATAGCGAATAGCGATGCAATCCTGCGAG
GGAAGGGAAGACTTCAGCCTTAGGGGCAGGGTTCTCCTGATCACAGCGAACTAGAGAGTGCTGGCAAAAAC
GAAAGGTTTATAGAGACAGAGGCCGATGTTTCGTGGCAAGGGACGTGGTTACAAGTATTCAGTATTCATGTG
TGCGAACGTGCCCGACAAACTTGGCCACCTCAAGCCTGCTCGCCCTGCCTTCCGACCACTCTCCAGAGCCTCAC
CGTGTCTCATCCGCTAGCAtggtgagcaagggcgaggaggataacatggccatcatcaaggagttcatgcgcttcaaggtgcacatgg
agggctccgtgaacggccacgagttcgagatcgagggcgagggcgagggccgcccctacgagggcacccagaccgccaagctgaaggtgacc
aagggtggccccctgcccttcgcctgggacatcctgtcccctcagttcatgtacggctccaaggcctacgtgaagcaccccgccgacatccccga
ctacttgaagctgtccttccccgagggcttcaagtgggagcgcgtgatgaacttcgaggacggcggcgtggtgaccgtgacccaggactcctccc
tgcaggacggcgagttcatctacaaggtgaagctgcgcggcaccaacttcccctccgacggccccgtaatgcagaagaagaccatgggctggg
aggcctcctccgagcggatgtaccccgaggacggcgccctgaagggcgagatcaagcagaggctgaagctgaaggacggcggccactacgac
gctgaggtcaagaccacctacaaggccaagaagcccgtgcagctgcccggcgcctacaacgtcaacatcaagttggacatcacctcccacaac
gaggactacaccatcgtggaacagtacgaacgcgccgagggccgccactccaccggcggcatggacgagctgtacaagccgcggTTCGAA
GGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTACGCGTACCGGTCATCATCACCATCACCATTGAGC
GGCCGCGCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCA
TCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTG

Sequence for 500bp_lnats4_pA-mCherry-v5-His construct

ATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGC
TGCTTCGCGATGTACGGGCCAGATATAACGCGTTTGGCGTTTCGGGAGGAGCGTATGACTTAAGCCCAAACG
CAAAGCTGTCCGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGGTTTGGGGGTTTGAGGGTTTGAGGGTTTAGG
GTGGCGCATGTAGCGCTTTCTGTAGGCAGGGGCAACGAAAGAAGCGGATAGCATGTACTCGTGTAACCTTAA
CTAAGCCCTTAACCCATCTAAACCCGAAGCCTAATTCCCTCATCTCCGAAACCCTGCACCCGACAGGCCAATCT
GTAGACGCCAAACCCCAACCTTGAAACACGAAGCCCAAAGCGTCTAAATGCCCAGAGCCCTTGATACCAAAAA
CCCTAAACCCCAAACCCTAATCCCTAAACCCTAGTCCCTAAACCCTAAACCCGAAACCCTAAACCCCAAACCCTA
AACCCCAAACCCTAGTCCCTAAACCCTAAACCCTAACCCTAAACCCCTAAATTCTAATCCCTAAACCCGCAAGCC
CGGCAGCCTTTGAGGGTGAGGCTAGCAtggtgagcaagggcgaggaggataacatggccatcatcaaggagttcatgcgcttcaa
ggtgcacatggagggctccgtgaacggccacgagttcgagatcgagggcgagggcgagggccgcccctacgagggcacccagaccgccaagc
tgaaggtgaccaagggtggcccccctgcccttcgcctgggacatcctgtcccctcagttcatgtacggctccaaggcctacgtgaagcaccccgcc
gacatccccgactacttgaagctgtccttccccgagggcttcaagtgggagcgcgtgatgaacttcgaggacggcggcgtggtgaccgtgaccc
aggactcctccctgcaggacggcgagttcatctacaaggtgaagctgcgcggcaccaacttcccctccgacggccccgtaatgcagaagaagac
catgggctgggaggcctcctccgagcggatgtaccccgaggacggcgccctgaagggcgagatcaagcagaggctgaagctgaaggacggcg
gccactacgacgctgaggtcaagaccacctacaaggccaagaagcccgtgcagctgcccggcgcctacaacgtcaacatcaagttggacatca
cctcccacaacgaggactacaccatcgtggaacagtacgaacgcgccgagggccgccactccaccggcggcatggacgagctgtacaagccg
cggTTCGAAGGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTACGCGTACCGGTCATCATCACCATCAC
CATTGAGCGGCCGCGCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTT
GCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTG

Supplementary figure 1: alignment of the theoretical sequence for the 500bp_lnats4_pA-mCherry-v5-His construct and the 500bp_lnats4_pA forward sequence



Supplementary figure 2: alignment of the theoretical sequence for the 1kb_hp_p-mCherry-v5-His construct and the 1kb_hp_p forward sequence