Screening for Fluorescent and Chromoproteins from South African Sea Anemones

By

Tanya Nyman

Submitted in partial fulfillment of the requirements for the degree of Magister Scientiae (M.Sc) in the department of Biotechnology, University of the Western Cape

Supervisor: Prof. DA Cowan

Co-Supervisors: Dr IM Tuffin

Dr W Loftie-Eaton

Declaration

I declare that *"Screening for Fluorescent and Chromoproteins from South African Sea Anemones"* is my own work, that it has not been submitted for any degree or examination in any other university, and the at all the sources I have used or quoted have been indicated and acknowledged by complete references.

Tanya Nyman

Abstract

Sea anemones (O*rder Actinaria*) are a diverse order from the C*lass Anthozoa*. They are found in all marine habitats at all depths and their symbiotic relationships play an important role in energy transfers especially in the benthic-pelagic community. The evolutionary background and phylogenetics of the class is poorly understood due to a lack of correspondence between taxonomic and molecular data (Daly et al. 2008). Therefore, a deeper exploration into Cnidarian molecular biology is needed to establish these as an evolutionary model organism.

Gene discovery from various marine invertebrates has facilitated the recovery of anti-cancer drugs, antibiotics and reporter genes (Faulkner, 2000; Allen and Jaspars, 2009). The most commercially lucrative products from sea anemones are fluorescent and chromoproteins (FP/ CP), which are used as non-invasive real-time reporter genes. The applications for these proteins are extensive and range from monitoring cellular processes such as protein localisation and interactions to imaging (Alieva et al. 2008). Therefore, novel FP and CPs have potential for commercialization.

The aims of the project were to analyze basic molecular diversity of the sea anemones *Pseudactinia varia*, *Pseudactina flagellifera* and *Bunodosoma capensis* and evaluate a new screening method to isolate novel FP and CPs.

To assess the basic molecular diversity, of the sea anemones and their associated symbionts 16S rRNA and 18S rRNA clone libraries were generated. The sea anemones used in this study clustered together with those of the *Family Actiniidae*. The bacterial associations observed based on the closest relative BLAST analysis were dominated by *Proteobacteria* (gamma, alpha and epsilon) as well as *Bacteroides*. The associate bacterial symbionts

possibly produce compounds that range from polyunsaturated fatty acids, poly-hydroxyalkanoates to anti-microbial compounds that aid the host in various processes.

In order to screen for FPs and CPs from sea anemones three types of cDNA libraries were generated to be screened either by sequence based or activity based approaches. Novel primers were designed which could be applied for the screening of a variety of Anthozoans. A positive control was also designed and synthesised in order to test the capability of the designed primers and optimise the amplification. Although amplicons were generated from gDNA and cDNA libraries from each of the sea anemones they were found to be non-specific products. The detection limit is likely to be the limiting factor. The construction of an activity based library was not achieved due to technical constraints, which highlights the need for new molecular tools in this field or improvements to the existing ones.

Acknowledgements

I would first like to acknowledge my supervisor, **Professor Don Cowan** for the opportunity to undertake research that I am very passionate about. His motivation and guidance throughout my academic training has been inspiring. You will be sorely missed at IMBM.

To **Associate Professor Marla Tuffin**, I am grateful for all your patience and encouragement when I complained about my lack of results, failed experiments and bad writing skills. Your enthusiasm for science is infectious and definitely drives each and every student in IMBM to do their best.

The maternal figure of the lab is most certainly **Dr Heide Goodman**. Thank you for taking care of us, we would be lost without you and your care and concern for each person in the lab. We often take you for granted and should show our appreciation for all your hard work more regularly.

Dr Wesley Loftie-Eaton, for putting up with such a temperamental student you deserves a medal! Thank you for all your effort and input throughout my thesis. Your advice and suggestions in my drafts were extremely valuable.

To my partners in crime **Jerry** and **Biff**, the other beer O clock members **Pete** and **François** and the other loud IMBM members (no-it wasn't just me and jerry) **JB, James, Lonnie, Francesca, Sam, Dom, Bron, Colin, Thuls, Timi** and **Dean** thanks for the fun times and making it enjoyable to work in IMBM.

My family - **Dad** and **Mom** and **Melly** and **Roxy**, living with a scientist in training is not the easiest. You have my eternal gratitude for dealing with my late nights up disturbing the quiet in the house and my moods when my experiments weren't going as planned. You all deserve a break more than me after this thesis.

To everyone (past and present) in **IMBM** and in the **Department** of **Biotechnology** as a whole I would like to thank you all for the years of experiences.

Thanks to the **NRF** for personal funding.

Dedication

To my Father, Mother, Sisters and Friends - My support system that keeps me together when everything else is falling to pieces.

Table of contents

List of Tables

List of Figures

List of Abbreviations

ACP	Anthozoan Chromoprotein
AFP	Anthozoan Fluorescent Protein
ARDRA	Amplified Ribosomal Digest Restriction Analysis
BFP	Blue Fluorescent Protein
BLAST	Basic Local Alignment Search Tool
bp	Base pair
cDNA	Complementary DNA
CFP	Cyan Fluorescent Protein
CFU	Colony Forming Unit
CP	Chromoprotein
dNTP	DeoxyriboNucleotide TriPhosphate
ELISA	Enzyme-Linked ImmunoSorbent Assay
FACS	Fluorescent Activated Cell Sorting
FP	Fluorescent Protein
FRET	Fluorescent Resonance Energy Transfer
gDNA	Genomic DNA
GFP	Green Fluorescent Protein
IPTG	IsoPropyl-Beta-D-1-ThioGalactopyranoside
kb	Kilobase pair
LB	Luria Bertani
M	Molar
MALDI-TOF-MS	Matrix Assisted Laser Desorption/Ionisation-Time Of Flight-Mass Spectrometry

Mbp	Megabase pair
µg	Microgram
mg	Milligram
µL	Microlitre
µM	Micromole
mM	MilliMolar
M-MLV	Moloney-Murine Leukaemia virus
MOPS	3-Morpholinopropane-1-sulfonic acid
mQ	Millipore Water
mRNA	Messenger RNA
nm	Nanometre
OFP	Orange Fluorescent Protein
OH	Optical Highlighter
PAFP	Photo-Activated Fluorescent Protein
PBS	Phosphate Buffered Saline
PCFP	Photo-Convertible Fluorescent Protein
PCR	Polymerase Chain Reaction
rDNA	Ribosomal DNA
RFLP	Restriction Fragment Length Polymorphism
RFP	Red Fluorescent Protein
rRNA	Ribosomal RNA
RSFP	Reversibly Switchable Fluorescent Protein
SDS	Sodium Dodecyl Sulfate
SIGEX	Substrate Induced Gene Expression
SMARTer	Switching Mechanism At the 5'RNA Transcript

SOB	Super Broth
SOC	Super Optimal Culture Broth
TE	Tris-EDTA Buffer
UV	UltraViolet
X-Gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside
YFP	Yellow Fluorescent Protein

1. Chapter One: Literature Review

1.1. Marine "blue" biotechnology

Marine or "blue" biotechnology is defined as the use of marine organisms, at the whole, cellular or molecular level to provide solutions benefiting society (Pomponi, 1998). Focused interest in blue biotechnology has recently shown practical consequences in many fields, particularly in pharmacology and ecology. The current global market for blue biotechnology is estimated at US $2.4 billion with an expected 10% annual growth (Allen and Jaspars 2009).

The oceans cover approximately 70% of the earth and represent the largest biome. They are characterised by extreme environmental factors such as high pressure, very high or low temperatures and low nutrient content. The vast majority of its biodiversity remains unexplored and untapped. South Africa is surrounded by approximately 3 800 Km of coastline and has easy access to diverse marine habitats rich in endemic species. Together with the growing South African science community, this presents a large potential for innovative and sustainable development of a marine biotechnology industry.

1.2. *Cnidaria: Order Actinaria*

Marine invertebrates such as sponges, corals, sea anemones, and jellyfish are a diverse and ecologically important group of animals. These cnidarians are simple askeletal animals which posses radial symmetry. Their body plan is divided into two forms, namely medusa (jellyfish) and polyp (sea anemones and corals). The polyp form (a stage which all medusa undergo) is generally sessile or has very slow and limited movement (Bridge et al. 1995). Their soft bodies and lack of an acquired immune system as well as physical protections (such as spikes, shell etc.) makes any Cnidarian a prime candidate for predation. Their

survival can, however, be attributed to their extremely well developed chemical defences that are now being more fully realised (Pawlik, 1993).

Members of the *Order Actinaria* (sea anemones) are a diverse and successful group of invertebrates found across a variety of marine and freshwater bodies, ranging from the tropical and temperate to cold waters. They are noted for their diverse symbiotic relationships which form an integral part of energy transfer especially in the benthic-palegic zone. Approximately 1 100 species of *Actinaria* have been described, but the evolutionary component of their ecological success is yet to be understood (Daly et al. 2008).

The established taxonomic classification is based on morphological characters such as polyp organization, tentacle/coelenteron relationship, the siphonoglyph and mesenteries characteristics etc. (for a comprehensive list see Daly et al. 2003). The challenge relating to taxonomy and assigning phylogenetic relationships includes the inconsistent patterns, lack of correspondence between ecological, morphological and biological variables and the drawing on presence or absence of features instead of synapomorphies. This has led to poorly distinguished families, genera and species within the group. Molecular techniques that use key marker genes to determine relatedness and resolve evolutionary lineages have recently been employed to aid in taxonomic classification (Daly et al. 2008).

Many drugs, secondary metabolites and useful peptides have been isolated from marine invertebrates and are currently under clinical trials for treatment of cancer, HIV and other diseases (for reviews see Hay and Fenical 1996; Faulkner, 2000; Proksch et al. 2002). Of these examples, the most recognised molecular tool to have emerged from the *phylum Cnidaria* and marine invertebrates is the Green Fluorescent Protein (GFP).

1.3. History of fluorescent and chromoproteins

1.3.1. Green fluorescent protein from *Aequorea victoria*

The GFP, as named by Morin and Hastings (1971), was discovered by Shimomura et al. (1962) during their work on the characterization of aequorin, a 21.4KDa calcium-dependent photoprotein from the bioluminescent *Aequorea* jellyfish (Niwa et al. 1996). These researchers demonstrated that pure aequorin results in blue chemiluminescence with an emission peak at 407 nm (Johnson et al. 1962). However, *A. victoria* tissues had a green glow. In response to this finding a second light emitting protein, the GFP, consisting of 238 amino acids with an approximate mass of 26 kDa was also isolated. This protein has the natural ability to emit light without being involved in an enzymatic reaction. The fluorescence is generated by absorbing light of a specific wavelength and re-emitting it in another. The characterized excitation (395 nm and 470 nm) and emission (508 nm) spectra of GFP (Johnson et al. 1962) was linked to that of the green luminescence of living *Aequorea* tissues. Morin and Hastings (1971) suggested a radiationless energy transfer mechanism whereby the GFP acts as a secondary emitter which is excited by light emitted by aequorin.

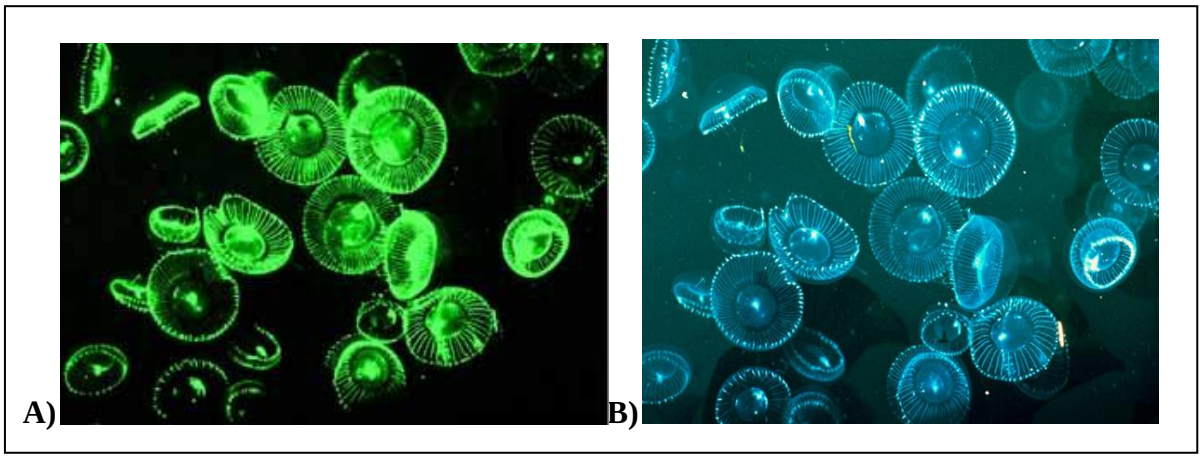


Figure 1. Photographs of *Aequorea victoria* illustrating its various biological light systems. A) the glow of GFP fluorescence under black light and B) illumination of aequorin bioluminescence.

The gene encoding GFP contains all the necessary information for synthesis of the chromophore which is responsible for the fluorescence observed. Post-translational modifications that make fluorescence possible are processed without the need of coelenterate-specific enzymes, external co-factors (other than oxygen) or chaperones (Prasher et al. 1992; Tsien, 1998). This was demonstrated when the gene for GFP was cloned (Prasher et al. 1992) and shown to efficiently express in other eukaryotic and prokaryotic organisms with resulting fluorescence (Chalfie et al. 1994; Inouye and Tsuji 1994; Inouye et al. 1997). Subsequently, through several site directed mutations, genetic variants of GFP were developed (see Figure 2) emitting in the blue (Yang et al. 1998), cyan (Heim and Tsien 1996), enhanced green (Heim et al. 1995), yellow (Wachter et al. 1998) and red (Mishin et al. 2008) regions of the visible spectrum.

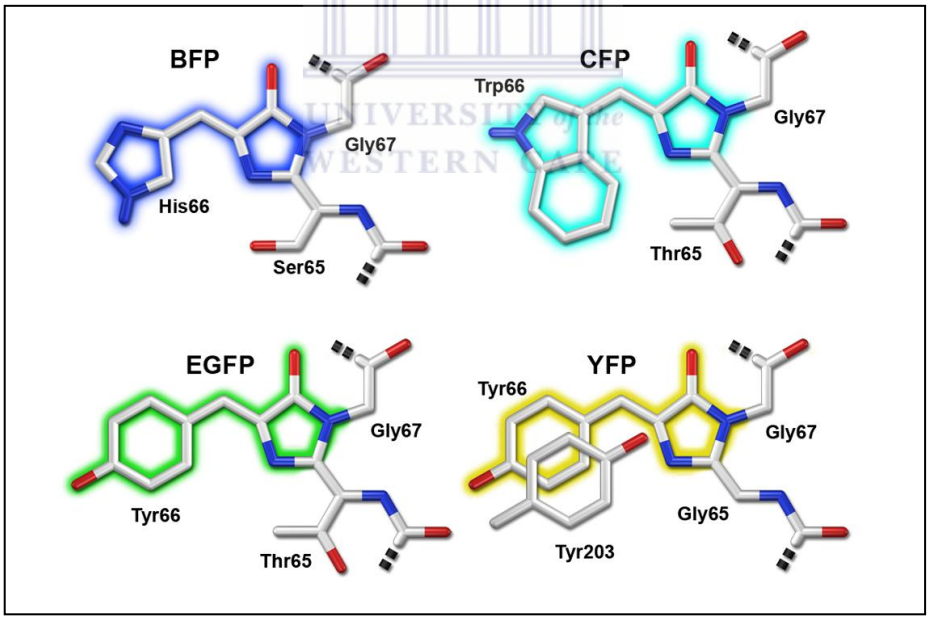


Figure 2. Chromophore structures resulting from site directed mutations of *A. victoria's* GFP. Various chromophore colours are labelled accordingly. The image is from (Shaner et al. 2007).

1.3.2. Anthozoan fluorescent and chromoproteins

In 1999 Matz and co-workers published a study detailing a group of proteins in non-bioluminescent anthozoans which absorb and re-emit light as fluorescence similar to the

Aequorea GFP. Despite the similar tertiary structure the isolated anthozoan fluorescent proteins (AFPs) share low (approximately 26-30%) amino acid sequence homology to GFP. AFPs have a monomeric size ranging between 25-30 kDa, share considerable homology with each other and comprise a diverse spectrum of colours. Brightly coloured proteins sharing substantial homology to anthozoan fluorescent proteins have also been isolated from anthozoans. These anthozoan chromoproteins (ACP) are able to absorb light but not re-emit it, thus resulting in displays of colour or very weak fluorescence (Verkhusha and Lukyanov 2004).

Anthozoan fluorescent and chromoproteins display unique characteristics such as optical highlighting (Ando et al. 2002; Ando et al. 2004; Chudakov et al. 2007; Patterson et al. 1997) and novel chromophore structures (Pletneva et al. 2007). Additionally their extensive spectral diversity is promising for multi-labelled imaging as well as Fluorescent Resonance Energy Transfer (FRET) analysis (Pollok and Heim 1999; Karasawa et al. 2004; Chudakov et al. 2005; Verkhusha and Lukyanov 2004; Shaner et al. 2007). Most wild-type fluorescent proteins are, however, unsuited to biotechnological applications and need to undergo extensive mutagenesis to achieve the desired qualities. Through rational design and directed evolution, characteristics such as oligomerisation, fluorescent brightness, photostability, temperature or pH sensitivity etc. can be improved (Yang et al. 1998; Nagai et al. 2002; Miyawaki et al. 2003; Chudakov et al. 2006; Chan et al. 2006; Nienhaus et al. 2006; Stiel et al. 2008; Piatkevich and Verkhusha 2009).

To date, approximately 100 fluorescent and chromoproteins have been characterised in only 4 phyla: Cnidaria (Matz et al. 1999; Fradkov et al. 2000; Gurskaya et al. 2001; Yanushevich et al. 2003; Wiedenmann et al. 2004; Shkrob et al. 2008; Delong et al. 2006; Ip et al. 2007; Vogt et al. 2008), Ctenophora (Haddock et al. 2009), Anthropoda (Masuda et al. 2006; Wilmann et al., 2006), and Chordata (Baumann et al. 2008; Bomati et al. 2009). Anthozoan

fluorescent and chromoproteins represent the largest percentage and widest spectral diversity of such proteins isolated thus far (Labas et al. 2002; Alieva et al. 2008).

1.3.3. Photoconvertable fluorescent and chromoproteins

Optical highlighters are FPs and ACPs that can be converted to a fluorescent state or shift the fluorescent emission spectrum when excited with different wavelengths of light (Miyawaki, 2002; Ando et al. 2004; Shaner et al. 2007; Vogt et al. 2008). The process can be either irreversible, such as in the case of photoactivated and photoconvertable proteins, or reversible in the reversibly switchable proteins.

The photoactivated fluorescent proteins (PAFP) and photoconverted fluorescent proteins (PCFP) are triggered by illumination with UltraViolet (UV) or violet light. Photoactivated FPs are switched from a dark to fluorescent state like the example PA-GFP (Patterson and Lippincott-Schwartz 2002). Photoconverted fluoresecent proteins undergo a shift in fluorescence emission from green to red and examples include Kaede (see Figure 3 A) (Ando et al. 2002) and Eos (Wiedenmann et al. 2004; Nienhaus et al. 2006). As the light stimulus in both groups is UV or violet it results in cleavage of a bond, thus rendering the process irreversible.

Reversibly switchable fluorescent proteins (RSFP) are triggered by laser illumination which induces rotation of a bond that results in a *trans* to *cis* isomerisation, as illustrated in Figures 3 B and 4 (Nienhaus et al. 2008). The bond rotation is reversible by either switching the wavelength of the light or by turning off the light source. Examples of proteins capable of reversible dark to fluorescent switching include Dronpa (Figure 4) (Ando et al. 2004; Shaner et al. 2007; Stiel et al. 2007; Asselberghs et al. 2008) and KFP (Figure 3B) (Henderson and Remington 2006; Lukyanov et al. 2000). Iris, a mutant of Eos, is the only current example of a RSFP able to switch from a dark state to either green or red fluorescence depending on the wavelength of illumination (Adam et al. 2008).

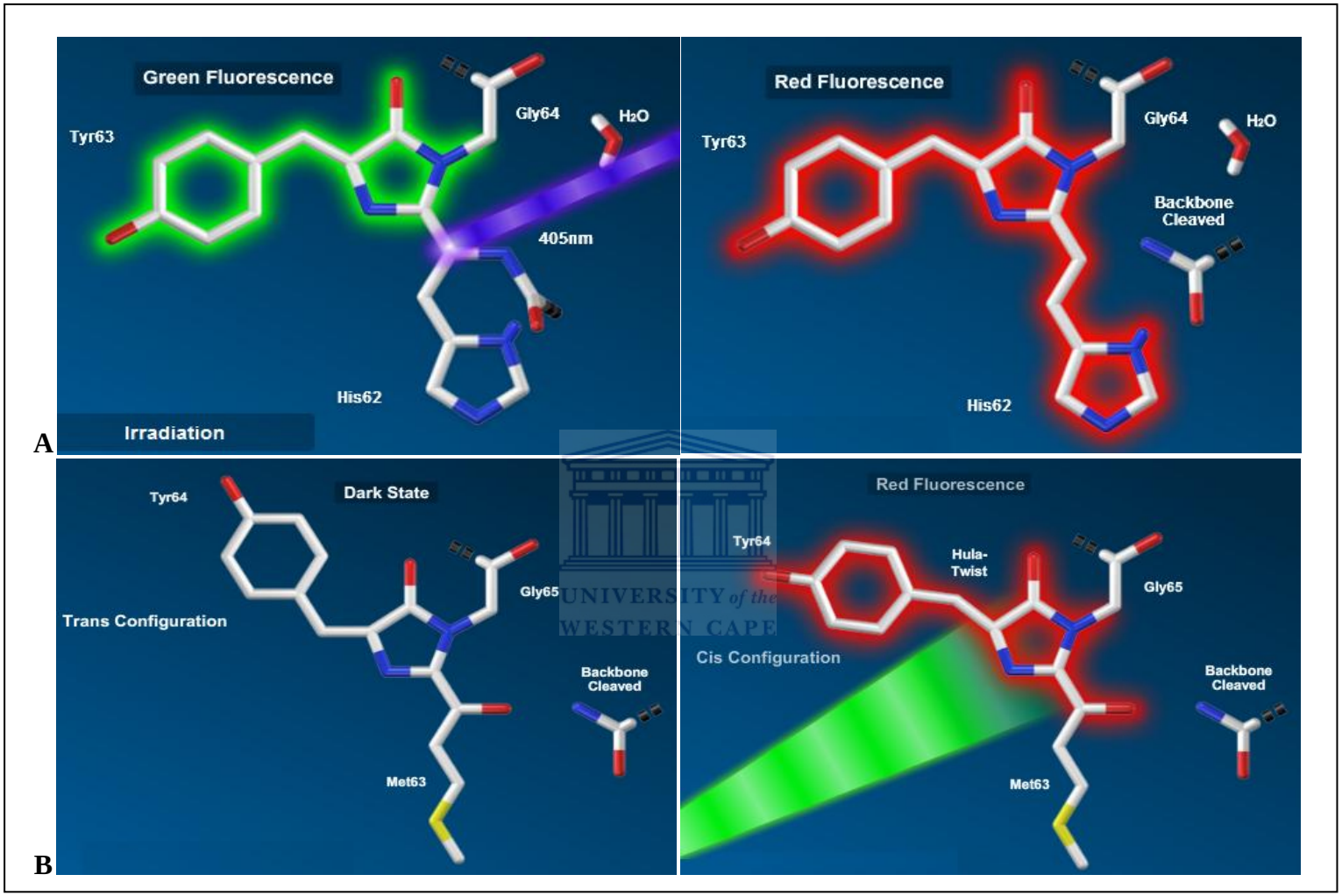


Figure 3. Chromophore structures of Optical Highlighters Kaede and the kindling fluorescent protein (KFP). A) Cleavage due to irradiation with UV light results in a irreversible red-shift for the Kaede protein's chromophore. B) After yellow light (525-580 nm) laser excitation, KFP undergoes *trans/cis* isomerisation and red fluorescence is observed. The chromophore returns to the dark state soon after light stimulus is switched off. Images adapted zeiss (http://zeiss-campus.magnet.fsu.edu/articles/probes/index.html)

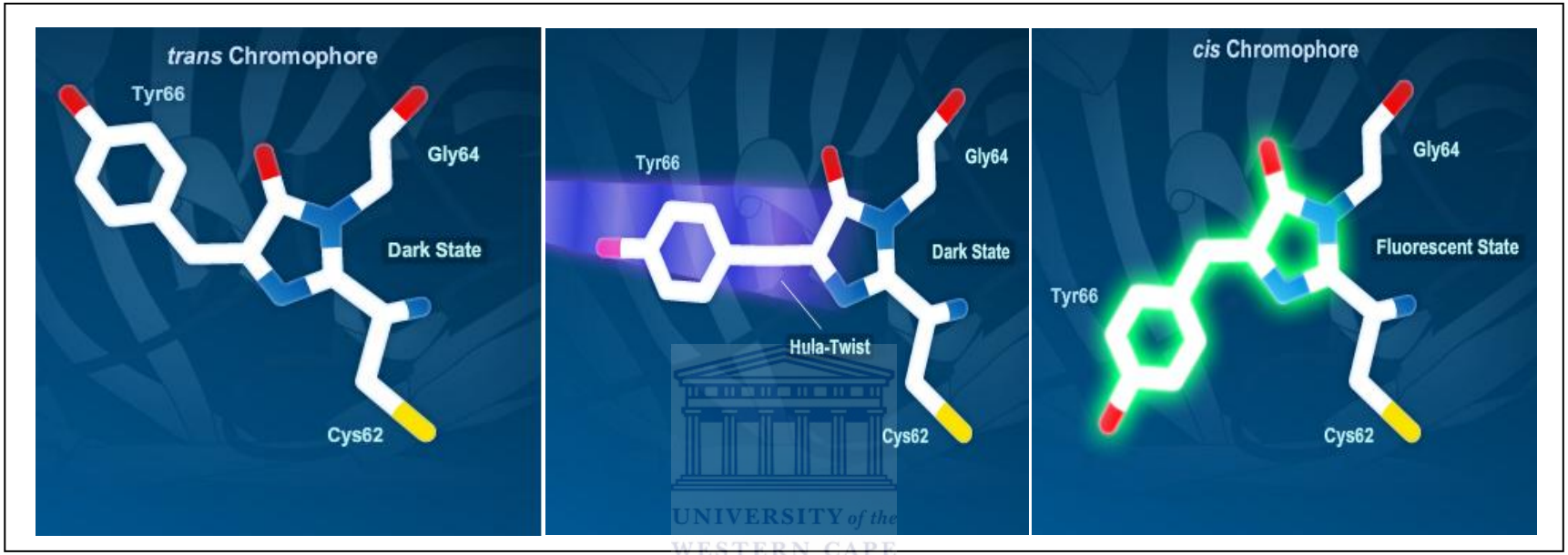


Figure 4. Chromophore state during reversible switching in Dronpa. From left to right: the mature chromophore before laser excitation; (middle) at 405 nm illumination inducing a *trans/cis* rotation; (far right) the resulting green fluorescence. Illumination with light at 488 nm results in a cis/trans isomerisation resulting in loss of fluorescence. Images adapted zeiss (http://zeiss-campus.magnet.fsu.edu/articles/probes/index.html)

1.4. Applications of fluorescent proteins

Research in molecular biology and related fields has been accelerated through the use of fluorescent proteins as non-invasive probes. Researchers are able to investigate cellular processes in living systems using fluorescence microscopy and novel molecular techniques. FPs are used to monitor the movements and interactions of individual or groups of protein(s), cell(s) and whole organisms. GFP and its spectral mutants are also commonly used as tools to quantitate experimental data (Chudakov et al. 2005).

Typical experiments include quantitative gene expression, detection and quantification of promoter activity (Figure 5), protein localization and interactions (Figure 6), cell division, intracellular transport pathways, and chromosome replication and organisation (March et al. 2003; Verkhusha and Lukyanov 2004; Wachter, 2006; Ehrenberg, 2008). Additional examples of FPs applications, as shown in Table 1, include their use as biosensors for monitoring pH, Ca^{2+}(Stepanenko et al. 2008) or arsenic concentrations (Hu et al. 2010). The application of FPs in imaging has greatly improved spatial resolution of cellular structures (Chudakov et al. 2007).

The success of FPs as molecular markers can be attributed to their non-toxicity, small monomeric size, high stability, lack of requirement of external co-factors besides oxygen and the capacity to be engineered as fusion constructs (March et al. 2003). The natural oligomeric state, low intrinsic fluorescent intensity and overlapping excitation and emission spectra are, however, obstacles to the use of FPs as tools in molecular biology.

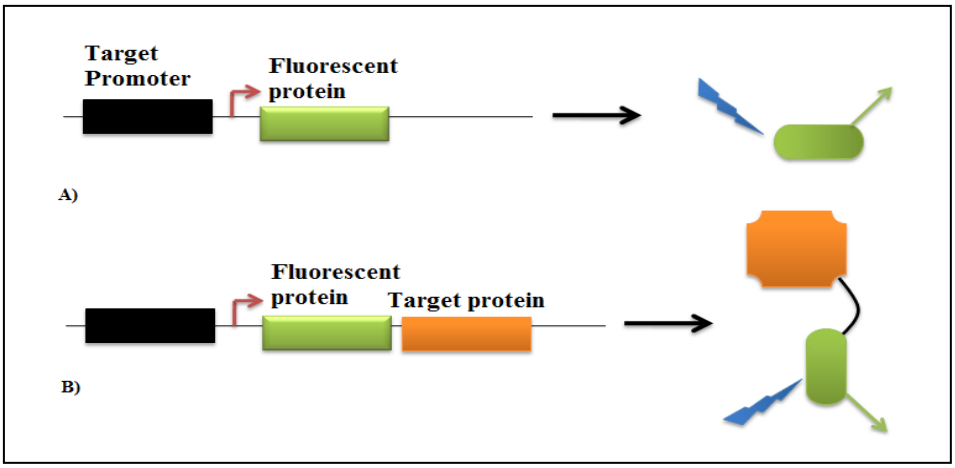


Figure 5. Applications of GFP and FPs. A) use of fluorescent markers in detection of promoter activity and B) protein tagging for monitoring heterologous expression or localisation in live cell imaging. Adapted from Chudakov et al. 2005.

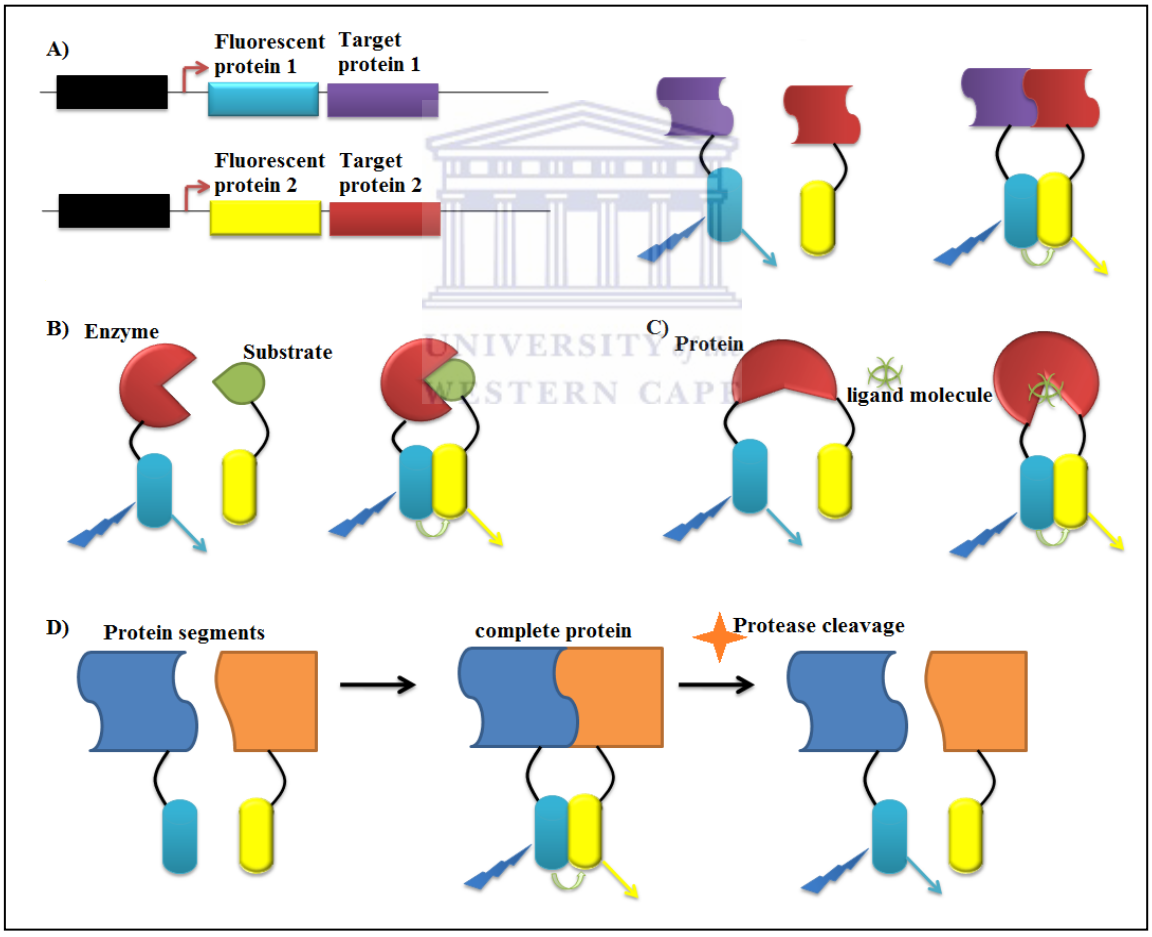


Figure 6. The use of fluorescent markers in FRET analysis to explore A) protein-protein interactions B) enzyme substrate interactions or enzyme activity B) ligand interactions C) substrate ligand interactions and D) protease activity. Adapted from Chudakov et al. 2005.

Table 1 Examples of fluorescent proteins as tools in molecular biology

Application	Experimental procedure and results	Reference
Biosensor for arsenic	The ArsR and its promoter region were cloned as a fusion construct into a vector containing phiYFP. The system was able to detect arsenite (As^{3+}) and arsenate (As^{5+}) in a time and dose-dependent manner.	(Hu et al. 2010)
FACS and SIGEX high-throughput screening	A novel technique for high-throughput screening of catabolic genes was developed. An operon trap-GFP vector was used to create a shotgun clone library which was screened by substrate induced catabolic gene expression in a semi-automated fashion using fluorescent activated cell sorting.	(Uchiyama et al. 2005; Yun and Ryu 2005)
Biosensor for pathogenic *Listeria monocytogenes*	A sensitive screen to distinguish various strains of *L. monocytogenes* was developed by mating different fluorescent vector constructs with bacteria isolated from a sample. This screen can be used under non-sterile conditions to distinguish between various strains of *L. monoctogenes*.	(Andersen et al. 2006)
Detection of DNA methylation states	A fluorescent protein was split into two sections; one half was fused to target DNA and the other to the *trans* interacting element. If the two meet the fluorescent protein would form a functional unit and the interaction would be detected by fluorescent microscopy. If the site was methylated no fusion would occur and no fluorescence would be observed.	(Furman et al. 2009)

1.5. Chromophore characterisation, synthesis and structure

1.5.1. Chromophore characterisation and synthesis

All GFP homologs, including anthozoan fluorescent and chromoproteins, synthesize a chromophore capable of fluorescence or vivid colour without any external cofactors other than oxygen (Inouye and Tsuji 1994; Cody et al. 1993; Ehrenberg, 2008). The chromophore is formed through multiple autocatalytic reactions of an internal tripeptide Xaa-Tyr-Gly (Figure 7) to generate a *p*-hydroxybenzylidene imidazolinone functional group (Tsien, 1998).

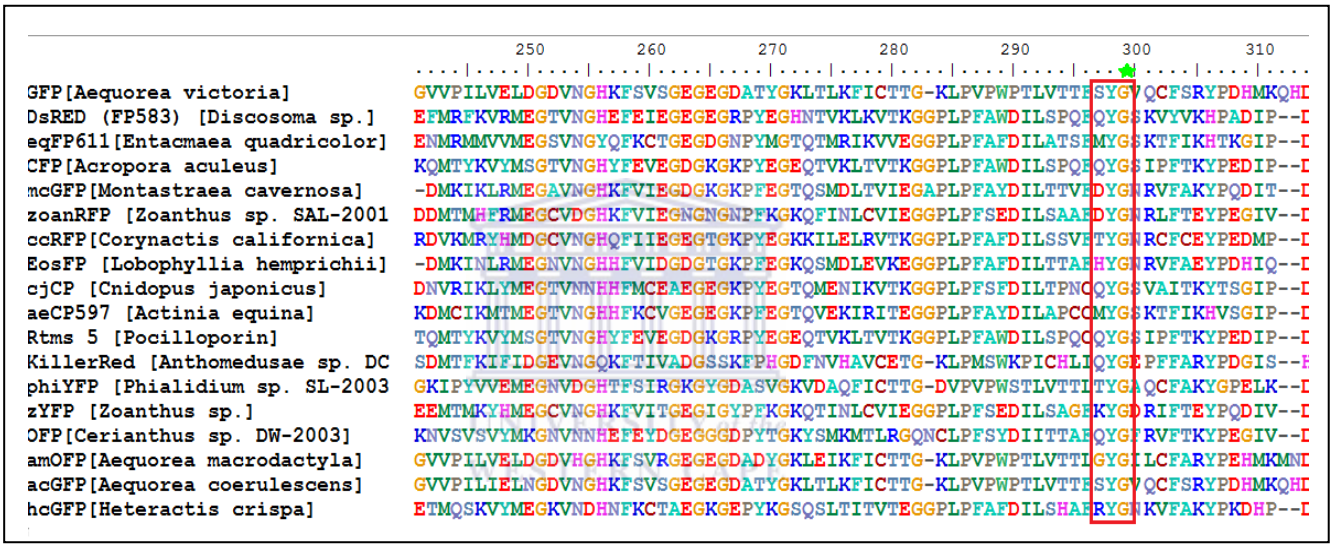


Figure 7. ClustalX alignment of protein sequence from fluorescent and chromoproteins across *Cnidarian* spp. The red box highlights the chromophore residues, while the green star indicates the highly conserved Gly residue.

The chromophore of GFP was first isolated in 1993 (Tsien, 1998) and its chemistry described in 1996 (Niwa et al. 1996). Models describing the chromophore synthesis were published as early as 1994 (Heim et al. 1994; Cubitt et al. 1995). Experimental studies to verify the models utilized sequential kinetics to resolve the steps involved in chromophore synthesis (Reid and Flynn 1997; Tsien, 1998). Recent studies made use of site directed mutagenesis and high-resolution crystallography to test the hypotheses and ascertain whether or not the proposed models are accurate (Barondeau et al. 2003; Barondeau et al. 2005).

The process of autocatalytic chromophore formation as per the original model by Heim and co-workers is illustrated in Figure 6. The protein folds into its native 3D conformation after which the tripeptide amino acid sequence (Xaa-Tyr-Gly) undergoes torsional rearrangement and cyclization. In GFP this rearrangement facilitates nucleophilic attack of the amide nitrogen of Gly on the carbonyl carbon of Ser (representing Xaa in the tripeptide sequence; see Table 2) to form an imidazolinone heterocyclic ring system. The newly formed double bond is dehydrated and the imidazolinone is conjugated to the phenol ring of Tyr by reducing the *alpha/beta* carbon bond with the oxygen as an electron acceptor. This generates a *pi*-conjugated system due to the delocalized electrons that are part of, and in the surrounding environment of, the chromophore (Heim et al. 1994; Reid and Flynn 1997; Tsien, 1998; Barondeau et al. 2003; Barondeau et al. 2005).

The wavelength of light that is absorbed is dependent on local chemistry and amino acid substitutions in and near the chromophore (Turcic et al. 2006; Miyawaki et al. 2003), as well as the planarity of the chromophore. Mature red chromophores and red shift optical highlighters undergo an additional oxidation reaction that extends the *pi*-conjuation system of the chromophore which results in a colour shift to red fluorescence (Miyawaki et al. 2003).

It is thought that chromophores emitting light in the range 500-525 nm (green light) are in the simplest form. This is supported by the fact that yellow, orange and red chromophores all contain a green intermediate chromophore. This may also be why it has been suggested the cyan fluorescent protein, CFP, (and its chromophore) are the result of independent convergent evolution. The CFP chromophore has a structure similar to the GFP's chromophore, however, none of the other (YFP, OFP or RFP) proteins are observed to have this chromophore as an intermediate (Shagin et al. 2004).

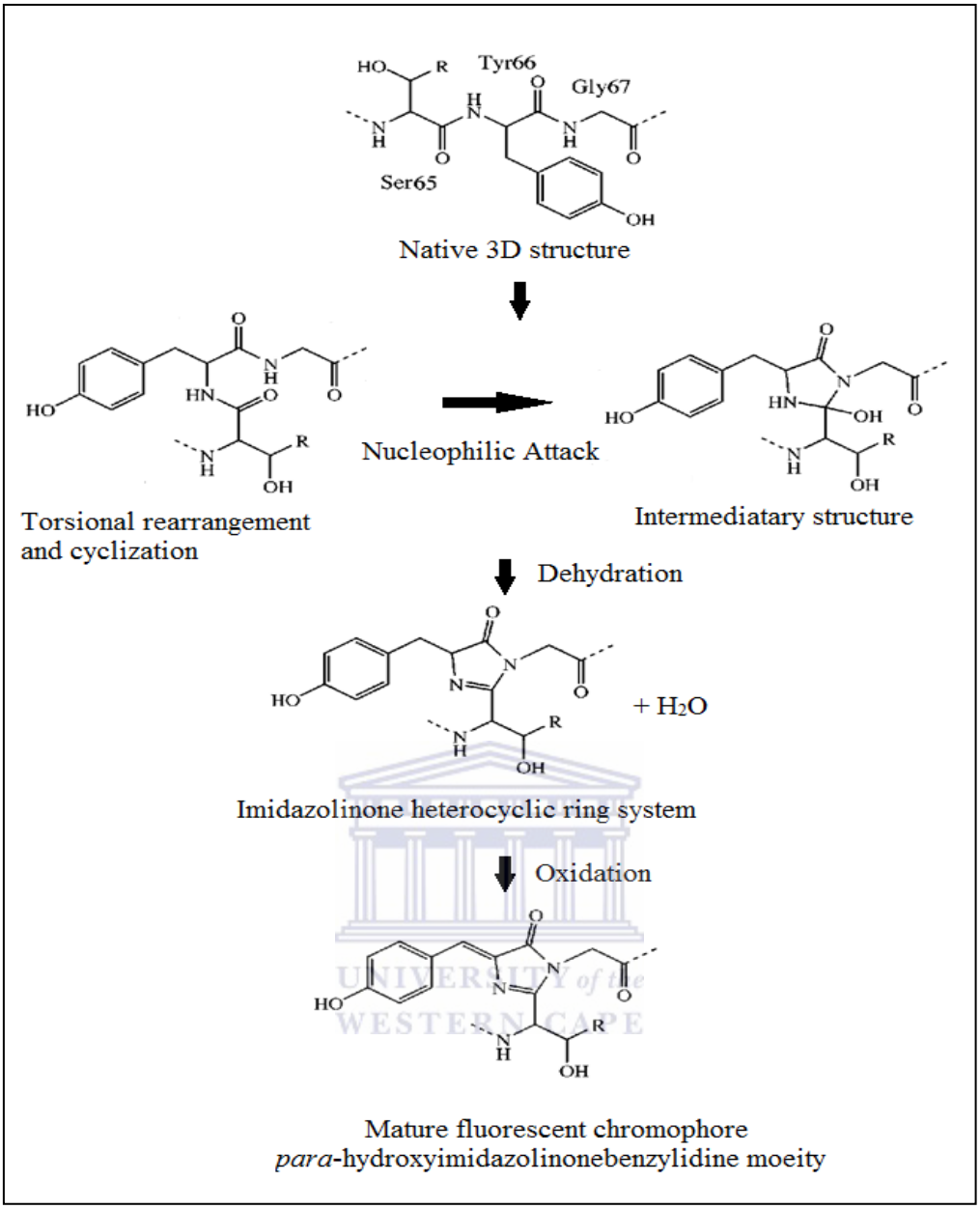


Figure 8. The autocatalytic process of chromophore formation. The above model is based on GFP isolated from *A. victoria*, adapted from Tsien (1998).

1.5.2. Structural differences between FP and CP chromophores

The chromophore structure of chromoproteins and fluorescent proteins differs across all phyla. Chromoproteins possess a *trans non-coplanar* chromophore while, with one exception (eqFP611), fluorescent proteins have a *cis coplanar* chromophore (Shagin et al. 2004). The chromophore structure of Anthozoan fluorescent proteins can be divided into four groups: GFP-like (Figure 9), DsRED-type (Figure 10), Optical Highlighters (Figures 3 and 4) and

zoanFP538 (Figure 9) (Cody et al. 1993; Inouye and Tsuji 1994; Field et al. 2006; Alieva et al. 2008; Ehrenberg, 2008).

Typically cyan and green fluorescent proteins in Anthozoa have a GFP-like chromophore structure (Figure 9). ZoanFP538, the only naturally occurring anthozoan yellow fluorescent protein, has a unique 3 ring chromophore structure as illustrated in Figure 9. This is due to a secondary nucleophilic attack, which occurs after dehydration and before oxidation that results in cleavage of the polypeptide backbone and yellow fluorescence (Alieva et al. 2008). The red chromophores contain two subsets: those that have a DsRED chromophore and then the eqFP611 chromophore, which is a unique exception. The difference between the two is illustrated in Figure 10.

Additional research on Scyphozoan/hydroid fluorescent and chromoproteins led to the discovery of a novel yellow fluorescent protein (phiYFP) and a purple chromoprotein (anm2CP) (Yanushevich et al. 2005). Furthermore, a mutant of *A. victoria* emitting in the red spectrum was recently engineered in addition to those that already exist; and the GFP mutants now cover the entire visible light spectrum (Mishin et al. 2008). Thus there is a high degree of spectral diversity in groups other than Anthozoans. However, as the chromophores are highly conserved, there is not a high degree of structural diversity of the chromophores of hydroid FPs and CPs.

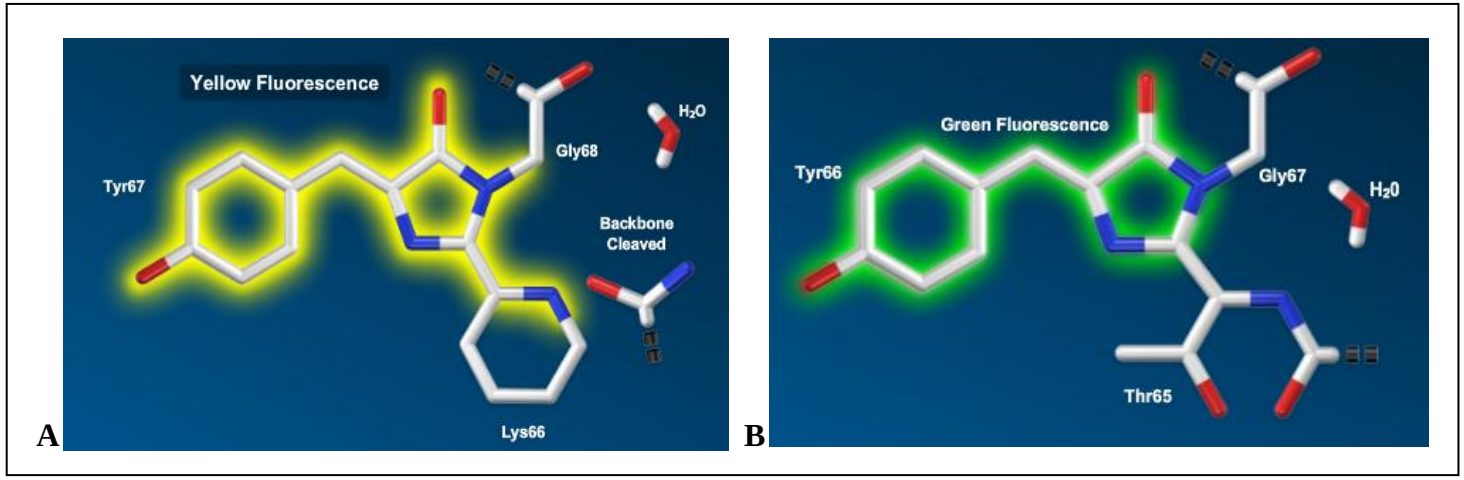


Figure 9. Mature chromophore structures of a YFP and GFP. A) zoanFP538 has a unique three-ring system chromophore, a stark contrast in comparison to B) the GFP-like chromophore structure which is observed in all hydroid (jelly), green and cyan chromophore structures. Images adapted zeiss (http://zeiss-campus.magnet.fsu.edu/articles/probes/index.html)

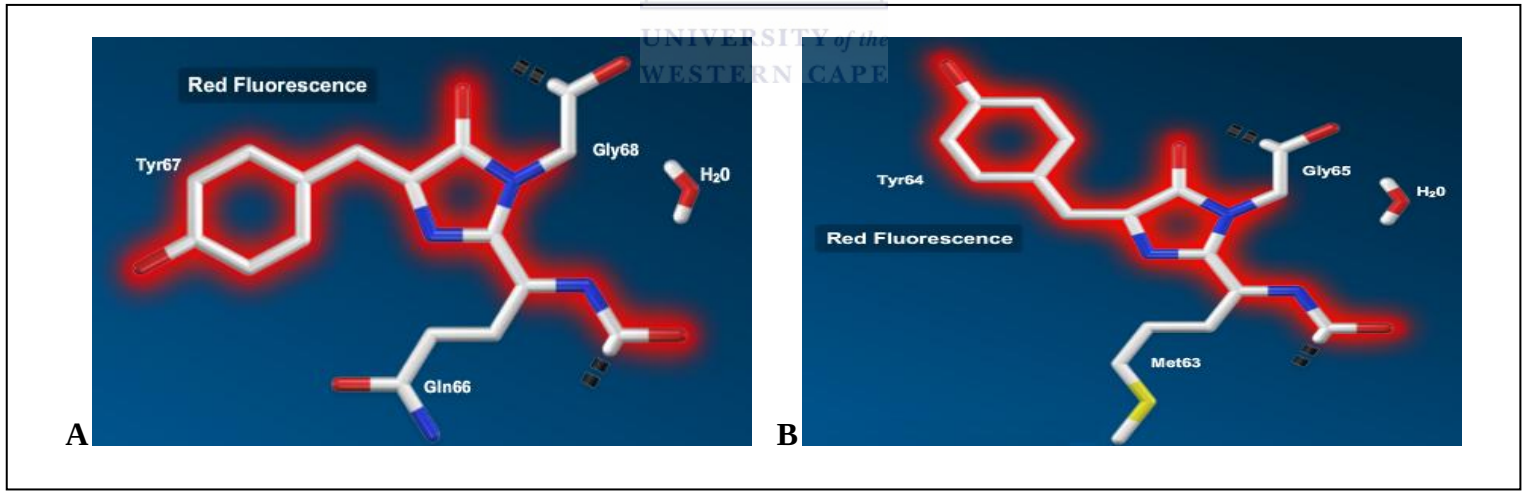


Figure 10. The structural differences in RFP chromophores. A) The chormophore of DsRED is *cis coplanar* and B) the chromophore of the far-red eqFP611*trans coplanar*. Both RFPs were isolated from Anthozoans. Images adapted zeiss (http://zeiss-campus.magnet.fsu.edu/articles/probes/index.html)

Table 2 Spectral and structural characteristics of fluorescent and chromoproteins

	Excitation	Emission	Tripeptide	Features	Structure	Reference
Blue fluorescent proteins (440-470nm)						
EBFP	383	445	Ser-His-Gly	FRET partner for GFP	Monomer*	(Patterson et al. 1997;Yang et al. 1998)
mTagBlue				Brightest blue probe	Monomer*	(Subach et al. 2008)
Cyan (470-500nm)						
ECFP	439	476	Ser-Trp-Gly	Trp instead of Tyr	Monomer*	(Cubitt et al. 1995; Heim and Tsien 1996)
AmCyan1	454	486	Lys-Tyr-Gly	Model of FP evolutionary studies	Tetramer	(Matz et al. 1999)
MiCy	440	495	Gln-Tyr-Gly	Only dimeric AFP	Dimer	(Karasawa et al. 2004)
Green (500-520nm)						
GFP	395 475	508	Ser-Tyr-Gly	The first fluorescent protein discovered	Weak dimer	(Heim et al. 1994; Cubitt et al. 1995)
EGFP	488	508	Thr-Tyr-Gly	"Gold standard" and most widely used fluorescent protein	Monomer*	(Cubitt et al. 1995)
Amanzi-green	492	505	Gln-Tyr-Gly	< 6% sequence homology to GFP	Tetramer	(Karasawa et al. 2003)
ZoanGFP	496	506	Asn-Tyr-Gly	Novel chromophore structure	Tetramer	(Matz et al. 1999)

Yellow (525-555nm)						
EYFP	514	527	Ser-Tyr-Gly	Mutations not of chromophore tripeptide	Monomer*	(Nagai et al. 2002)
zYFP538	528	538	Lys-Tyr-Gly	Novel chromophore structure	Tetramer	(Matz et al. 1999; Pletneva et al. 2007)
PhiYFP	525	537	Thr-Tyr-Gly	Unique example of a mutagenesis study and natural protein structure overlapping	Monomer*	(Shagin et al. 2004)
Orange (555-580nm)						
Kusabira orange	548	559	Cys-Tyr-Gly	FRET partner for CFPs	Tetramer	(Karasawa et al. 2004)
OFP	548	573	Gln-Tyr-Gly	FRET partner for CFPs	Tetramer	(Ip et al. 2007)
Red (580-630nm)						
DsRed	558	583	Gln-Tyr-Gly	Most commercially valuable RFP	Tetramer	(Matz et al. 1999)
zRFP574	553	574	Asp-Tyr-Gly	Novel chromophore structure	Tetramer	(Matz et al. 1999; (N. Pletneva et al. 2006; N. Pletneva et al. 2007)
eqFP611	595	611	Met-Tyr-Gly	The most far red RFP naturally occurring	Tetramer	(Wiedenmann et al. 2002)
Chromoprotein						

asCP	595		Met-Tyr-Gly	Kindles naturally	Tetramer	(Lukyanov et al. 2000)
Rtms5	591		Gln-Tyr-Gly	Unique environment surrounding chromophore	Tetramer	(Beddoe et al. 2003; Turcic et al. 2006)
gtCP	580		Gln-Tyr-Gly	coral CP	Tetramer	(Gurskaya et al. 2001; Martynov et al. 2001)
aeCP597	597		Met-Tyr-Gly	Positive control used in this study	Tetramer	(Poponov et al. 2005)
hcCP	578		Glu-Tyr-Gly	Mutated to Optical Highlighter	Tetramer	(Gurskaya et al. 2001)
cjBlue	610		Gln-Tyr-Gly	Longest absorption spectra	Tetramer	(Chan et al. 2006)
Optical Highlighters						
Kaede	508 572	518 580	His-Tyr-Gly	Green to red fluorescent shift	Tetramer	(Ando et al. 2002)
Eos	505 569	581 600	His-Tyr-Gly	Green to red	Tetramer	(Wiedenmann et al. 2004; Nienhaus et al. 2006)
Dronpa	503	518	Cys-Tyr-Gly	Dark to green fluorescent reversible shift and natural monomer	Monomer	(Ando et al. 2004)
HcRED	598	645	Glu-Tyr-Gly	Dark to red fluorescent kindling	Tetramer	(Gurskaya et al. 2001)
KFP	580	600	Met-Tyr-Gly	Dark to red kindling reaction	Tetramer	(Yanushevich et al. 2003)

*Engineered to monomeric state; Red text indicates Sycphozoan/Hydroid FPs

1.6. Spectral diversity of anthozoan FP and CP

The range of visible light extends from ~400 – 780 nm. The fluorescent proteins emitting in this range are blue (440-470 nm), cyan (470–500 nm), green (500-525 nm), yellow (525-550 nm), orange (555-580 nm) and red (< 580 nm). Blue and cyan FP are excited by near UV-range and blue spectrums of light, green FPs by blue light, yellow FPs by cyan light, orange and red FPs by green and yellow light, respectively (Wiedenmann et al. 2002).

No naturally occurring BFPs have yet been discovered. The first engineered BFP was an *A. victoria* derivative and was designed as a FRET partner for GFP (Yang et al. 1998). The major concerns with using BFPs in research are background autofluorescence, light scattering and the use of UV-excitation as this is detrimental to living organisms even in low doses (Olenych et al. 2005; Shaner et al. 2007).

As for BFPs, autofluorescent background is a challenge to applications employing CFP and GFPs. The most noteworthy feature of CFP is its application in FRET analysis together with YFPs and OFPs. Despite being very similar to GFPs, CFPs are thought to have evolved separately (as discussed previously) and this is why they have been used in evolutionary FP studies (Henderson and Remington 2005). GFPs are the most abundantly dispersed FP throughout all the species studied (Alieva et al. 2008). This group is the most widely used molecular probe and EGFP has long been the "gold standard" to which the fluorescent brightness of all other FPs are compared (Heim and Tsien 1996).

ZoanFP538 is the only existing anthozoan YFP and emits at 538 nm, the midpoint of the overlapping emission spectra of GFP and RFPs. This may be why YFPs are potentially considered to be the most versatile FP for use as probes and very important for FRET analysis due to their ability to couple GFP to RFP as an intermediate (Shaner et al. 2007).

Two naturally occurring OFPs, namely Kusabira orange and orange fluorescent protein (see Table 2), have been documented thus far and both are from anthozoan species. These OFPs

are used in FRET analysis together with CFPs, but besides this no other outstanding usable features have been highlighted (Karasawa et al. 2004; Chudakov et al. 2005; Ip et al. 2007). A bright stable far red fluorescent protein would be the pinnacle of FP technology for imaging deep tissues since RFPs have low autofluorescence and there would be little background interference from cells. This group of FPs also displays the potential for use in multi-labelled FRET analysis in combination with cyan, green and yellow FPs (Fradkov et al. 2002).

Blue and purple chromoproteins contain a similar *trans* planar chromophore to eqFP611. They have also been successfully mutated to produce FPs that emit further in the red spectrum, perhaps due to the long wavelength absorption that they demonstrate (Shaner et al. 2007). Future application of ACP may include fusion targets that are monitored in a stable and permanent fashion (Yanushevich et al. 2003; Shkrob et al. 2008). Finally, each class of fluorescent protein adds another dimension to multiple labelled imaging (Krishna and Ingole 2009).

1.7. Structure

Fluorescent and chromoproteins share a general tertiary structure known as a β-can structure. This structure is composed of 11 anti-parallel β-sheets which form a cylinder through which an α-helix containing the tripeptide sequence (responsible for fluorescence/colour) is threaded (Figure 11 A) (Ormo et al. 1996). Anthozoan fluorescent and chromoproteins are different from GFP in that they have a quaternary structure (Figure 11 B). The tetramer of AFP/ACP is a dimer of dimers where each monomeric subunit forms contacts with 2 other subunits. These contacts play a role in stabilizing the *pi*-conjugation structures of the chromophores (Nienhaus et al. 2006).

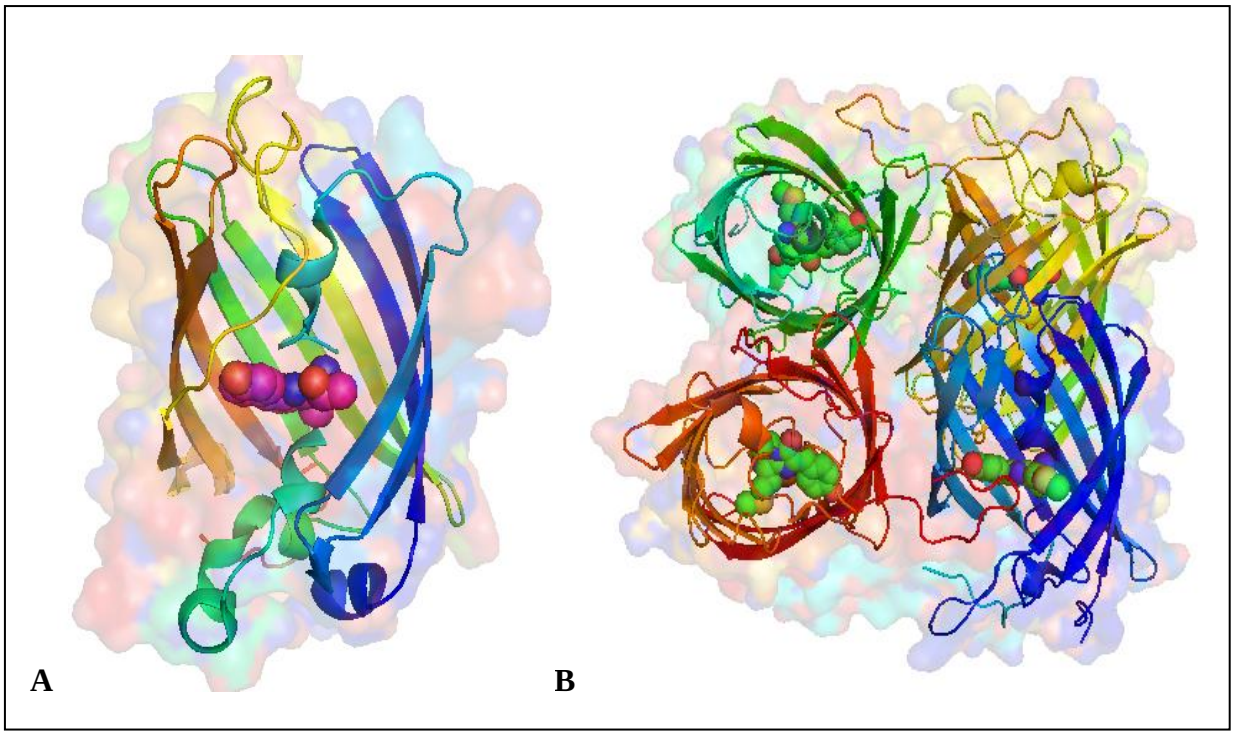


Figure 11. Tertiary and quaternary structure of hydroid and anthozoan FPs. A) General structure of FP (2Y0G from PDB). Anti-parallel β-Sheets form a barrel through which an α-helix containing the chromophore moiety (ball stick structure), is threaded. B) Anthozoan FPs, such as the red FP (3E5W) from *Entacmaea quadricolor*, have a quarternary structure as opposed to the monomeric or dimeric structure of hydroid FPs.

1.8. Speculated function

Exact functions of fluorescent and chromoproteins are not known. It is thought that fluorescent proteins and chromoproteins have distinctive purposes in their respective host organisms and that within each group a variety of functions are supported.

Speculative functions have included signalling and communication (Field et al. 2006; Gruber et al. 2008). This may, however, be speculation as a result of inferring function based on bioluminescent systems. The possibility that FPs are used as a UV filter-protection for endosymbionts, particularly in the case of GFP which occurs in high abundances, is another function that may be proven in future. However, examples where lower concentrations of FPs are observed have yet to be assigned a potential function (Salih et al. 2000; Labas et al. 2002).

More recently, suggested functions of fluorescent proteins included roles in the host stress and innate immune response as antioxidants or oxygen radical quenchers (Palmer et al. 2009). One study found that purified FP extracts from infected or stressed tissues were able to scavenge harmful reactive oxygen species. Another study suggested that FPs generally have the ability to act as light-induced electron donors (Bogdanov et al. 2009; Bogdanov et al. 2010). The research showed that a diverse range of FPs was able to undergo photoreduction and photooxidation, depending on external factors such as oxygen concentration and the presence of appropriate electron donors or acceptors. The study included biologically relevant examples such as NAD+, FAD and CytC, all of which were reduced through the photooxidation activity of EGFP. Though promising, much more work is needed to infer biological function.

1.9. Engineering fluorescent and chromoproteins

The majority of FP and CP are not innately suitable for biotechnological applications. Specifically, the tetrameric structure of AFP and ACPs places a large metabolic burden on the heterologous host (March et al. 2003). Additionally fluorescent brightness, stability, maturation rates and temperature sensitivity are not always ideal for deep imaging (or imaging in general) when background autofluorescence is high (Heim and Tsien 1996; Karasawa et al. 2003; Shaner et al. 2007; Ehrenberg 2008). To improve the suitability of AFP and ACP for applications in molecular biology, techniques such as random mutagenesis (generally error-prone PCR) and directed evolution or a combination of both are used to engineer more robust fluorescent probes (Loos et al. 2006; Baird et al. 2000; Müller-taubenberger and Anderson 2007). Due to the intrinsically long wavelength absorption maxima of chromoproteins, these are ideal targets for creating far-red FP by conversion using site-directed and random PCR mutagenesis (Gurskaya et al. 2001; Bulina et al. 2002;

Poponov et al. 2005; Wilmann et al. 2005; Shkrob et al. 2008; Piatkevich and Verkhusha 2009).

To generate monomeric FPs, site directed mutagenesis, principally at the C-terminus of FP (Chudakov et al. 2005; Stepanenko et al. 2008), is used to disrupt the oligomerisation of the tetramer. However, as the quaternary structure is important for chromophore stabilization this generally results in a loss of fluorescence. To recover fluorescence, rescue mutagenesis, generally through error prone PCR, is performed and desired mutants are selected for further enhancements. The first mutant monomeric anthozoan FP was mRFP1, a DsRED derivative which was obtained after 33 mutations (Nienhaus et al. 2006; Shaner et al. 2007).

An unusual observation was made when phiYFP, the yellow fluorescent protein found in the jellyfish *Phillidium*, was discovered. Mutations induced in the lab resulting in an *A. victoria* GFP derivative, EYFP (see Table 2), were found to match the natural amino acid positions in native phiYFP. Furthermore, the crystal structure of the phiYFP indicated that the chromophore structure was the same as that of EYFP (Yanushevich et al. 2005).

A survey of the literature suggests that despite being well characterised, FP and CP are poorly understood in terms of biological context. A deeper understanding of the biological importance of the various structural states of both the protein and chromophore will greatly benefit mutagenesis efforts.

1.10. Aims and objectives

South African coastal waters are rich in sea anemone species which have largely been unexploited in biotechnology. FPs and CPs are present in most anemone species and the opportunity therefore, exists to find novel FP/CPs which could contribute to the South African blue biotechnology industry. Furthermore, characterisation of the FP/CPs isolated

from the organisms could lead to an increase in our understanding of their *in vivo* functions. The aims of this project were to:

1. Generate cDNA gene banks of coloured or fluorescent sea anemones using SMARTer^{TM} In-Fusion^{TM} technology.
2. Screen the libraries for fluorescent and chromoproteins based on sequence homology and functional characteristics.
3. Characterize the spectrophotometric and properties of isolated FP/CPs.
4. Identification of 18S rRNA and 16S rRNA products from sea anemones and associated symbionts

2. Chapter 2 - South African sea anemones of the *Order Actinaria*: Phylogeny and associated symbionts

Grouping or classification of organisms based on their DNA sequences, phylogenetics, has become a popular technique in taxonomy. The question of which DNA sequences to use has, however, been a constant debate. A good phylogenetic marker must meet various criteria. Firstly the marker must be conserved in all the organisms to be studied and contain variable regions for defining closely related species. Examples of popular phylogenetic markers include 16S / 23S (prokaryotic studies) and 18S / 28S (eukaryotic studies) rRNA genes, Cytochrome genes and 12S or 16S mitochondrial rRNA genes (Daly et al. 2008).

Ribosomal RNA genes serve as valuable phylogenetic markers. They are ubiquitously found in all cellular life-forms and have been used to infer universal relationships (Head et al. 1998). Using rDNA sequences as phylogenetic markers does, however, have drawbacks such as rDNA heterogeneity. Furthermore, the use of molecular typing is hampered by PCR bias and artifacts, contaminating DNA, interpretation of the data and relating it to existing taxonomic relationships (Ashelford et al. 2006; Daly et al. 2008; Haas et al. 2011).

Amplified ribosomal DNA restriction analysis (ARDRA) is a technique used to visualize variations in groups of similar sequences. The method relies on amplification of the rRNA gene and digestion of the product with various restriction endonucleases. This generates a unique banding pattern or fingerprint (Stakenborg et al. 2005). The fingerprint assists in preliminary identification as the banding pattern of an unknown sample may be compared to those of a reference organism. The weakness of this technique is that one pattern may represent more than one species-type or that two different patterns may exist for the same species-type and thus the results must be verified by sequencing (Dahllöf, 2002).

In this study, the phylogenetic relationships of *Pseudactina varia*, *Pseudactina flagellifera* and *Bunodosoma capensis* sea anemones from South African coastal waters were examined

using the 18S rRNA gene as a marker. Furthermore, 16S rRNA clone libraries provided a rudimentary evaluation of observed bacteria associated with the sea anemones.

2.1. Materials and Methods

2.1.1. Sample collection

Sea anemones were collected (from South African coastal waters) and stored by the UCT zoology dive team under the supervision of Andrea Plos. Once collected from UCT, samples were kept in purified sea water at room temperature overnight before dissection.

2.1.2. General microbiology

The bacterial strains, plasmids and primers used in this study are shown in Table 5 (Appendix B). *Esherichia coli* strains were grown on LB agar or in LB broth supplemented with appropriate antibiotic (see Appendix A). Strains were incubated at 37 °C for ~16 hours, unless otherwise stated. When incubating strains in broth cultures, the broth cultures were agitated by shaking at 150 to 225 rpm. Glycerol stocks were prepared in medium containing 50% glycerol and stored at -80°C.

2.1.2.1. Electrocompetent cells

GeneHogs and BL21 *E. coli* were streaked out to obtain single colonies. A starter culture, originating from a single colony that was inoculated in LB broth and grown overnight at 37 °C, was diluted 1 in 100 in fresh preheated LB broth. Cells were grown at 37 °C until early log phase (OD_{600} 0.4-0.45), chilled on ice for 30 minutes and then harvested at 4000 *g* for 15 min at 4 °C. The cells were made competent by washing them in ice cold sterile water 4 times followed by a wash in ice cold 8.7 % glycerol. All washing steps as well as aliquoting were done in a cold room on ice. Final aliquots were snap frozen in liquid nitrogen and stored at -80°C.

2.1.2.2. Electroporation

Electrocompetent cells were thawed on ice and upto 5µL of dialysed ligation was used in the standard reaction along with appropriate controls. Cells were electroporated using the GenePulser (Bio-Rad) with the following settings: 1.8 Volts, 200 Ohms (Ω) and 2.5 µFarads according to the manufacturer's instructions. Immediately after electroporation 950µL of SOC was added and cells were recovered at 37°C with shaking for 1 hour. To observe transformed colonies for screening, 100 µL of transformation mix was plated onto appropriate antibiotic containing media.

2.1.3. General molecular techniques

2.1.3.1. Genomic DNA extraction

Tissue samples were dissected and stored in RNA*later* according to manufacturer's instructions. To extract genomic DNA (gDNA), ~100mg of tissue was removed and washed 3 times using 1 X TE to remove any chemicals that could inhibit further downstream applications. The gDNA extraction protocol used was a modified version of the Pearson et al. 2002 protocol: Tissue was disrupted in liquid nitrogen using a pestle and mortar. Powdered tissue was homogenised in 500µL of lysis buffer and incubated at 37 °C for 30 minutes, followed by the addition of 300µg of Proteinase K. The samples were incubated for a further 2 hours 30 minutes with frequent mixing. The lysate was cleared and the supernatant extracted with equal volume of phenol and chloroform: iso-amyl alcohol (24:1). The phases were separated by centrifugation at 12 000 *g* for 10 minutes at 18°C and the aqueous phase was transferred and re-extracted with 500µL of chloroform: iso-amyl alcohol. The phases were separated as before and the aqueous phase transferred to a sterile tube where 1/10 volume 3M sodium acetate (pH 5.5) and 2.5 volumes ice-cold absolute ethanol were added. After gentle mixing, precipitated gDNA was transferred using a sterile bent Pasteur

pipette to 70 % ethanol and washed. gDNA was resuspended in 1X TE pH 8.0 and quantified (Pearson et al. 2002).

2.1.3.2. Agarose gel electrophoresis

The appropriate amount of agarose was weighed out and added to 1 X TAE and boiled until dissolved. The mix was cooled to ~50°C and ethidium bromide to a final concentration of 1 µg/mL was added after which the gel was cast and set. Gels were electrophoresised at 8 V/cm for approximately one hour.

2.1.3.3. Polymerase Chain Reaction (PCR)

For each PCR the reaction contained 1 X appropriate buffer, 0.2-0.8 µM of primers, 0.2-0.4 mM dNTP and appropriate units of DNA polymerase to amplify the respective target. PCR parameters included initial denaturation at 94°C for 2-5 minutes followed by 25-35 cycles of denaturation at 94°C for 30-45 seconds, annealing at variable temperatures appropriate for each primer set and elongation at 72°C (for the approximate time calculated as per Kbp/second for the polymerase used) and final elongation at 72°C for 5-20 minutes.

2.1.3.4. General DNA manipulation techniques

Plasmid isolations for sequencing were performed with Qiagen mini-prep kit (Qiagen) according to manufacturer's specifications. All other plasmid preparations were extracted using the alkaline lysis protocol. All digests were performed with enzymes from Fermentas (Inqaba Biotech, SA, Jhb), according to manufacturer's instructions, unless otherwise stated. DNA was quantified on the NanoDrop ND8000 (Thermo Scientific, USA,) according to the manufacturer's instructions. All DNA purifications were performed using the Nucleospin extract II kit (Machery-Nagal) according to manufacturer's instructions.

2.1.4. Construction of 16S and 18S rRNA clone libraries

Genomic DNA was extracted as in section 2.1.3.1 and analysed on gels cast as in 2.1.3.2. The gDNA extraction protocol of Pearson et al. (2002) was modified to extract

polysaccharide-free genomic DNA from sea anemones (Figure 12). Decreasing the amount and time of sample incubation from 72 hours to 2½ hours in the presence of SDS during the lysis resulted in a decrease in the amount of co-extracted polysaccharides that was co-extracted. The purified gDNA was suitable for downstream applications such as PCR.

To survey the associated microbial diversity and to identify closely related sea anemone species, gDNA was extracted from tissue of the foot or wall from each sea anemone and subjected to 18S rRNA and 16S rRNA PCR. The 16S and 18S rRNA genes were amplified using the F1 and R5 and EukA and EukB primer sets, respectively. PCR reactions to amplify the 18S and 16S rRNA genes were performed as described in section 2.1.3.3. The template amount varied in the PCR reactions: 10 ng of gDNA was used in the 16S rRNA and 100 ng in the 18S rRNA; DNA template, forward and reverse primers were substituted with mQ water in control reactions. General microbiology and DNA molecular techniques and plasmid isolations were performed as in sections 2.1.2 and 2.1.3.4. Sequencing was done by Macrogen (Korea).

2.1.5. DNA sequence analysis

Representative DNA sequences of the *Order Actinaria* were retrieved from GenBank (http://www.ncbi.nlm.nih.gov/genbank/) and compared to the 18S rRNA sequences from the sea anemone clone libraries. Nucleotide sequences were aligned using MAFFT online tool (http://mafft.cbrc.jp/alignment/server/) I-GNS-i algorithm and 20PAM/k=2 model. The resulting alignment was used to construct a phylogenetic tree in MEGA 4.0 using the Neighbor-joining algorithm under the Maximum Composite Likelihood model and a gamma distribution of 1.0 performed with a bootstrap value of 1000. BLAST analysis was done using sequences of approximately 1.5kb with sequence coverage of 97-100% and yielded matches between 94-99%.

2.2. Results: Phylogeny and associated symbionts of South African sea anemones

2.2.1. Construction of diversity clone libraries

Except for *P. flagellifera*, a single product of ~1.5kb and ~1.8kb were obtained when the 16S and 18S rRNA genes were amplified, respectively (Figure 13). The 16S rRNA amplicons are expectedly fainter than the 18S rRNA amplicons because of the gDNA extraction protocol and tissue used. The amplified products were cloned into pGEM-T-easy, transformed into *E. coli* GeneHogs and spread onto LB agar supplemented with 100 µg/µL ampicillin, 80 µg/µL X-gal and 1 mM IPTG. Following blue white selection, 48 putative clones from each library were verified by M13 PCR (see Figure 42 in Appendix D). The resulting amplicons were subjected to ARDRA using Alu*I* and resolved on a 2 % agarose gel. 38 unique ribotypes (22 for the 18S and 16 for the 16S) 17 (*B. capensis*) 9 (*P. varia*) and 12 (*P. flagellifera*) were sequenced to obtain phylogenetic data and used in subsequent analyses. Any sequences shorter than the expected products were discarded as non-specific products.

In the 18S ARDRA, a dominant ribotype was observed within all the samples (Figure 14). Representatives of this dominant riboytpe in each sample along with other unique ribotypes were sequenced and used in BLAST analysis. All of the 18S sequences compared best to 18S rRNA genes from various other anemones. There were no 18S rRNA sequences indicating the presence of *Symbiodinium* spp, dinoflagelletes and micro-eukaryotic algae, a surprising result given as these are known symbiotic associations (Hill et al. 2011)

Sequenced clones from the 16S rRNA libraries of *P. flagellifera* and B. *capensis* were similar to various known endosymionts associated with marine invertebrates (Rosa et al. 2000; Ivanova and Mikhailov 2001; Verlag et al. 2003; Schuett et al. 2005; Haine 2008; Lopez et al. 2008; Wang et al. 2008). The clones originating from the *P. varia* 16S rRNA library, however, contained only vector sequence. Results from the NCBI BLASTn analysis of 16S rRNA sequences from the *P. flagellifera* library showed *Colwellia* sp. BSw20195 and

Arcobacter sp. MA5 as the closest known relatives. Similarly the NCBI BLASTn results from 16S rRNA sequences in the *B. capensis* library showed *Brevundimonas* sp. Tibet-IX23, *Shewanella* sp. N6, *Polaribacter* sp. NF3-11, *Colwellia* sp. BSw20188, *Alteromonas sp*. SN2, uncultured *Bacteroidetes* and uncultured *Flavobacteriales* phylotypes as the closest known relatives to the sequenced clones.

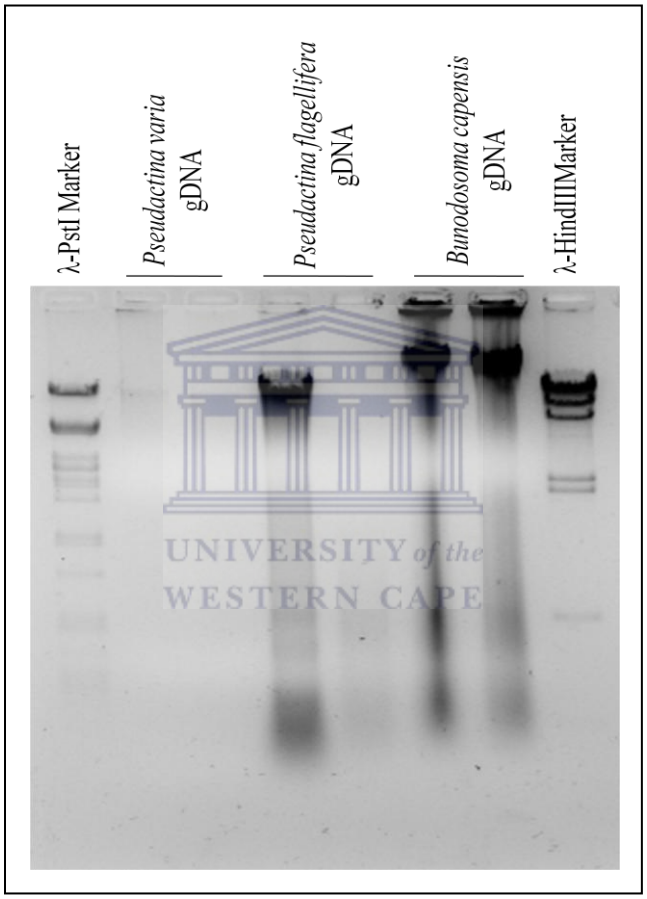


Figure 12. Genomic DNA extracted from sea anemones using a modified Pearson et al. 2002 method. Approximately 1 ng of *P.varia* and 100 ng each of *P. flagellifera* and *B. capensis* were loaded on a 1% agarose gel. For λ-*Pst*I and λ-HindIII marker sizes see Appedix C.

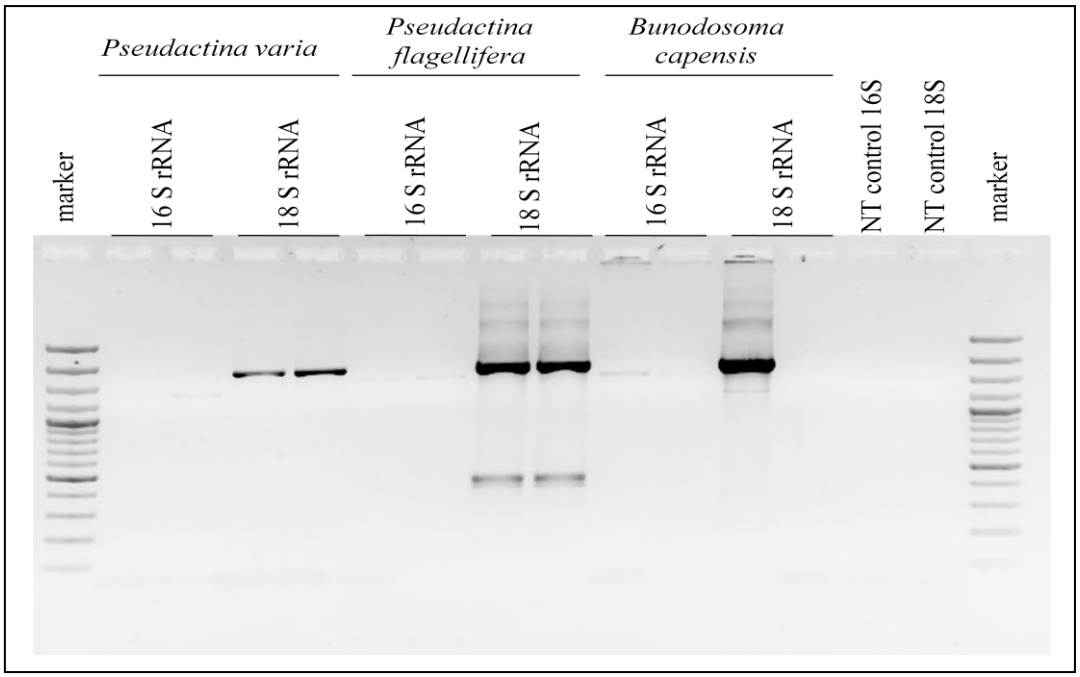
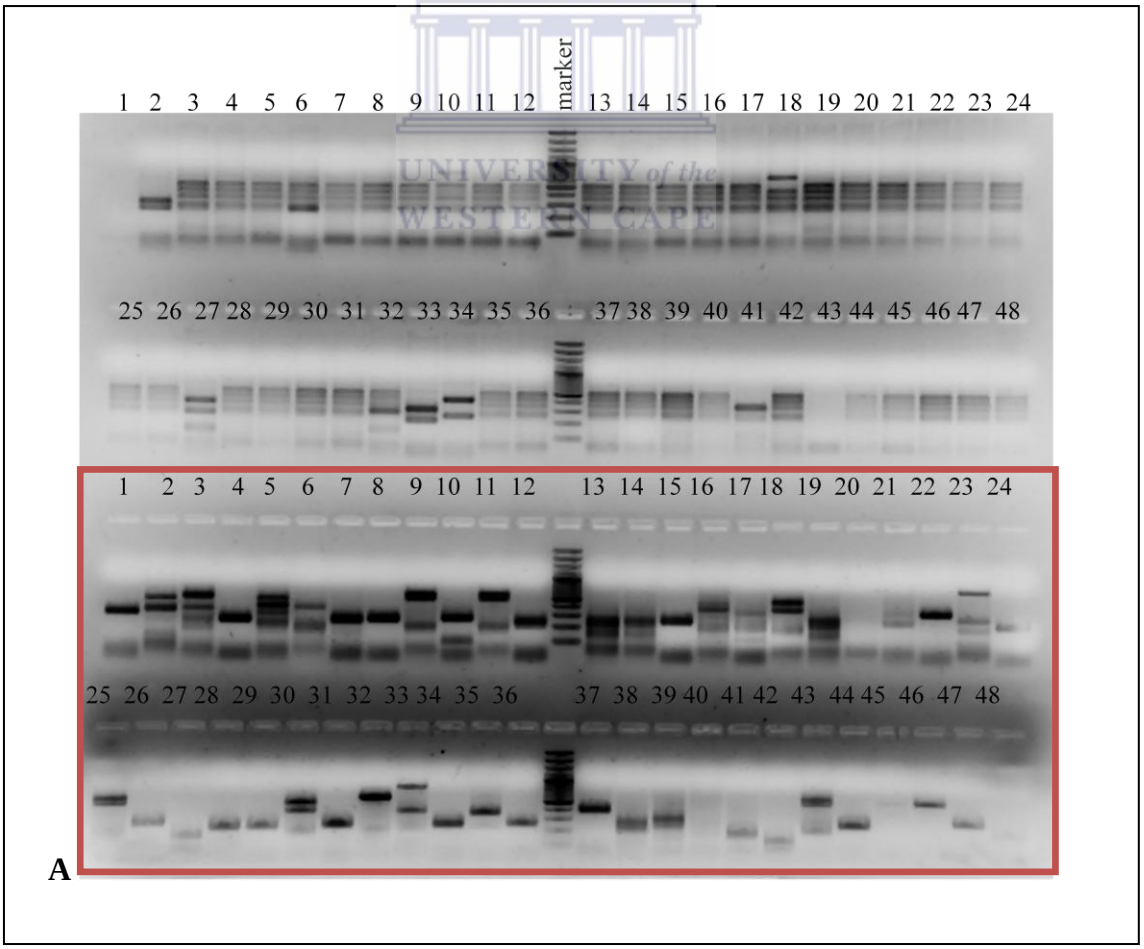


Figure 13. PCR amplicons of 16S and 18S rRNA genes from gDNA extracted from various sea anemone species. The marker is 100 bp plus O'generuler (Appendix C).



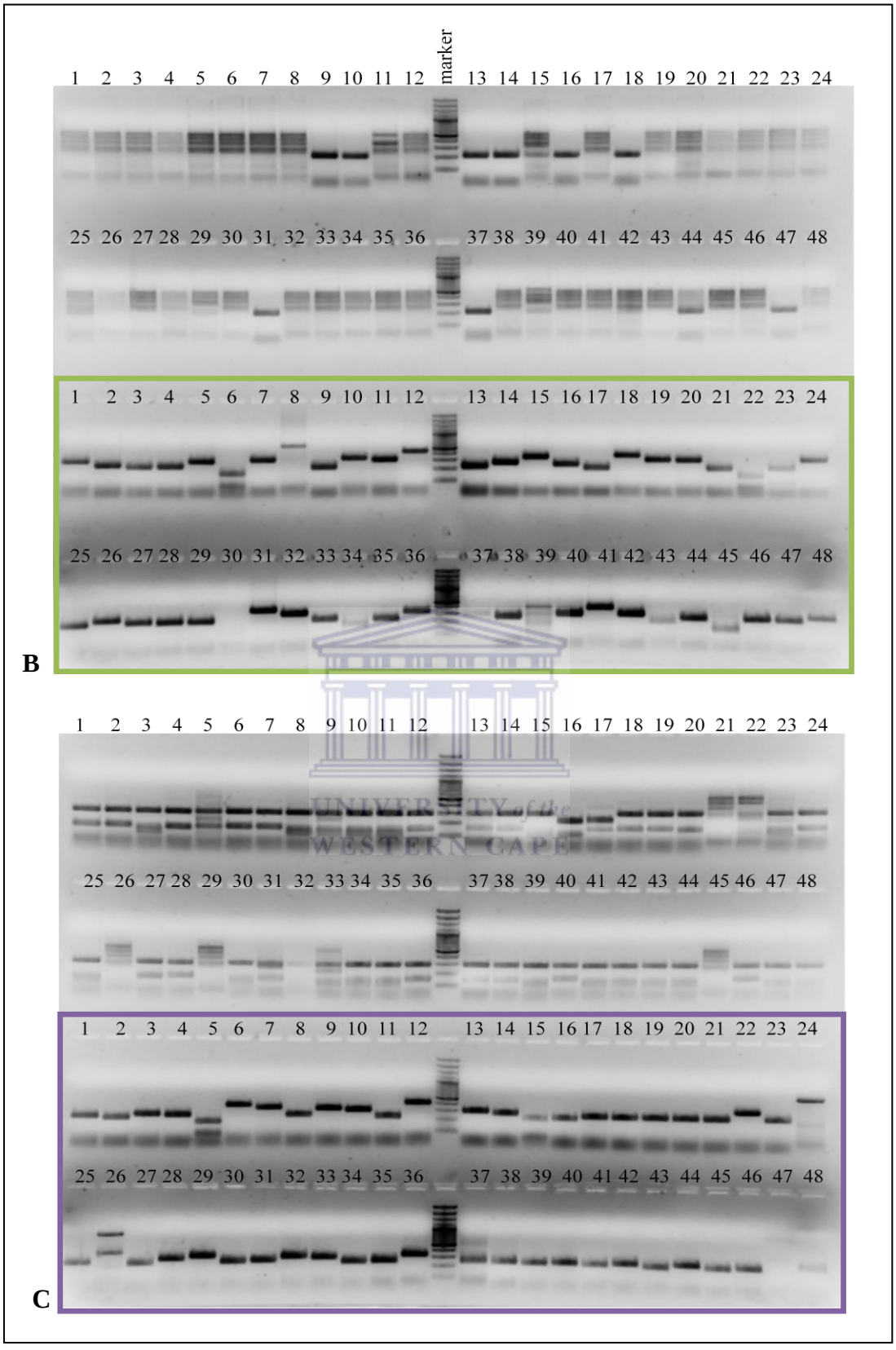


Figure 14. Survey of prokaryotic and eukaryotic diversity associated with sea anemones species. ARDRA patterns from clone libraries of A) *B. capensis* B) *P. varia* and C) *P. flagellifera*. The top 48 lanes (unblocked) are the 18S rRNA clones and the bottom 48 blocked lanes are 16S rRNA clones. The marker in all gels is 100bp plus O'generuler (see Appendix C)

2.2.2. DNA sequence analysis

The phylogenetic trees in Figures 15 and 16 are consistent with trees found in literature (Daly et al. 2003; Daly et al. 2008). The *Order Actinaria* clusters in a monophyletic nature to the *Order Scleractina*. However, within *Actinaria* the families demonstrate examples of para- and polyphyletic disposition. As expected, the 18S rRNA sequences from the sea anemones *P. varia*, *P. flagellifera* and *B. capensis* clustered with other sequences from the *family Actiniidae*. *P. varia* and *P. flagellifera* cluster together as members of the same genus. This is not the case with *B. capensis*. Even though the sequence of *B. grandis* is included in the tree *B. capensis* does not group closest with it, but rather with *Anthopleura krebsi*. However, according to literature, *Havoclava* and *Bundosoma* (Daly et al. 2008) as well as *Anthopleura* and *Bunodosoma* (Heestand, 2009) are known to cluster with each other. Furthermore other examples of species form the same genus not clustering together can be seen with *Antipathes* and *Hormanthia*. Even within the *Family Actiniidae* a similar example can be seen where *Actina midori* does not cluster together with *A.equina* and *A. fragecea* but rather with *Metedwardsia akkeshi*.

Figure 15. Phylogenetic tree constructed using NJ in Mega 4.0 to show the relationships within the *Order Actiniaria*. The orange block highlights the *Family Actiniidae* and is enlarged in Figure 16. The red blocks indicate sea anemone species studied in this thesis.

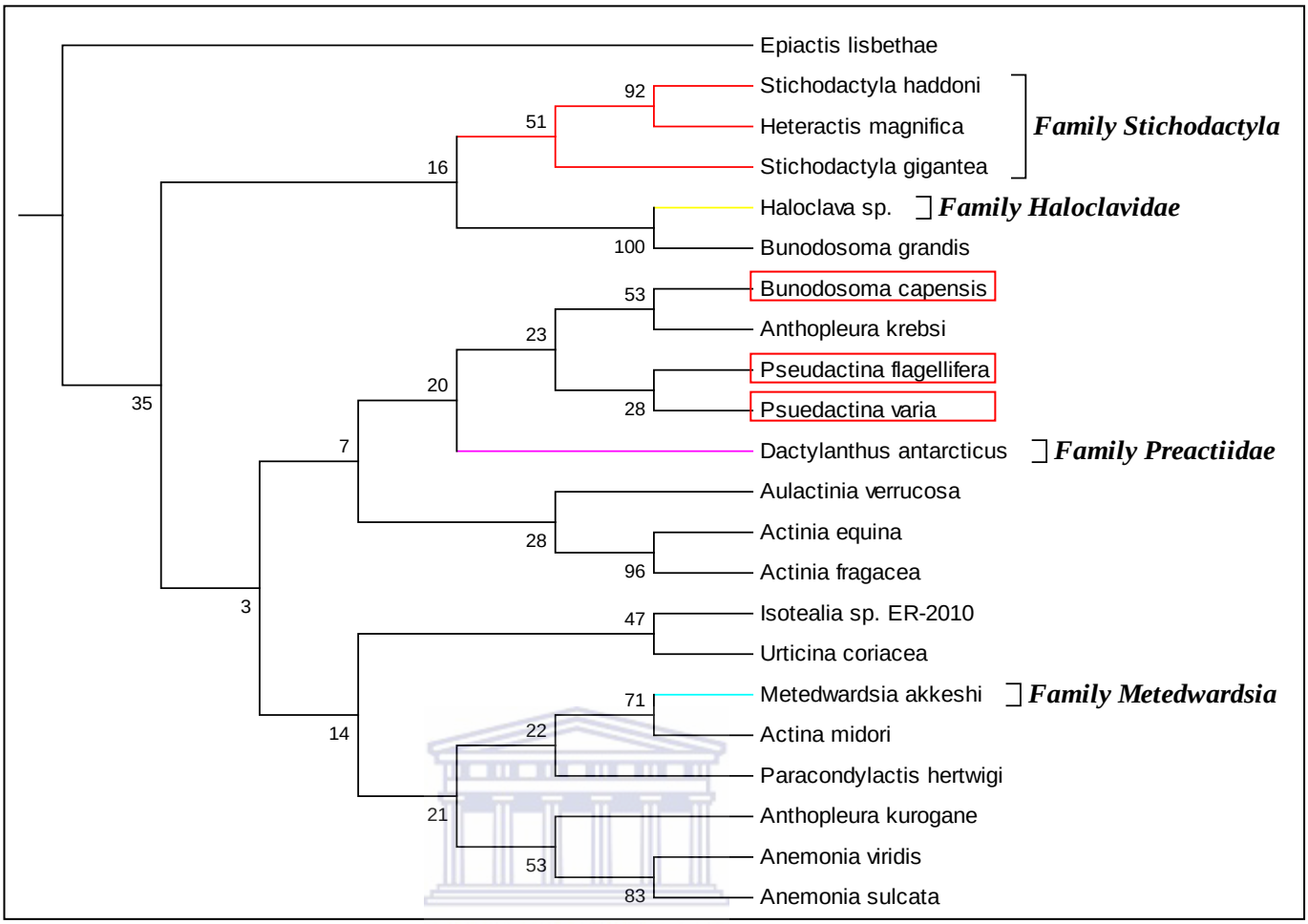


Figure 16. Sub-tree of *Family Actiniidae* (from Figure 15 of *Order Actinaria* tree). The tree shows the polyphyletic nature of the Family because within the *family Actiniidae* four other *Actinaria* families are clustered. Samples from this thesis are blocked in red.

2.3. Discussion

Unique ribotypes from *P. varia*, *P. flagellifera* and *B. capensis* 18S clone libraries were sequenced. These were found to cluster with sequences of the *Family Actiniidae*. The *family Actiniidae* contains 44 genera and upwards of 200 species. There is no distinguishing characteristic that clearly describes this family, as opposed to *Actinernidae* which is distinguished by an unusual arrangement of mesenteries (Daly et al. 2007). It is the general consensus that *Actiniidae* is probably not monophyletic as phylogenetic studies using DNA sequences that include representatives from *Actiniidae* and those from other families in

Endomyaria (super family) have failed to recover its members as sister taxa (Daly et al. 2007).

The use of rRNA and other mitochondrial genes as phylogenetic markers has been reliable as data generated display similar trends using various genes (Daly et al. 2003; Daly et al. 2008). However, when trying to relate phylogenetic studies to an evolutionary time line for various characters e.g. development of basilar muscles, a trait which is assumed to be lost and recovered numerous times in *Actinaria*, the use of rRNA and mitochondrial genes does not seem logical (Won et al. 2001). This is a possible reason why the data presented in this and other studies is contradictory. Furthermore, the morphological characters used in distinguishing various groups generate lineages that are misleading (Heestand, 2009). Concerns about heteroplasy make phylogenetic analysis more challenging as generally characters used in phylogenetic analysis are derived from a single common ancestor.

The sequenced 16S rRNA showed a range of organisms which are thought to have endosymbiotic relations with marine invertebrates. A study by Rhower et al. (2002) on the bacterial diversity in corals showed that marine invertebrates provide a structural matrix that host diverse and ecologically structured prokaryotic communities. Additionally, many of the bacterial associates are previously and yet uncultured species. Further analysis in the study revealed that different species (regardless of geographical location and developmental stage) host distinctive communities illustrating that the association is non-random. Lastly, the study reported that bacterial communities were spatially arranged in an organ specific manner within the corals (Rohwer et al. 2002). This shows the importance of the site where the tissue is taken from and was confirmed in another study by Zongjin et al. 2010 which showed that different bacterial communities dominate different sections of the sea anemone (Zongjun et al. 2010).

When investigating the characteristics of the 16S clones' closest related organisms the most notable is that almost all display prominent molecular adaptations for psychrotolerance (Murray et al. 2007). This was not unexpected as the samples originate from the southern Atlantic oceans. The symbiotic associations observed in this study are, like in most marine environments, dominated by *Proteobacteria* (gamma, alpha and epsilon) as well as *Bacteroidetes* and play various roles in the sea anemone ranging from protection to nutrition. The bacteria are either transferred within the planula stage of reproduction with the sea anemone or can be acquired by ingestion (Werren and Neill 1879).

The psychrotolerant microbes such as *Cowellia* and *Shwanella* are known to produce and accumulate compounds such as PUFA and poly-hydroxyalkanoate compounds, which can serve as both carbon and energy stores for starvation periods (Russell and Nichols 1999; Murray et al. 2007). *Shewanella* has an additional suggested role of metal detoxification (Wang et al. 2008). In addition, some microbes such as *Alteromonas* play a role in host defences by producing various anti-microbials (Ivanova and Mikhailov 2001) or toxins (particularly) in the tentacles of the anemone (Schuett and Doepke 2010).

3. Chapter Three - cDNA technology and eukaryotic gene discovery

Recovery of bioactive compounds such as proteins (including enzymes which serve as biocatalysts or hormones for their medicinal uses) or polysaccharides (for drug discovery) can be achieved by various methods. Organic extractions are common and well established in the chemical industry (Gold and Bethell 1977) but gene discovery, a fairly new technique in comparison, offers the alternative of retrieving the genetic component the results in the specific activity (Gu and Wang 2011).

Unlike for prokaryotes a genomic approach for gene discovery in eukaryotes is not often successful. This is because eukaryotic genes often contain large intronic sequences; they lack conserved motifs in promoter sequences (Bailly et al. 2007) and therefore cannot be expressed by bacterial hosts. Furthermore, due to the large size and variation of eukaryotic genomes (Wood et al. 2002; Aury et al. 2006) it is unrealistic that a substantial fraction of gene content would be expressed and recovered in various bacterial or many eukaryotic heterologous hosts (Tringe et al. 2005; Delong et al. 2006). To circumvent these problems mature mRNA molecules (containing no introns) can be converted to double stranded complementary DNA (cDNA) using reverse transcription. The processed cDNA can be cloned, expressed and screened for the desired gene or function by sequence homology or detection of resulting bioactivity in prokaryotic hosts (Grant et al. 2006).

The general protocol for construction of cDNA libraries is described in Figure 17. The most critical step when isolating intact RNA is to rapidly inactivate the RNases present in the cell (Chomczynski and Sacchi 1987). This is achieved by the addition of guanidium thioisocyanate. The process of generating cDNA is performed by a multidomain enzyme, reverse transcriptase, which has both RNA and DNA dependant polymerase activity and RNaseH specific activity (Das and Georgiadis 2004). Reverse transcriptases are found in all retroviruses and are used to convert the single stranded RNA genome into DNA that can be

intergrated into the host's DNA (Georgiadis et al. 1995). They have been adapted for biotechnological applications in cDNA synthesis wherein mRNA template is converted to single stranded DNA that can be amplified and cloned as double stranded cDNA (Izuno et al. 2010).

The culture independent approach where cDNA of a sample is directly cloned into appropriate vectors for expression studies presents a workable alternative for novel eukaryotic gene recovery (Grant et al. 2006). This has been made possible due to advances in RNA stabilisation and isolation; mRNA enrichment and cDNA technology enabling the isolation of full-length cDNA for library construction (Grant et al. 2006). Today commercial enzymes such as MonsterScript (EPICENTRE) and other polymerases are able to reverse transcribe up to 15 Kb.

Screening cDNA libraries is dominated by two strategies, namely PCR-based amplification and probe hybridization. PCR-based approaches rely on prior sequence knowledge and primer design for successful screening (Munroe et al. 1995). Advances in PCR screening have been possible through new primer design strategies, pooling schemes, and improved thermal cyclers. The improvements to thermal cyclers include decreased sample volumes, increased sample numbers per run, and decreased required cycling time (Alfandari and Darribere 1994; Campbell and Choy 2002).

Techniques such as ELISA, western blots and other immunoscreening assays make use of probe hybridization to screen for structurally intact protein targets. Hybridization, however, is not restricted to protein or metabolite detection and can also be used in sequence-based screens such as in southern blots (Onishi et al. 1995; Campbell and Choy 2002). A major bottle neck in screening active extracts has been recovery of sufficient proteins or active compound for detection. To simplify hybridization-based approaches, which are labour-intensive and time consuming, high-throughput expression screening and purification of

proteins in *E. coli* (Vincentelli et al. 2011) has been applied. Furthermore, automated robot technology such as robotic plaque/colony pickers and gridders reduce the labour involved in handling and arraying cDNA libraries. The application of high-throughput microtiter plate techniques and MALDI–TOF-MS stream-line the analysis of gene products and detection of new biological activity of yet uncharacterized proteins (Kornbluth et al. 1988; Tanaka et al. 1999; Bussow et al. 2000).

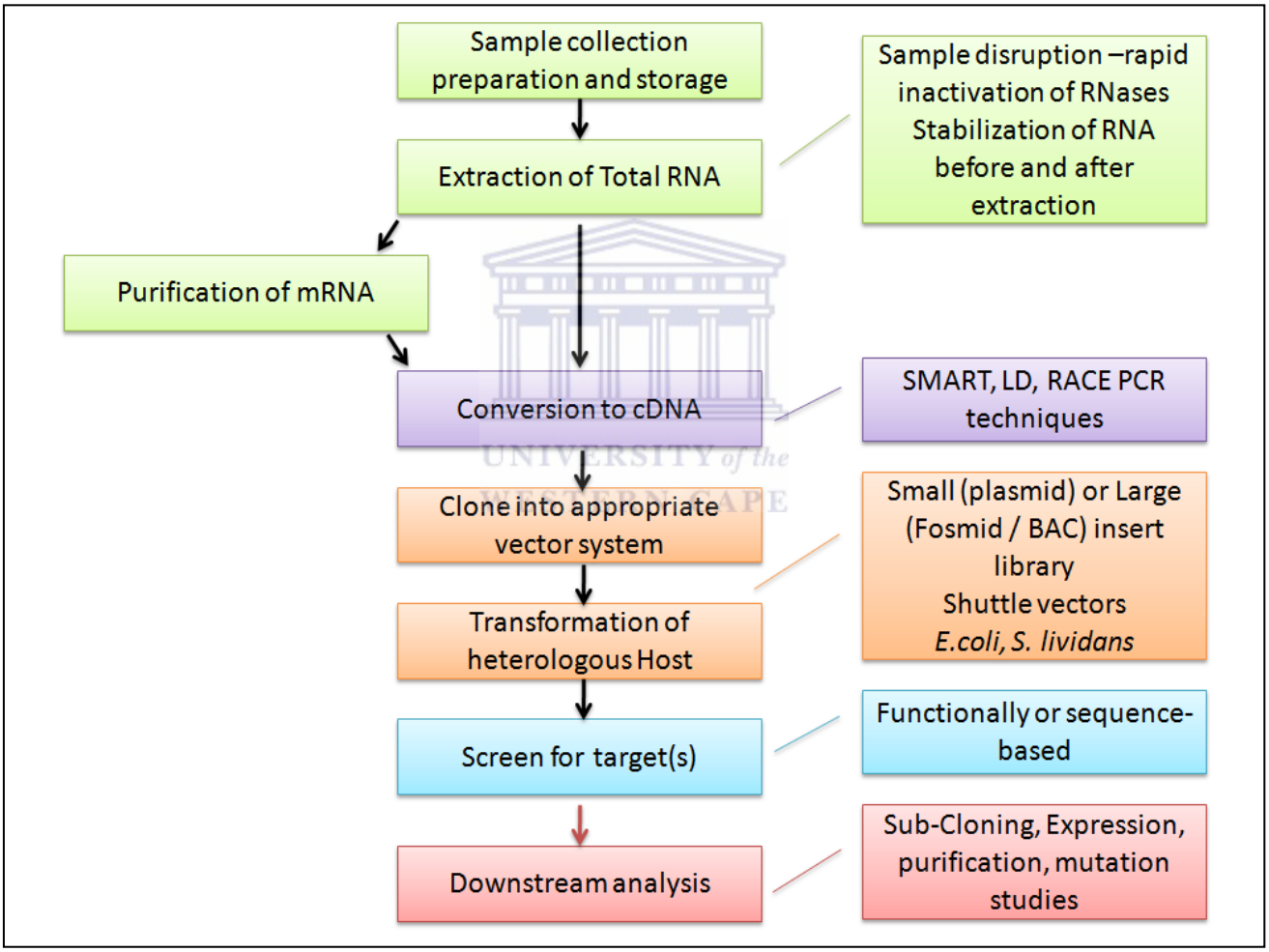


Figure 17. The major steps required to construct a cDNA library.

Vectors used to construct libraries that would be screened based on detection of activity have promoters on the plasmid that allow the recombinant DNA target to be transcribed and expressed from the host's machinery. The vectors used in constructing libraries for sequence

based screening do not have promoter sequences and only allow for the recombinant DNA target to be amplified as the vector is replicated (Garrett and Grisham 2005).

Table 3 Vectors used to construct cDNA libraries for various screening approaches

Vector	Screening approach	Vector source	cDNA synthesis
pSMART2IF	Sequence based	Clontech	SMARTer
pLINKet	Activity detection	Modified pET21a	SMARTer
pEGFP-N3	Activity detection	Clontech	Modified SMARTer

In this study functional screening and sequence base approaches were designed. This chapter describes the construction of 3 cDNA libraries for each of the sea anemone samples and the modifications to the techniques.

3.1. Materials and Methods

3.1.1. RNA extraction

Total RNA was extracted using a single step acid phenol-chloroform guanidium-thiocyanate protocol (Chomanzyski and Sacchi 1987). All tips, tubes, glassware, solutions and equipment used were RNase free. Tips and solutions were double autoclaved and most solutions were treated with 0.1% DEPC. Glassware was washed with 0.5M sodium hydroxide and rinsed using double autoclaved DEPC treated water. All surfaces were wiped down using 10 % bleach and an RNase removal solution (3 % hydrogen peroxide, 0.1M Sodium hydroxide and 0.1 % SDS). Tissue samples stored in RNA*later* (Sigma) were washed three times in 1X PBS pH 7.2 to remove any inhibitory chemicals.

Tissue (500mg) was ground to a fine powder in liquid nitrogen with a pre-chilled mortar and pestle. The powdered tissue was homogenized in 5 mL Solution D (see Appendix A) after which the following were sequentially added on ice: 1/10 volume 2 M sodium acetate (pH 4.0), 5 mL water-equilibrated phenol (~pH 4.3) and 1 mL chloroform: isoamyl Alcohol (24:1). The tubes were inverted to mix after each addition. The lysate-mixture was vortexed for 10 seconds and placed on ice for 15 minutes. The organic and aqueous phases were separated by centrifugation at 10 000 *g* and 4°C for 20 minutes. The aqueous phase (i.e. top layer) was transferred to a fresh tube without any of the interphase to which 1 mL chloroform: isoamyl alcohol (24:1) was added. After thorough mixing the sample was placed in ice for 15 minutes. The phases were separated as before and again the aqueous phase was transferred to a new sterile tube. One volume of ice-cold isopropanol was added and the sample was inverted to mix. The samples were placed at -20^{o}C for approximately two hours to precipitate RNA. The RNA was collected as a pellet by centrifugation at 10 000 *g* and 4^{o}C for 20 minutes. The precipitated RNA pellet was re-dissolved in 1.5 mL Solution D and precipitated with 1 volume of ice-cold isopropanol at -20^{o}C for 1 hour. RNA was collected by centrifugation at 10 000 g and 4^{o}C for 20 minutes. The pellet was washed with 70% ethanol and collected as before, air-dried for 10 minutes and re-suspended in 100 µL of RNase-free mQ water. Aliquoted RNA was stored at -20°C until needed.

3.1.1.1. MOPS/Formaldehyde gel electrophoresis

A 1% MOPS formaldehyde gel was prepared by adding 1% [w/v] agarose to RNase free mQ water and boiled. After cooling 10 X MOPS running buffer, 37% formaldehyde and 10 µg/µl ethidium bromide were added to final concentrations of 1 X, 5% [v/v] and 1 µg/mL, respectively. The gel was cast, set and electrophoresed in 1 X MOPS running buffer at 50 V/cm for 3 hrs.

3.1.2. cDNA library construction in pSMART2IF

3.1.2.1. SMARTer™ cDNA synthesis

All cDNA libraries to be screened by sequence homology were constructed using the In-Fusion™ SMARTer™ cDNA library construction kit (Clontech, USA) according to manufacturer's instructions with minor modifications.

For first-strand cDNA synthesis 1 µg of total RNA was primed using the 3'CDS primer (supplied by manufacturer) at 72°C for 2 minutes followed by 42°C for 3 minutes. Immediately after priming a master mix containing first strand reaction buffer, 2.5 mM DTT, 1 mM dNTP mix, 1.2 µM SMARTer II A oligonucleotide, 10 units RNase inhibitor and 100 units SMARTscribe™ reverse transcriptase was added and the samples were incubated at 42°C for 90 minutes to complete first strand extension. Unused first strand cDNA was stored at -80°C.

Second-strand synthesis was carried out by long distance PCR. Duplicate reactions were set up each containing: 2 µL of first strand cDNA, 1 X HF Phusion buffer (Finnzymes, OY, Finland), 200 µM dNTP (Fermentas), 240 µM 5' PCR primer II A (Clontech) and 4 units Phusion polymerase (Finnzymes, OY, Finland) in a final 100 µL reaction volume. The cycling parameters for synthesis were as follows: Initial denaturation at 95°C for 1 minute, denaturation at 95°C for 15 seconds, annealing at 65°C for 30 seconds, extension at 68°C for 6 min. The number of cycles was determined experimentally by monitoring formation of product of the appropriate size by gel electrophoresis over an increasing number of cycles.

3.1.2.2. Size fractionation and purification

Double stranded cDNA was electrophoresised on a 1% agarose gel. Fragments between 1 and 3-Kb were excised and purified from the gel using the Nucleospin Extract II kit (Machery-Nagel) according to the manufacturer's instructions.

3.1.2.3. In-Fusion™ Cloning and cDNA library recovery

Cloning reactions for each sample were set up according to manufacturer's instructions with vector to cDNA ratios of 1:1.5, 1:2 and 1:2.5. The entire ligation reaction was drop dialysed on a 0.025 µm VSWP membrane (Millipore), diluted to a final volume of 12 µL and transferred for storage until needed. The dialysed recombinant vector was transformed into *E. coli* GeneHogs as in section 2.1.2.2.

The entire transformation mix containing the cDNA from *P. varia* was spread onto LB agar plates supplemented with 100 µg/µL ampicillin, 80 µg/µL X-gal and 1 mM IPTG and incubated overnight. Plasmid DNA containing the cloned cDNA was isolated from the transformed *E. coli* using the Qiagen midi-prep kit after the cells were scraped off the LB agar plates.

For the *P. flagellifera* and *B. capensis* libraries 100 µL of each was spread onto LB agar supplemented with 100 µg/µL ampicillin, 80 µg/µL X-gal and 1 mM IPTG for library verification. To reduce library amplification the remainder of the transformation mixture was added to 15 mL pre-warmed (37°C) SOC containing 150 µg/mL ampicillin and incubated at 37°C for an additional hour. The cDNA plasmid library was isolated using the Qiagen Mini-prep kit according to manufactures instructions.

3.1.2.4. PCR-based screening

Degenerate primers AfpF2, AfpF3, AfpF6, AfpR, dCPfwd and dCPrv were used to screen the cDNA libraries for anthozoan fluorescent and chromoproteins (AFP and CPs) (Table 5 in Appendix B). A touch-down PCR protocol such as outlined by Korbie and Mattick (2008) was applied when screening the cDNA libraries. The annealing temperatures for the various degenerate primers were optimised across a gradient ranging from 52 to 72°C. Optimal template and primer concentrations were determined using a positive control (dApc) and reactions were set up according to standard PCR (see 2.1.3.3).

3.1.3. cDNA library construction in pLINKet

3.1.3.1. pLINKet vector synthesis

A linearized In-Fusion cloning vector for construction of cDNA libraries, which could be screened based on function, was prepared by adding two 15 bp linker molecules (TTGATACCACTGCTT) to pET21a during an inverse PCR as illustrated in Figure 35. For a single reaction a standard PCR (see 2.1.3.3) was set up using Phusion polymerase in the HF buffer system. Template, primer concentrations and annealing temperature were optimized. Parental pET21a background was removed after PCR synthesis by treatment with *Dpn*I. The pLINKet vector was checked on an agarose gel (see 2.1.3.2) and a single band was gel excised and purified using the Nucleospin Extract II kit according to manufacturer's instructions.

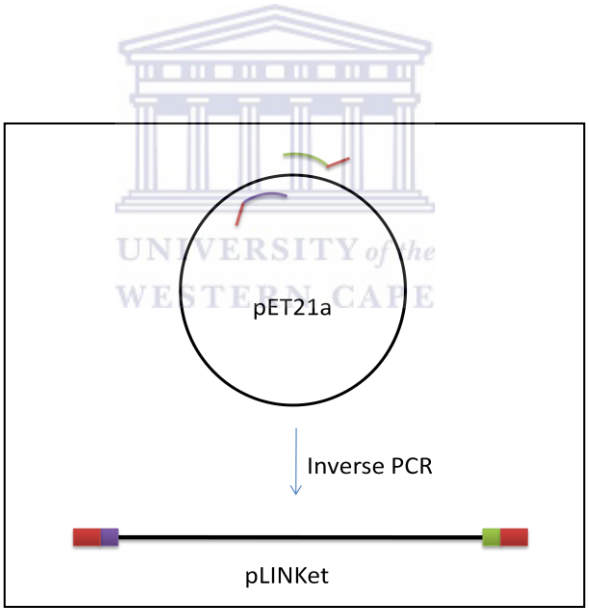


Figure 18. Synthesis of pLINKet. Primers including 15 bp linker sequences are used in inverse PCR with pET21a as template to generate a linear vector pLINKet.

3.1.3.2. cDNA synthesis

SMARTer cDNA synthesis as in 3.1.2.1 was performed using total RNA extracted in 3.1.1.

3.1.3.3. Size fractionation and In-Fusion cloning

cDNA was size fractionated by gel excision and purified as in 3.1.2.2. In-Fusion cloning was carried out as per SMARTer™ In-Fusion™ cDNA library construction protocol. Reaction buffer, cDNA and pLINKet vector in a 2:1 ratio were setup in parallel with various control reactions: (i) a positive control (2kb *lacZ* fragment) supplied with the In-Fusion cloning kit was used to verify that the pLINKet vector could clone insert, (ii) no insert and (iii) no enzyme reactions. The transformants were plated onto LB supplemented with IPTG, X-gal and ampicillin (as in Appendix A).

3.1.3.4. Vector and library verification

Prior to cloning, the pLINKet vector was verified for functionality (replication and antibiotic resistance). The vector was restricted using *Xba*I and *Hin*dIII, blunted using T4 DNA polymerase, re-circularized using T4 DNA ligase and transformed into *E. coli* GeneHogs. The transformation mix was spread onto LB agar supplemented with IPTG and ampicillin (as in Appendix A). Finally the vector ends were sequenced using the T7 promoter and terminator primers.

3.1.4. cDNA library construction in pEGFP-N3

3.1.4.1. Modified cDNA synthesis

An alternative approach to constructing a functional library was to modify the SMARTer™ cDNA synthesis reaction. This was achieved by designing primers FUNF and FUNR (see Table 5 in Addendum B) that contained *Sal*I and *Not*I restriction sites for directional cloning. The forward primer also contained an AGGAGG consensus ribosomal binding sequence.

An outline of the modified cDNA synthesis protocol is illustrated in Figure 19B together with the original SMARTer™ cDNA synthesis in Figure 19A. Briefly, all remaining first strand synthesis reactions were subjected to 3 rounds of second strand conversion as previously described in section 3.1.2.1. Excess 5'PCR primer (sequence not provided by Clonetech) and

dNTPs that could interfere with subsequent rounds of double stranded cDNA synthesis were removed by ExoSAP (GE Healthcare, USA) treatment according to manufacturer's instruction. Modified primers (100 µM) were added and cycled as in 3.1.2.1 for 21 cycles.

3.1.4.2. Cloning cDNA library into pEGFP-N3

Double stranded cDNA was digested with *Sal*I and *Not*I and prepared as in 3.1.2.2. T4 DNA ligase (Fermentas) was used to clone sticky end cDNA fragments into *Sal*I and *Not*I linearized pEFGP-N3 vector in a 2:1 insert: vector ratio. The libraries were transformed into *E. coli* BL21 DE3 pLysS and spread onto LB agar supplemented with kanamycin (as in Appdendix A).

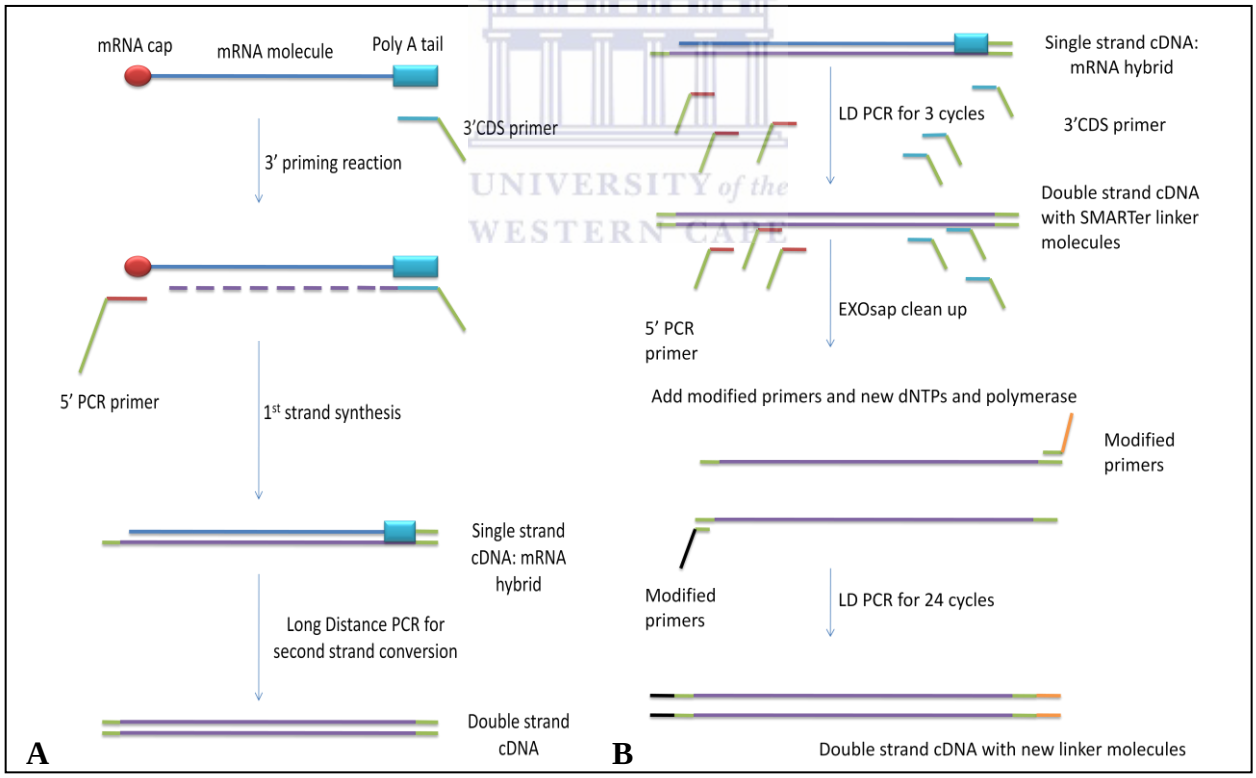


Figure 19. SMARTer cDNA synthesis. A) Illustration of the addition of linker molecules in the first strand synthesis. B) the modifications to the SMARTer cDNA synthesis

3.2. Results of cDNA library construction

3.2.1. Total RNA extraction

The integrity of the total RNA extracted from *P. varia*, *P. flagellifera* and *B. capensis* was checked on a 1% MOPS formaldehyde gel (Figure 18). The approximate 2:1 ratio of the 28S to 18S rRNA transcripts indicate that intact high quality RNA was extracted. To determine if gDNA was co-extracted with the RNA a PCR reaction was set up using eukA and eukB primers. After 35 cycles no amplification bands were present indicating the absence of gDNA (Figure 43 in Addendum D).

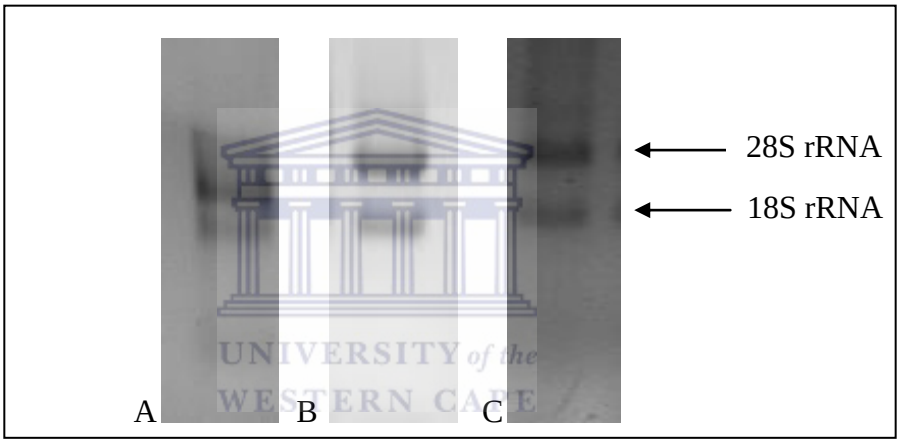


Figure 20. Total RNA extractions from sea anemones A) *Pseudactina varia* B) *Pseudactina flagellifera* and C) *Bunodosoma capensis*. The 18S and 28S rRNA transcripts are labelled accordingly.

3.2.2. cDNA synthesis

Approximately 1 µg of total RNA was used as template in the construction of each of the 3 cDNA libraries using the In-Fusion SMARTer cDNA synthesis library construction kit. Smaller DNA fragments are favourably biased in a PCR reaction. Thus the optimal number of cycles that would limit the bias effect as well as recover rare transcripts during the second strand synthesis was determined experimentally for each sample. The PCR parameters were set so that each sample was subjected to at least 15 cycles of amplification after which a small amount (30 µL) was removed and subjected to further cycling. Aliquots (5 µL) were

taken at 18, 21, 24 and 27 cycles. The cDNA products in the aliquots were visualized on an agarose gel and it was observed that 24 cycles produced the highest yield of cDNA in the correct size range (from 0.6 to 3kb) (Figure 21).

3.2.3. Insert size & diversity and library size

Colonies obtained from plating 100 µL of the total transformation mix were examined and >95% contained insert. Random clones obtained from the library transformations were selected and the plasmid DNA extracted. The size of cDNA inserts was determined by restriction endonuclease digestion using *Xba*I and *Eco*RI (Figures 22-24). Insert sizes ranged from 0.4 kb to 2.5 kb with an average insert size of 1.1 kb for *P. varia*, 0.9 kb for *P. flagellifera* and 0.8 kb for *B. capensis*. Additionally the forward and reverse screening primers (In-Fusion SMARTer cDNA synthesis library construction kit, Clonetch) were used to amplify cloned cDNA from each of the libraries in order to determine the size of inserts, which was found to vary between 0.5 kb to 2.8 kb (Figure 25).

Based on average insert size and the estimated number of clones in each library (~15 000 in *P. varia* and ~22 000 in *P. flagellifera*) it was found that both the amplified *P. varia* and non-amplified *P. flagellifera* libraries each covered approximately 15 Mbp of cDNA. Using the size of the nearest relative genome, namely *Nematostella vectensis* (the starlet anemone) which is 450 Mbp, the calculated genome coverage of the cDNA libraries was only 3%. However, based on the recent genome draft and transcriptomic data for *N. vectensis*, the starlet anemone was said to contain between 18 000 and 22 000 genes with an average insert size of 0.9 kb. Therefore, the calculated percentage of expressed genetic material captured in each cDNA library is between 76-93%.

To establish that the cDNA cloned was from the sea anemone and not from contamination, randomly selected clones were sequenced (Macrogen) using M13 forward and reverse primers. Table 4 contains the tBLASTx (http://www.ncbi. nlm nih.gov/BLAST/) results of

the sequence data from the 19 respective clones. The sequences returned showed a bias toward transcriptional regulators (CnidEFs), transcripts of small cysteine rich proteins (SCRiPs) unique to Cnidarians and surprisingly two 16S rRNA transcripts of uncultured bacteria.

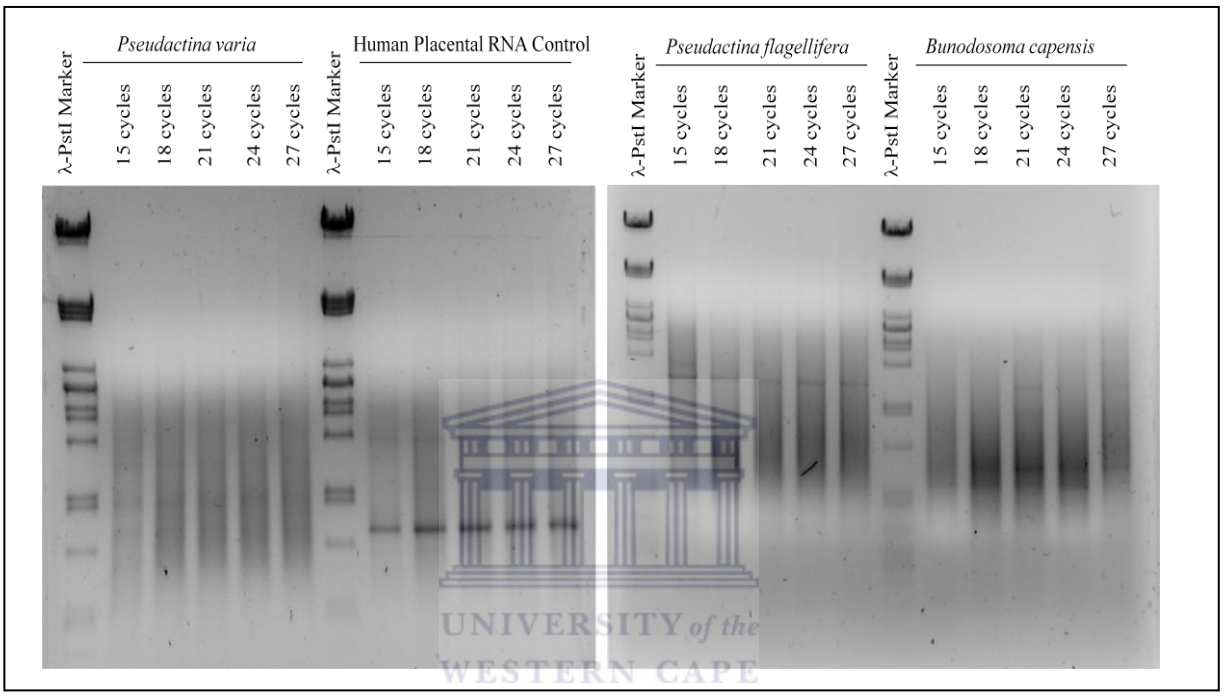


Figure 21. Experimental determination of cycling for double stranded cDNA. The control provided with the kit is human placental RNA which has a prominent transcript of ~900 bp. For λ-*Pst*I marker sizes see Appendix C.

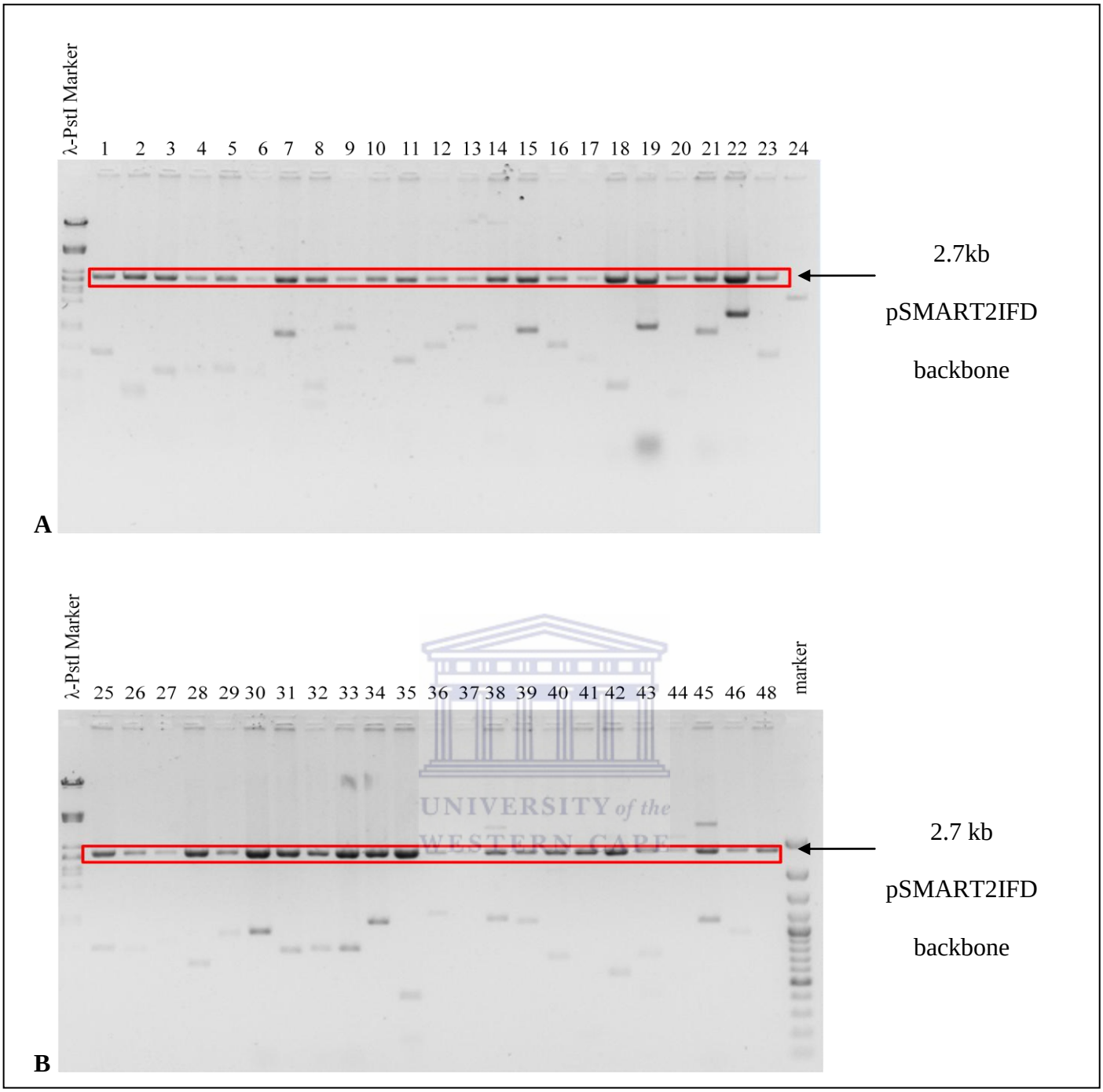


Figure 22. Double digests performed with *Xba*I and *Eco*RI on random colony plasmid preps to determine insert size for the library of *P. varia*. The pSMART2IF vector backbone is clearly marked and consistent in all digests. Clones contained inserts which ranged from 0.4 kb to 1.2 kb. The markers are 100 bp plus O'generuler and λ-*Pst*I for band sizes see Appendix C.

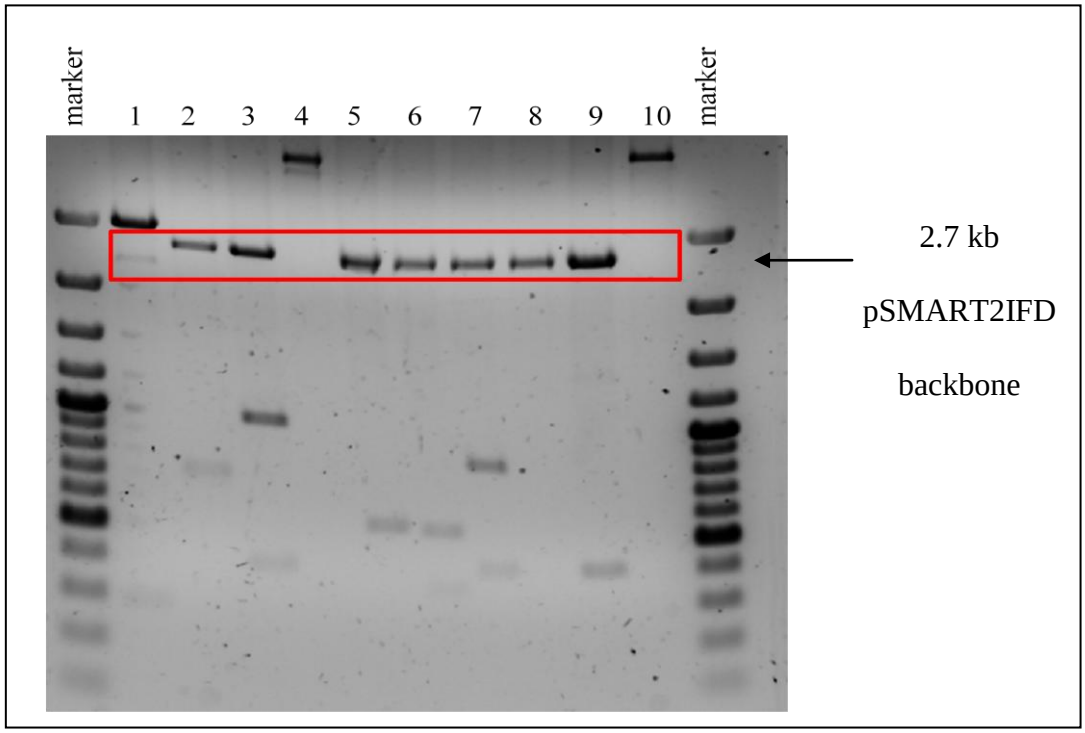


Figure 23. Double digests performed with *Xba*I and *Eco*RI on random colony plasmid preps to determine insert size for the library *P. flagellifera*. The pSMART2IF vector backbone is clearly marked. Colonies contained insert which ranged from 0.4 kb to 2.5 kb. Marker is 100 bp plus O'generuler (see Appendix C).

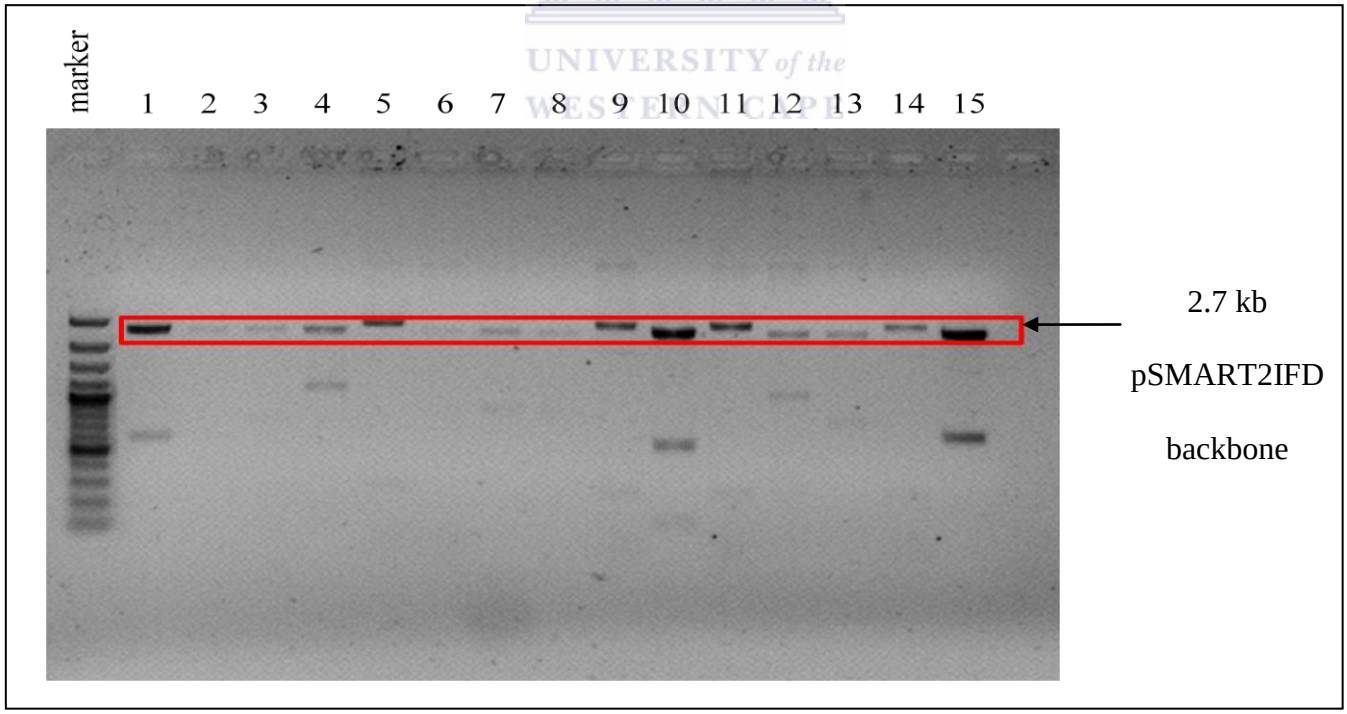


Figure 24. Double digests performed with *Xba*I and *Eco*RI on random colony plasmid preps to determine insert size for the *B. capensis* library. The pSMART2IF vector backbone is clearly marked. Colonies contained insert which ranged from 0.3 kb to 1.2 kb. Marker is 100 bp plus O'generuler (see Appendix C).

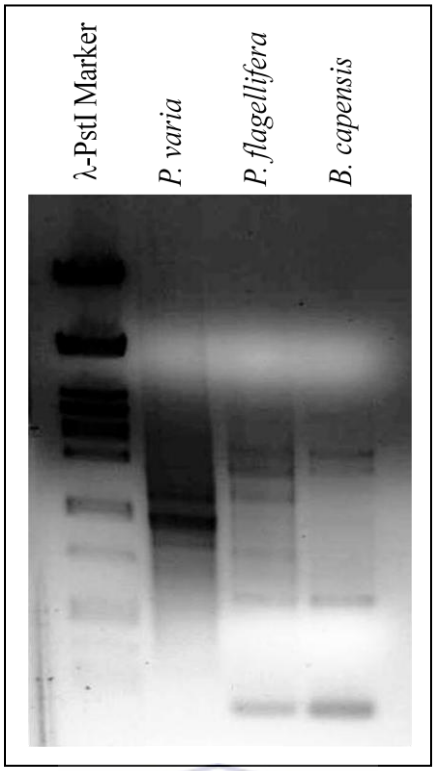


Figure 25. Diversity of insert size when 1 µL cDNA is used in a PCR using the forward and reverse screening primers from the SMARTer cDNA kit. (For λ-*Pst*I molecular band sizes see Appendix C)

Table 4 BLAST analysis of sequenced inserts from random pSMART2IF library clones

Closest tBLASTx results	**E-value**
Pseudactina varia	
Uncultured prokaryote clone deadsea_1992_DSAA-DSG2_3G03Rev genomic sequence	0.046
Mycobacterium marinum M, complete genome	1.0
Leishmania braziliensis complete genome, chromosome 18	2.6
Uncultured actinobacterium clone HSB-E05 16S ribosomal RNA gene, partial sequence	2e-12
Uncultured bacterium clone GL81-Act13 16S ribosomal RNA gene, partial sequence	4e-28
Nematostella vectensis predicted protein mRNA, complete cds	1e-95
Pseudactina flagellifera	
Nematostella vectensis predicted protein partial mRNA	1e-32

Aiptasia pulchella clone 2C7 mRNA sequence	2e-18
Aiptasia pallida CnidEF mRNA, partial cds	9e-14
Anthopleura elegantissima CnidEF mRNA, complete cds	4e-12
Gallus gallus hypothetical protein LOC772907 (LOC776475), partial mRNA	3e-09
Nematostella vectensis predicted protein mRNA, complete cds (BLASTx)	7e-14
Nematostella vectensis predicted protein partial mRNA	1e-19
Phyllostachys edulis cDNA clone: full insert sequence	0.18
Bunodosoma capensis	
Nematostella vectensis predicted protein partial mRNA	0.003
Symbiodinium sp. clade C3 from Acropora aspera CCAAT enhancer-binding protein-like mRNA sequence	3e-18
Nematostella vectensis predicted protein partial mRNA	2e-23
Acropora millepora small cysteine-rich protein 2 (SCRiP2) mRNA, complete cds	2e-04
Nematostella vectensis predicted protein partial mRNA	2e-66

3.3. PCR Screening of cDNA libraries and genomic DNA

3.3.1. Degenerate primers for screening for anthozoan fluorescent and chromoproteins

To screen the cDNA libraries based on sequence homology, degenerate primers were designed specific for AFPs and ACPs. More than 80 sequences representative of all fluorescent protein colours as well as chromoproteins from sea anemones and corals were collected from NCBI genebank and aligned using MAFFT's online (http://mafft.cbrc.jp/alignment/) G-INS-i algorithm and the BLOSUM80 scoring matrix. Two areas of conservation were selected as indicated in Figure 26 for the design of degenerate primers (Table 5 in Appendix B) which could potentially serve as "universal" screening primers. Furthermore, these regions are not conserved in fluorescent and chromoproteins from jellyfish and non-anthozoan organisms (Figure 27). Therefore, they would target a variety of AFPs and ACPs from the *Order Anthozoan*. Despite the high degree of

conservation at the protein level an alignment of the nucleotide sequences of the conserved regions identified (Figure 28) revealed that the sequences were more degenerate than initially thought. The degeneracy of the screening primers designed ranged from 48 to 256.

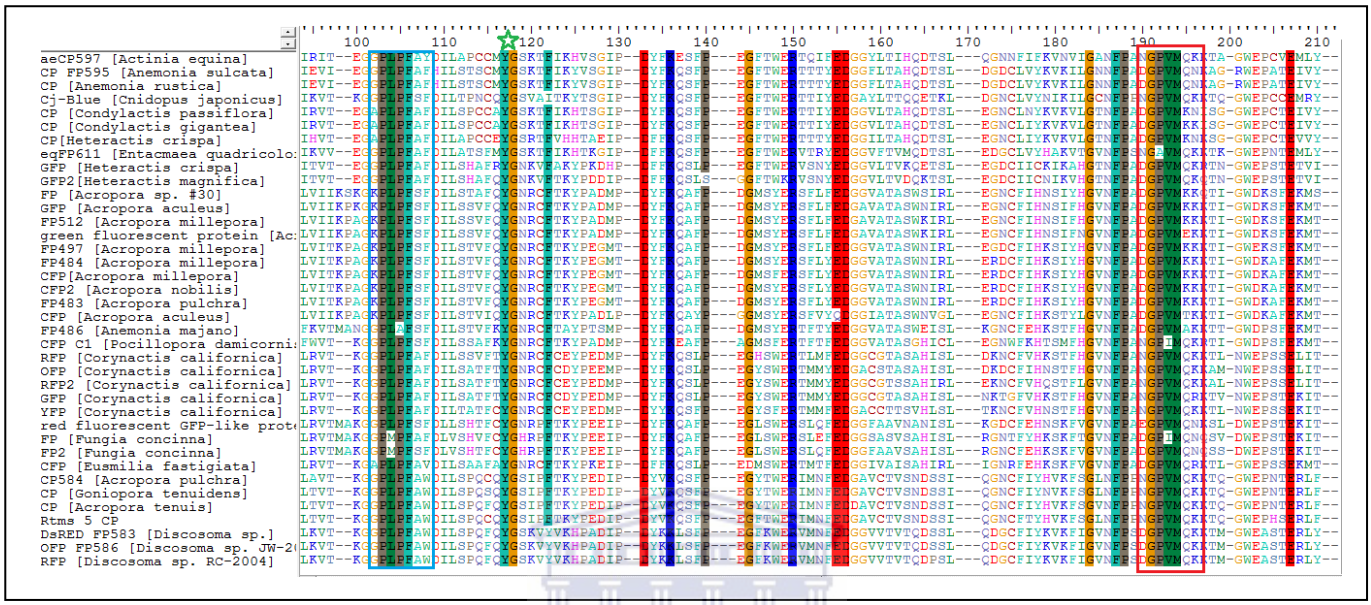


Figure 26. Exert of MAFFT alignment of various AFPs and ACPs. Blocked sequences are the sites that primers were designed to and star indicates the XYG motif of the tripeptide sequence.

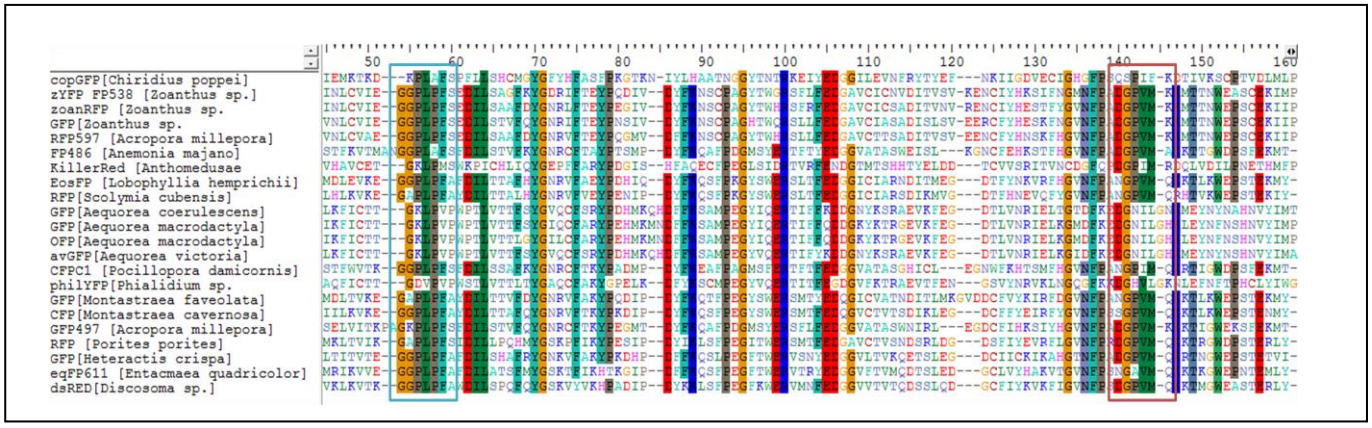


Figure 27. Exert of MAFFT alignment of fluorescent and chromoproteins from *Cnidaria*. Blocked regions show the difference in the conserved regions selected for primer design of AFP and ACPs screening.

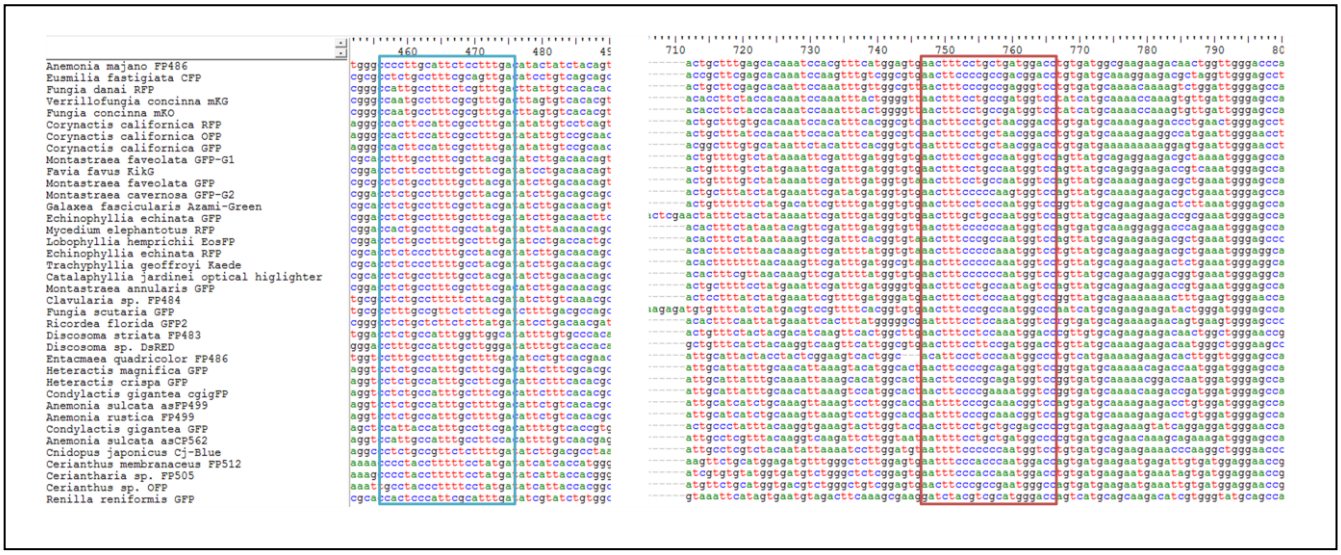


Figure 28. Exert of MAFFT alignment of nucleotide level to which primers were designed. The blue block is where the forward primers were designed to and the red block the reverse primers. Codon bias of each organism resulted in the degeneracy of the primers that were designed.

3.3.2. PCR optimization on dApc positive control

After being unable to procure an AFP or ACP as positive control, a synthetic positive control (dApc) was designed and synthesised to be used as template in a PCR reaction. The purpose of the positive control was to prove that the degenerate primers were indeed capable of ampilifying AFP and ACP cDNA targets. The synthetic positive control template was based on the sequence from *Actina equina* aeCP597 (blue chromoprotein) and formed the main scaffold. Additional conserved sequences from other fluorescent proteins (corresponding to the region the primers were designed to) were incorporated onto the scaffold sequence (Figure 29). The motivation for using the aeCP597 sequence was because *A.equina* is from the same Family (*Actiniidae*) as the sea anemones used in the libraries constructed. The synthesized positive control dApc (Biomatik, USA) and the expected amplicon sizes resulting from amplification using the various primer sets based on the sequence are shown in Figure 29.

The effects of the type of polymerase used, primer concentration, template amount and influence of non-target DNA were explored. It was observed that high fidelity polymerases like Phusion

(Finnzymes) (in GC buffer only), and high fidelity hot start polymerases such as PrimeSTAR (Takara) and AmpliTaq Gold (Applied Biosystems) were able to amplify the target from the positive control. The PrimerSTAR polymerase was most successful in amplification and was therefore, used in subsequent screening (unless otherwise stated).

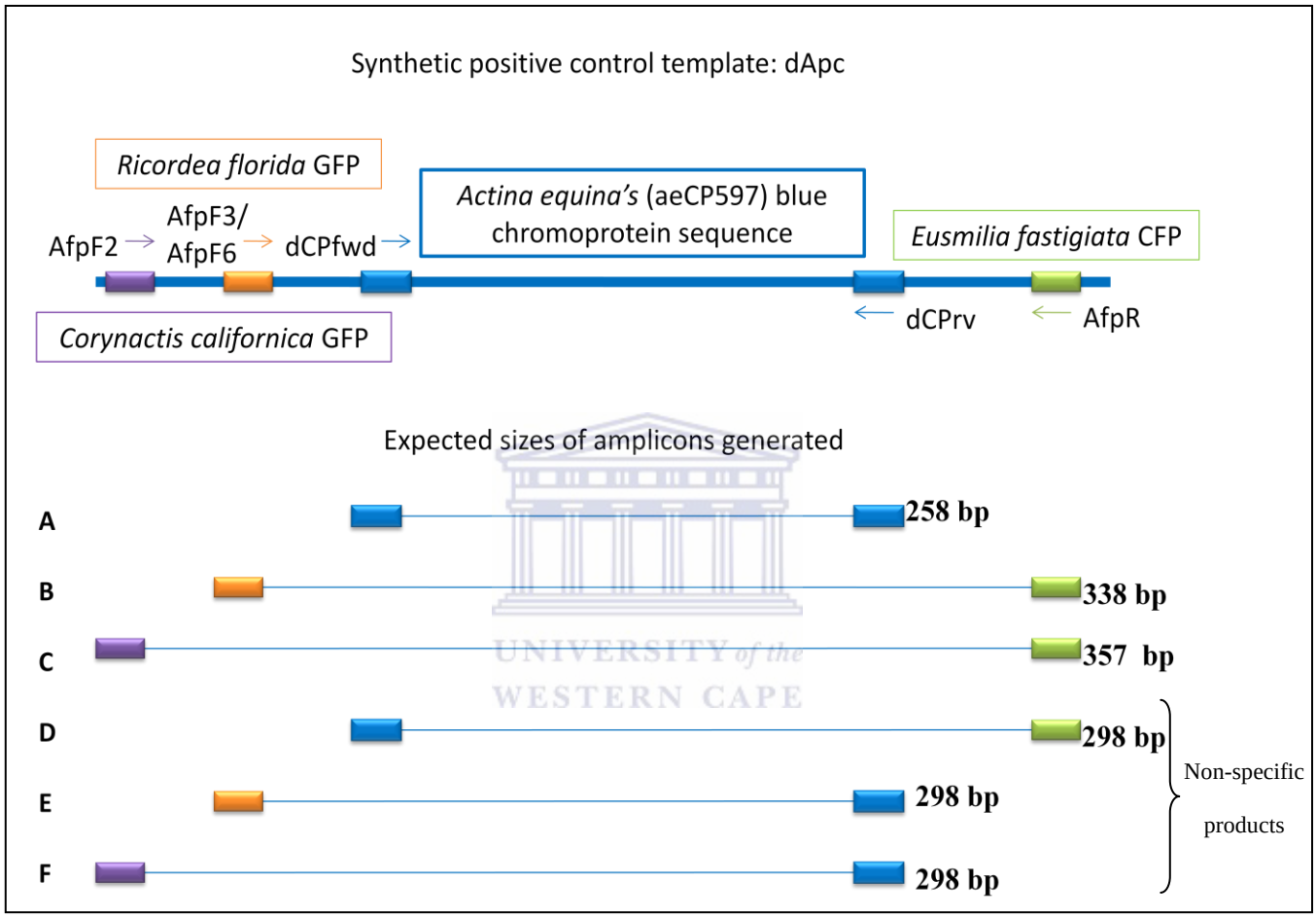


Figure 29. Positive control dApc and expected product sizes. Primer binding sites are indicated by respective arrows. The blue line and blocks represents sequence from *A. equina*'s aeCP957sequence. The various blocks represent where other non-aeCP597 sequence was inserted: purple-sequence from *Corynactis californica* GFP; orange-sequence from *Ricordea florida* GFP and green-AfpR (sequence from *Eusmilia fastigiata* CFP. A-F are possible amplicons.

The forward and reverse primer concentrations were individually increased from 0.2 µM to 0.8 µM and used at a concentration of 0.6 µM unless otherwise stated. Figures 30 and 31 illustrate that the primers were specific to AFPs and ACPs as they were unable to bind and amplify EGFP (jellyfish FP). However, the primer sensitivity was low as the minimum amount of template that could be successfully amplified was 0.5 ng (see Figure 32) which equals approximately to 1.27^{9} copies of the

target DNA added per reaction. In each of the various primer sets, however, a secondary non-specific product is amplified (Figure 30).

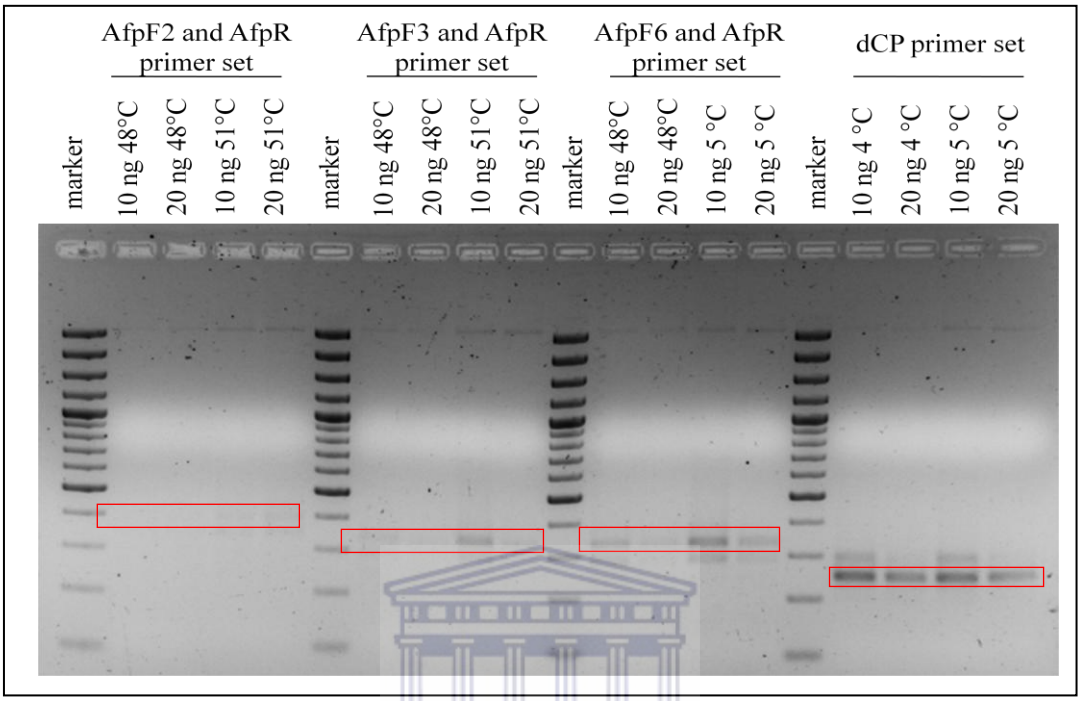


Figure 30. cDNA screening optimisation of template concentrations and annealing temperatures. Positive control dApc was used as template in a PCR screen using PrimeSTAR and 0.4 µM of each primer. The red boxes indicate expected amplicons. Marker is 100 bp plus (see Appendix C).

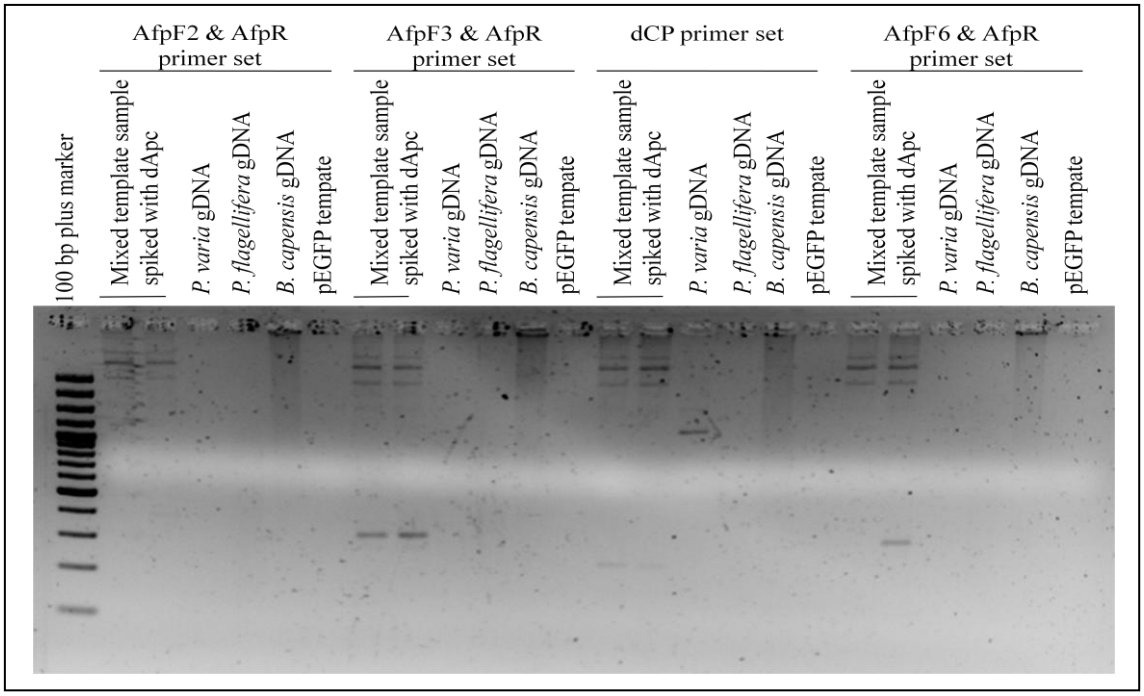


Figure 31. Assessment of degenerate primer sets' binding capability and specificity on different templates using PrimeSTAR and 0.4 µM of each primer. Marker is 100 bp plus (see Appendix C).

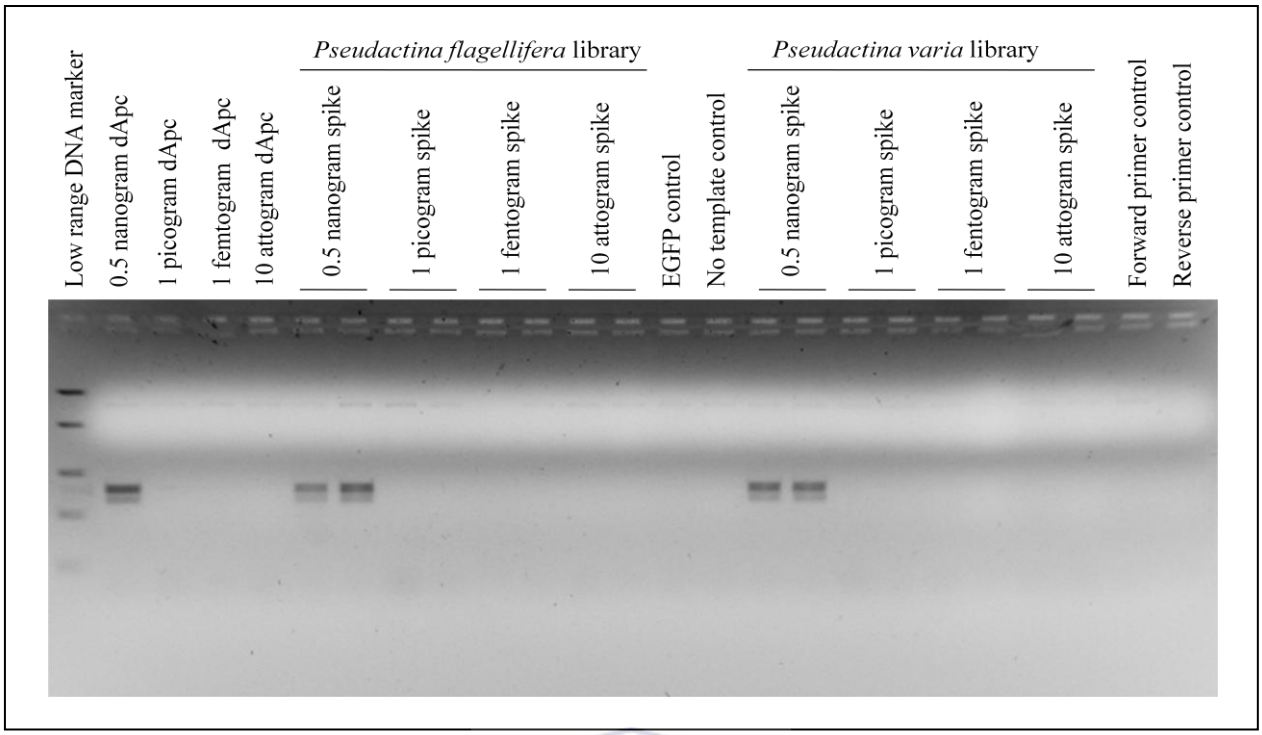


Figure 32. Determining detection sensitivity of dCP primer set within cDNA libraries by spiking with dApc- using PrimeSTAR and 0.6 µM of each primer. Low range DNA marker molecular sizes are in Appendix C.

3.3.3. PCR-based cDNA library screening using degenerate primers

PCR amplification using cloned cDNA as template with the respective primer sets under optimised conditions did not yield any observable product. This may have been due to the stringency of the cycling parameters or that the products were not at a level detectable by agarose gel electrophoresis. Therefore, touch-down PCR, a technique which accounts for the variations in the annealing temperatures of the degenerate primers was used to increase binding and amplification of target sequences. Additionally second round PCR, where a previous PCR reaction serves as template, was used to increase the yield of product so that it would be detectable by agarose gel electrophoresis. After applying both techniques amplicons of correct size as well as larger non-specific products were observed. Figure 33 shows that nonspecific products were amplified when cDNA was screened with various primer sets in a touchdown PCR using an initial annealing at 63°C decreasing at 1°C per cycle over 15 cycles thereafter. The alternative approach where PCR amplicons from an initial library screen were used (3 µL of "dirty" PCR reaction) in a second touch-down PCR yielded more

promising results. The products were cloned into pJET and the size was verified by amplifying the insert from the clone using pJET primers (Figure 34). The cloned products were sequenced and were found to be either primer dimers or non-specific products of approximately 250 bp and 1.2 kb.

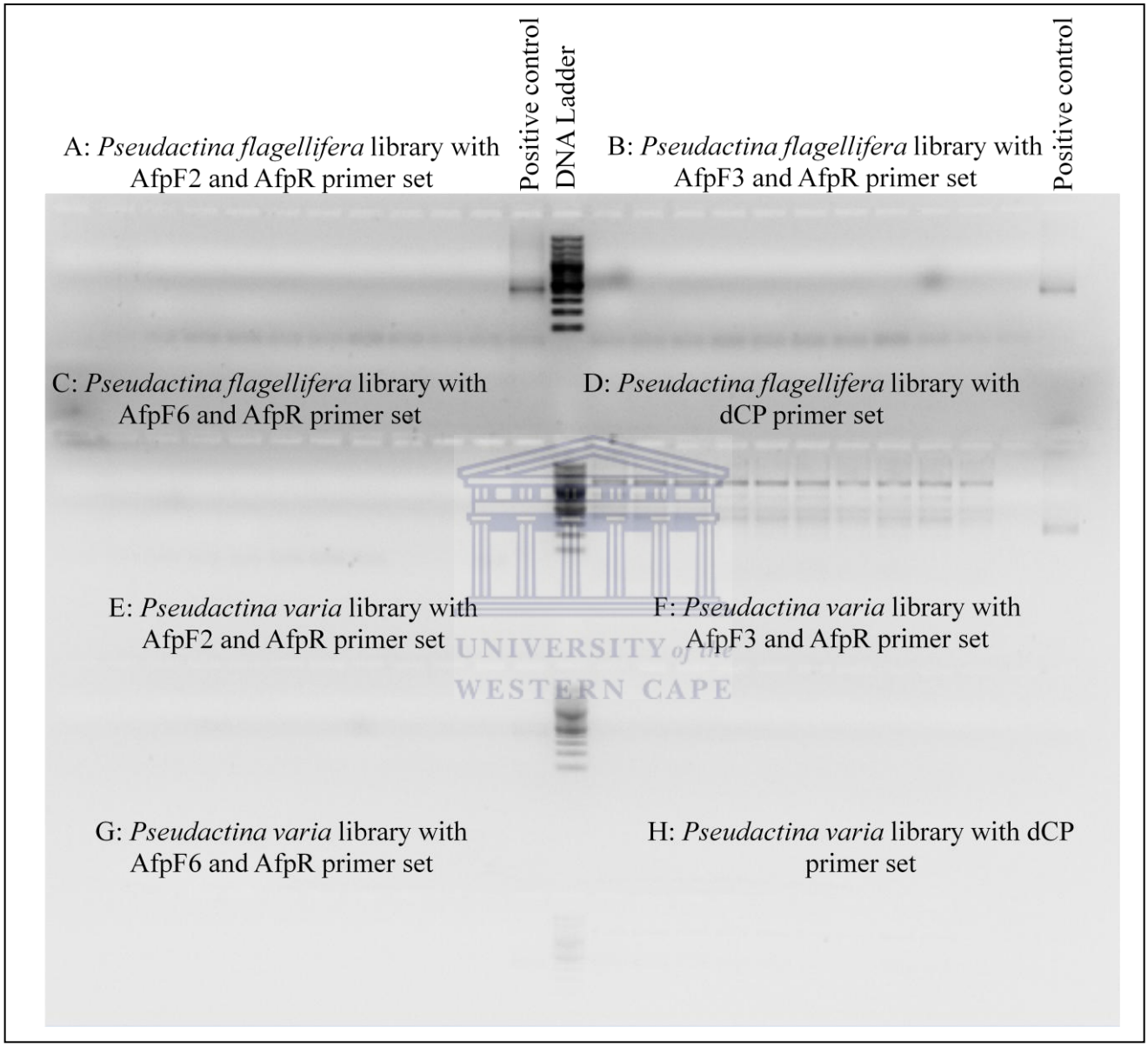


Figure 33. Results of the touch-down PCR screen on *P. varia* and *P. flagellifera* cDNA libraries reveal that amplification is observed only with the dCP primer set. The product size ranging from ~300-1200 bp was, however, an indication the amplification is non-specific. DNA ladder is 100 bp O'generuler (see Appendix C).

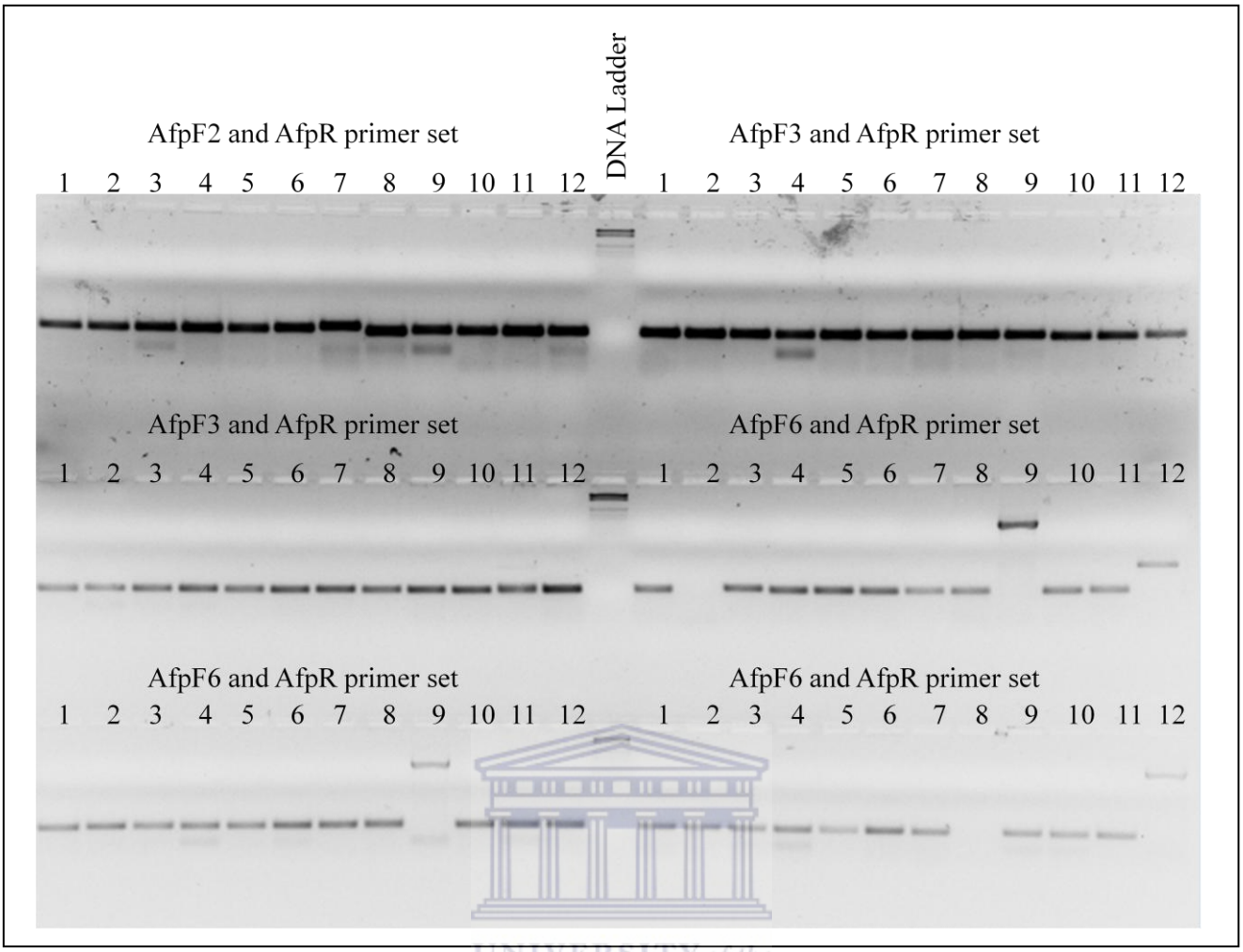


Figure 34. Products from a PCR with cDNA template and dCP primer set which was used in a subsequent touch-down PCR with various primer sets. The resulting amplicons where cloned into pJET and colony PCR with vector primers was performed. Clones A7, A9, C11, D9 and D12 were sequenced. DNA ladder is 100 bp O'generuler (see Appendix C)

3.3.4. PCR-based screening of gDNA using degenerate primers

The gDNA was screened using the previously designed degenerate primers in an attempt to recover a partial portion of the gene for AFP or ACP. Figure 35 shows a 1.2 kb product was consistently amplified in each sample using the dCP primers in a touch-down PCR protocol using initial annealing conditions between 70.5-71.5°C (depending on primer and gDNA sample used) with a decrease of 1°C/cycle for 15cycles.

The amplified fragment was cloned (see Figure 36) and subjected to Restriction Fragment length polymorphism (RFLP) using *Alu*I. The RFLP fingerprints, as seen in Figure 37, delineate three distinctive phylotypes with one being unique to *B. capensis*. One clone of each phylotype from each

of the sea anemone samples was sequenced. Two additional clones of each phylotype originating from the *P. varia* gDNA were also sequenced to confirm that the RFLP patterns were indeed unique sequences.

Sequence comparisons using the BLASTx and tBLASTx algorithms against that the NCBI database revealed the gDNA amplicons were nonspecific PCR products. For all sequences, except one, the closest BLAST relative was a viral insert similar to gag-pol of *Schistosoma japonicum*. However, the unique clone from *B. capensis* showed similarity to *N. vectensis* Zinc finger proteins.

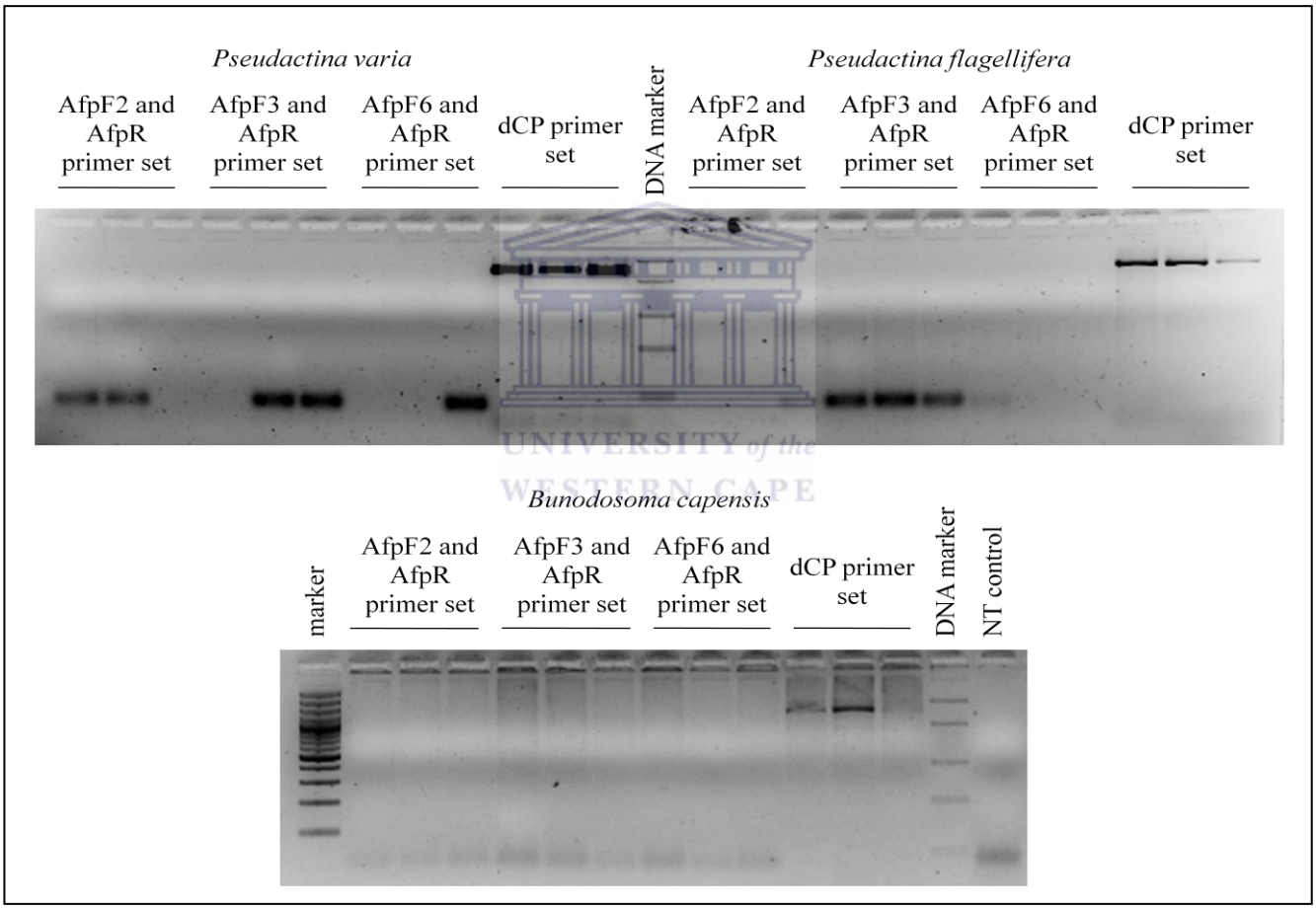


Figure 35. Gel images of the degenerate PCR using touch-down PCR on gDNA from sea anemones. Low range DNA marker molecular sizes and DNA ladder is 100 bp O'generuler are in Appendix C.

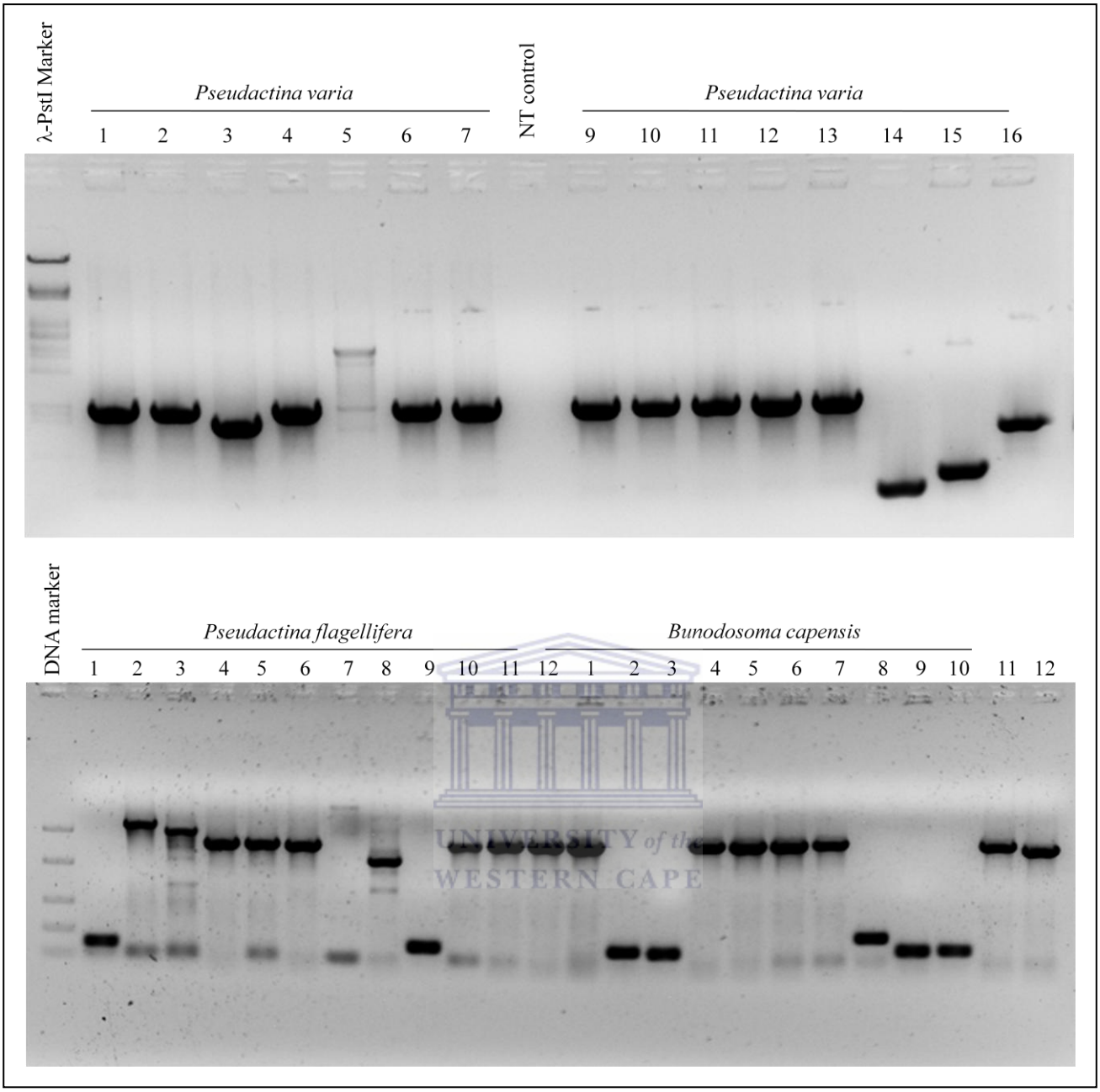


Figure 36. A and B are gel images of a pJET colony PCR – the amplicons from clones containing the correct insert size were subjected to RFLP using *Alu*I. Molecular weight markers λ-*Pst*I and Low range DNA marker sizes are in Appendix C.

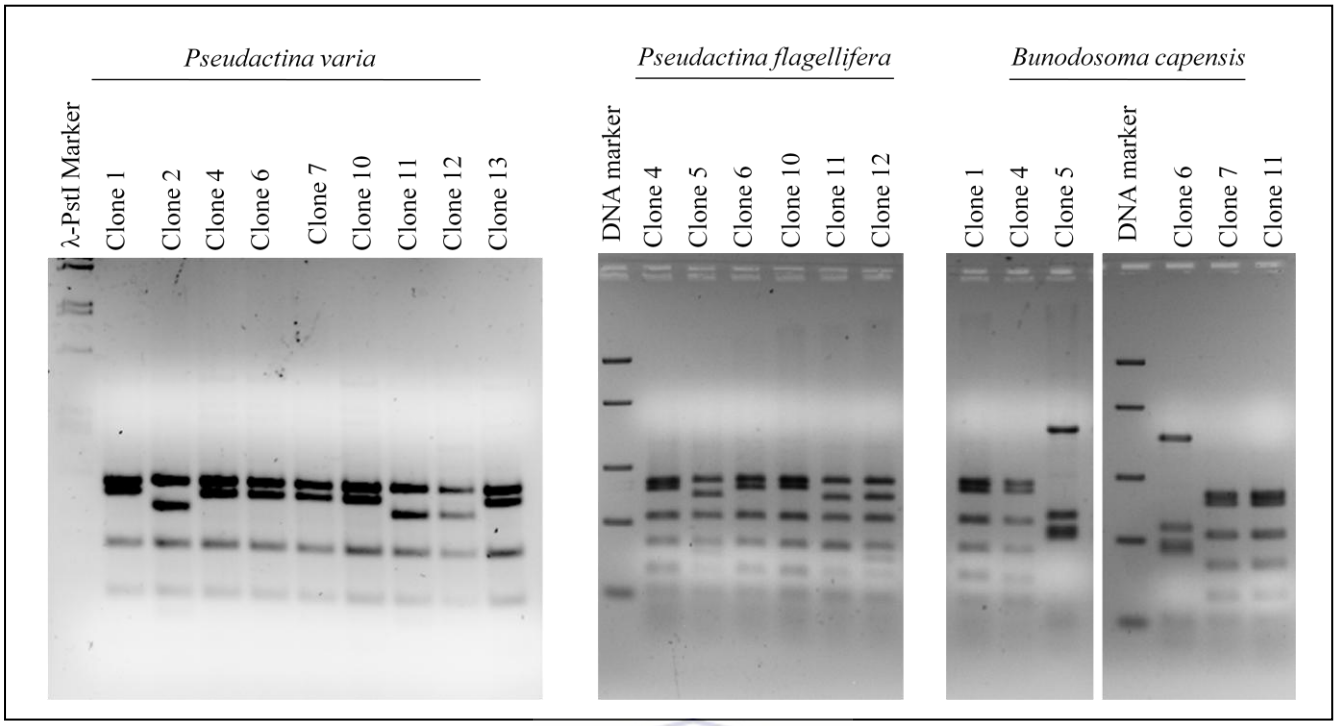


Figure 37. RFLP fingerprints of amplicons from gDNA_dCP. Gel images are labelled accordingly and the molecular weight markers λ-*Pst*I and Low range DNA marker sizes are in Appendix C.

3.4. Construction of Functional cDNA libraries

The following sections will cover the construction of cDNA libraries for activity based screening in expression vectors pLINKet and pEGFP-N3. The linker molecules necessary for the In-Fusion cloning are added by inverse PCR to the chosen expression vector. The pLINKet vector was synthesised in the course of this study to be used in combination with the In-Fusion SMARTer cDNA synthesis library construction kit (Clonetech) as the kit does not have a functional vector. pEGFP-N3 is a commercially available vector that encodes a fluorescent protein that is used as a tag to track protein expression and localization in cells. The fluorescent protein was excised and cDNA was cloned in its place. In order to accomplish this, the SMARTer protocol was modified as later described.

3.4.1. Functional library construction: pLINKet synthesis

A single 5 473 bp product was amplified under the following experimentally determined conditions (see Figures 38 and 39): 2 ng pET21a, 0.4 µM of each primer, 0.5 mM dNTPs, 1 X HF phusion

buffer and 1 unit of Phusion in 100 µL final volume. The cycling parameters were initial denaturation at 96°C for 60 seconds followed by 25 cycles of denaturation at 95°C 10 seconds, annealing at 60°C for 20 seconds and extension at 68°C for 5 minutes and 40 seconds ending with final extension at 68°C for 5 minutes.

Restriction endonuclease digestion of the linearized vector using *Xba*I and *Hin*dIII removed the linker molecules that were added during the PCR. The vector was then re-circularised using T4 ligase and transformed into *E. coli* GeneHogs and BL21 DE3 pLysS. The transformants obtained serve as an indication that the origin of replication and antibiotic resistance marker of the vector were still functional. Sequencing of the vector (Figures 38 and 39) revealed that the linker addition was successfully incorporated and in the correct orientation for In-Fusion cloning.

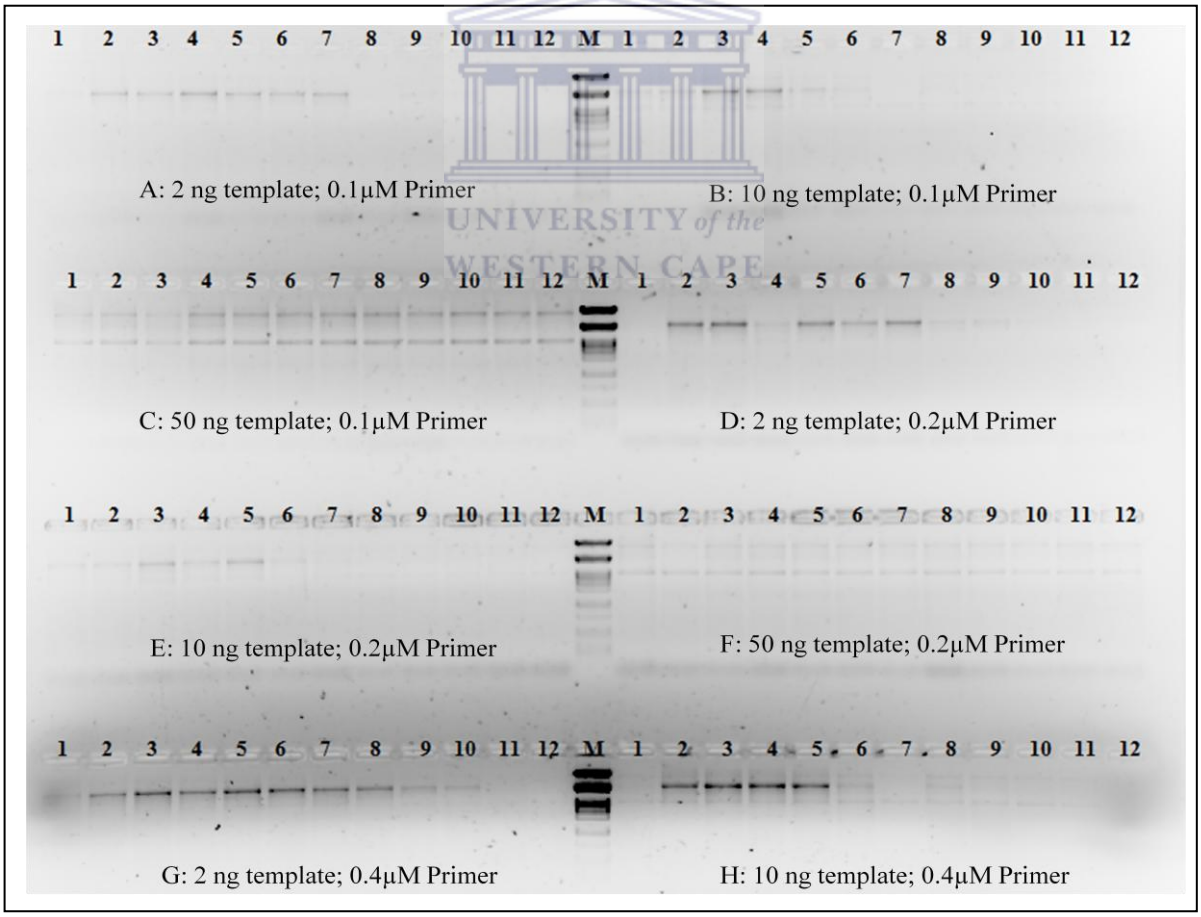


Figure 38. Optimization of pLINKet PCR synthesis. Effect of amount of template and primer concentration; lanes 1-12 represent a gradient of annealing temperatures from 52-62°C. Molecular weight marker λ-*Pst*I is in Appendix C.

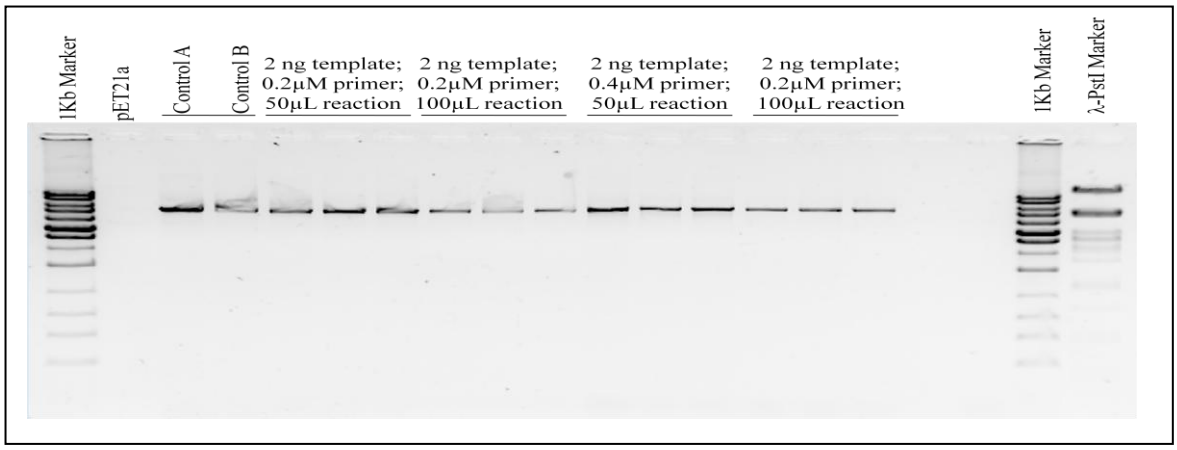


Figure 39. Optimization of PCR reaction volume and primer concentration with annealing temperature fixed at 60°C. Control A-10 µl of and Control B-5 µL loaded from a previous PCR optimization in a final volume of 25 µL for comparative analysis. Molecular weight marker λ-*Pst*I is in Appdenix C.

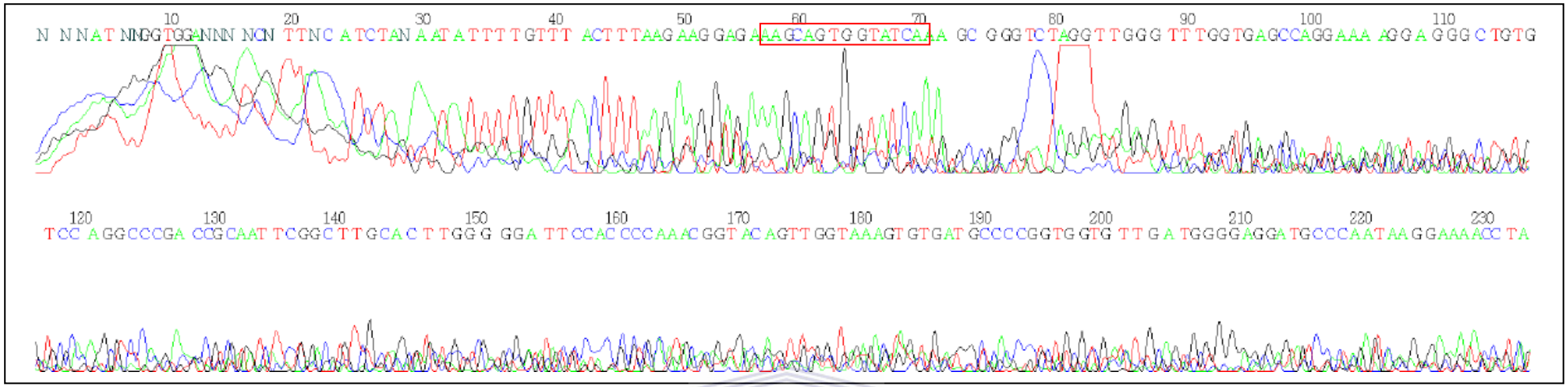


Figure 40. pLINKet sequencing results using the T7 promoter primer- Linker molecule sequence incorporated is indicated in red box.

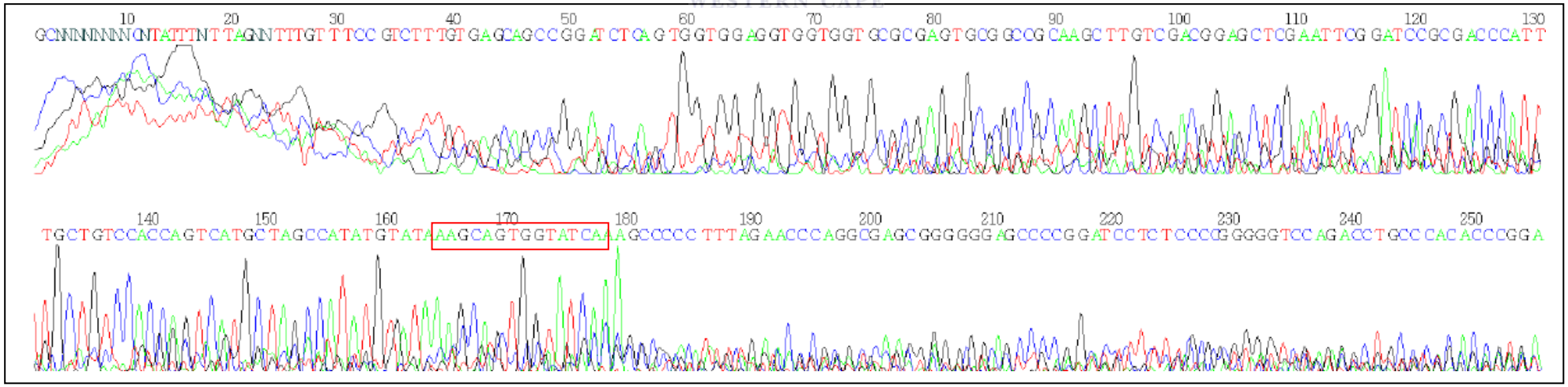


Figure 41. pLINKet sequencing results using the T7 terminator primer- Linker molecule sequence incorporated is indicated in red box.

3.4.2. pLINKet cloning using positive control *LacZ*

Attempts to clone with the synthesized pLINKet were largely unsuccessful. The transformation and cloning controls showed that the cells were highly competent (approximately $4.7X10^8$ CFU/µg of DNA), viable and sensitive to ampicillin. Ligation controls included reactions where either no enzyme or no insert were added. These respective controls were used to verify that the parental pET21a was digested to completion and that pLINKet was not re-circularising.

Of the several vector preparations that were synthesised and used in cloning experiments, only one of the preparations yielded parental colonies. The sequencing results of this vector preparation are shown in Figures 38 and 39. The rapid peak drop just after the linker sequence suggests that the vector product from the PCR synthesis is indeed a linear molecule. However, the continuation of the sequence after the linker region confirms that this preparation of vector did contain a proportion of parental molecules.

3.4.3. Functional library construction: modified SMARTer™ cDNA synthesis

As an alternative to the In-Fusion library cloning, conventional restriction digest cloning was used to construct libraries in pEGFP-N3. For this the cDNA synthesis needed to be modified to incorporate the restriction sites into cDNA linker molecule ends. The cDNA synthesis was successful as products were visualized by gel electrophoresis. Digested cDNA was cloned into pEFGP-N3 and successfully transformed into *E.coli* BL21 DE3 pLysS as colonies were obtained on LB agar supplemented with kanamycin (as in Appendix A). However, the libraries consisted mainly of vector without insert as the vector had not been completely digested and in the process of cloning had simply re-ligated. This was confirmed by both colony PCR using the modified SMARTer cDNA primers (Figure 42) and digestion of plasmid DNA using *Not*I and *Sal*I (Figure 43).

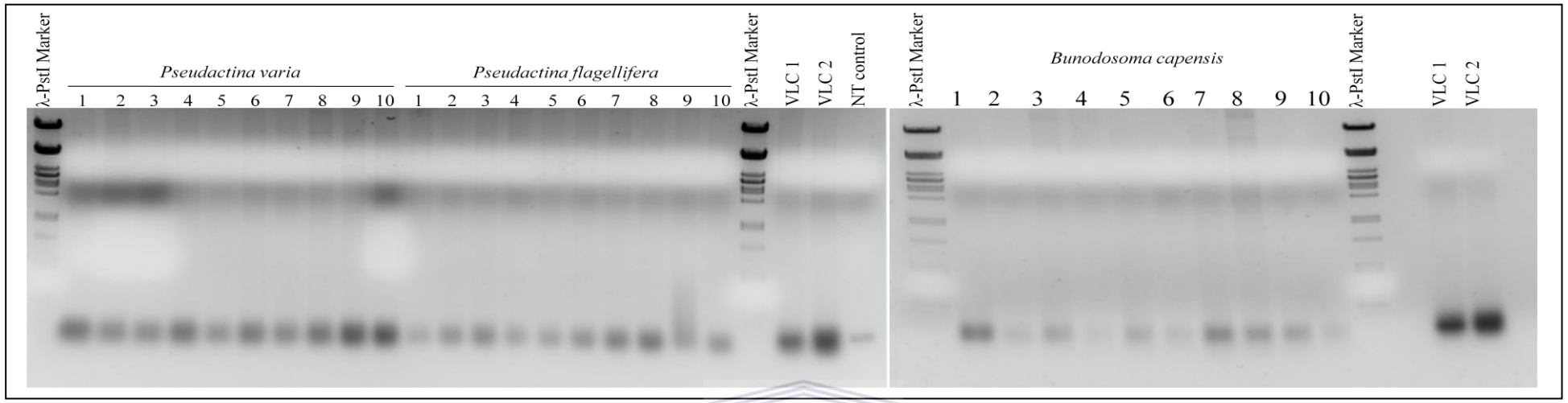


Figure 42. Colony PCR from each functional library created through the modified cDNA synthesis. No amplification was observed in any of the experimental reactions indicating no inserts were taken up. VLC -vector ligation control showing the background colonies were due to re-ligation of a partial vector digest. Molecular weight markers λ-*Pst*I is in Appdenix C.

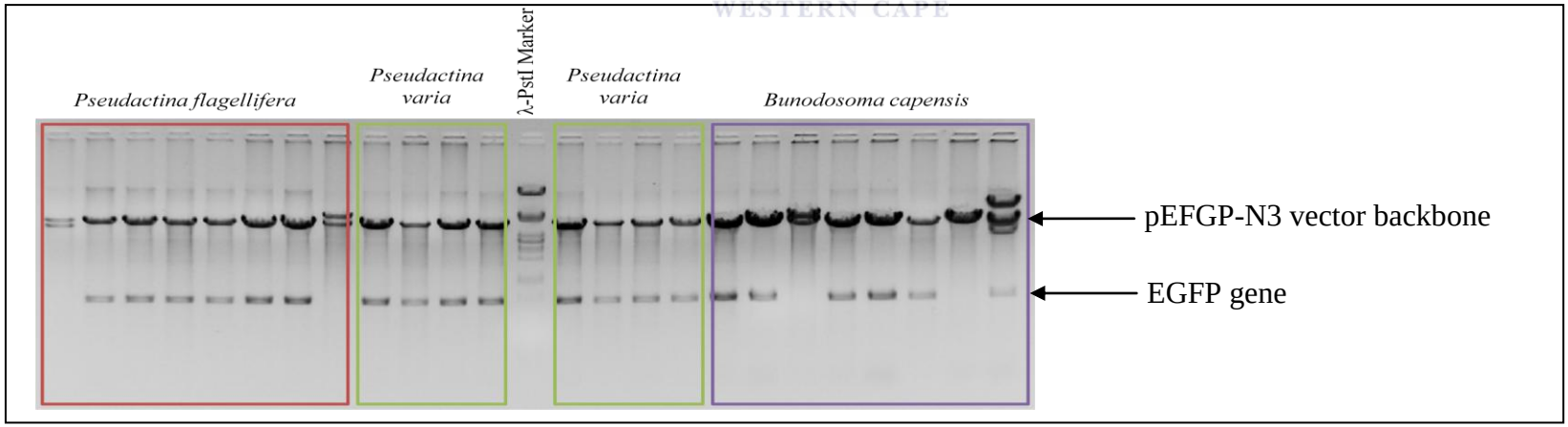


Figure 43. Plasmid DNA was extracted from individual colonies and digested with Not*I* and Sal*I*. The resulting digest confirms no insert was taken up and that partially digested vector had re-ligated. Molecular weight marker λ-*Pst*I sizes is in Appdenix C.

3.5. Discussion

The use of new cDNA technologies was evaluated with respects to its application in eukaryotic gene discovery with special focus on cloning AFP and ACP. The advances in both cDNA synthesis and library construction using novel methods were considered. cDNA technology is a fickle technique that requires high level transcription of the target gene (especially in the case of rare or less abundant transcripts), if this is not the case the likelihood of finding the target using this (cDNA) approach is unlikely (Gilbert et al. 2008). A specific and sensitive screen, especially when sequence homology based screens are performed, is critical to success (Stangegaard et al. 2006).

Unlike for genomic DNA library construction where partial restriction digests are performed to ensure the cloning of complete open reading frames, cDNA needs to be fully digested before cloning. This is problematic as there is the likelihood that the particular restriction enzymes chosen could cut within the target cDNA sequence. While prior knowledge of sequence would help, mutations and sequence variation between genes of interest are common and thus there would be no guaranteed prediction that a particular target gene would not be cut within the target sequence.

The addition of linker (or adapter) molecules to cDNA and ligation independent cloning is gaining popularity in the field of eukaryotic gene discovery (Bystrova et al. 2010). Linker addition negates the requirement for restriction digestion, while ligation independent cloning does not have the disadvantage of low efficiency cloning as with blunt end ligation. In the past, linker addition was a tedious and problematic stage during cDNA synthesis. The advances in these techniques over the years can be observed in the development of kits utilizing a mutant of the M-MLV reverse transcriptase with a deletion in its RNaseH activity (Georgiadis et al. 1995; Das and Georgiadis 2004; Izuno et al. 2010). This reverse transcriptase demonstrates the ability to remove the mRNA 5' cap and adds a sequence onto this end. This has allowed researchers to add linker molecules to both ends during a single reaction when synthesizing cDNA (Song et al. 2010). This maximizes yield and adds sensitivity to the construction of cDNA libraries. Old techniques required the synthesis of cDNA before linkers could be attached. These extra steps and complications to the process of

constructing cDNA libraries resulted in contamination and loss of yield. The advances in ligation independent cloning, when combined with the SMARTer^{TM} cDNA synthesis, removed the problem of restriction enzyme cloning. Additionally, only transcripts with a 5' cap are converted in the stage of second strand synthesis and thus the libraries created are selective for full length transcripts (Zhu et al. 2001).

The In-Fusion SMARTer cDNA library construction kit has proven to be an effective technique for creating libraries needing to be screened based on sequence homology. The most important considerations when using this method would be the preparations prior to cDNA synthesis: extraction of high quality mRNA; design and optimisation of primers or probes for the screening approach.

The SMARTer^{TM} cDNA synthesis is applicable for the next stage of eukaryotic gene discovery, namely deep sequencing of cDNA. A recent study has found this approach to gene discovery to be very promising (Johansen et al. 2010). The workflow outlined in the paper was to synthesise cDNA that could be sequenced using next-generation sequencing and following intensive bioinformatic analysis of select targets for expression trials. The SMARTer^{TM} synthesis could be modified and applied to this approach by incorporating the linker molecules necessary for sequencing in the first-strand reaction. As both linker molecules are attached in a single step directly to the starting material the bias introduced by PCR and other intermediary steps can be negated. The need to construct a library falls away and the screening process is streamlines.

In this study there are 3 possible reasons why no positive targets were identified. Firstly, the amount of starting material used in the RNA extraction was 5 times that recommended in the literature. This may mean the target transcripts were "diluted" to levels that could not be detected with the primers that were designed. Pigmented tissue (i.e. tissue that was brightly coloured or fluorescent) was used for the RNA extractions in the likelihood of increasing the presence of the transcripts for the protein targets. However, it has been noted that the AFP and ACP are extremely stable and possess a relatively long half-life (Leutenegger et al. 2007). The possibility that the mRNA transcripts for any

AFP or ACP were no longer in the cells is thus a likely explanation as to why the targets were not detected in the sequence based screens. This highlights the second reason that the primer design is of utmost importance when screening based on sequence homology. Even though the sequences were highly conserved at the amino acid level, the nucleic acid level was fairly degenerate due to codon biases in different species. The primers designed may not have been sensitive enough to detect the respective targets.

The cDNA libraries constructed to be screened by activity were also unsuccessful for various reasons. The libraries in pLINKet, a functional vector for In-Fusion^{TM} cloning, were unable to clone any insert. The In-Fusion^{TM} cloning is a form of ligation independent cloning. The enzymes involved in this process are polymerases that have exonuclease activity. The In-Fusion^{TM} enzyme has 3' exonuclease activity that creates sticky ends by removing the nucleotides at the end of the 3' strand. For this process to occur, the 3' ends need to be hydroxylated (Aslanidis et al. 1994; Li and Evans 1997; Berrow et al. 2007). One possibility that was not explored for the cloning failure was to test if the 3' ends of pLINKet were indeed hydroxylated. Libraries constructed in pEGFP-N3 and screened for activity yielded only background colonies. The most likely possibility for this was that the vector was not completely digested, resulting in religation.

Possible alternative routes that could have been taken in the sequence based screening approach are: Use the degenerate primers listed in some other published studies; redesign of primers using a close relative to make the screen more sensitive and specific. With regards to the libraries constructed for activity based screening, a different kit (using a different cDNA synthesis technique) could have been used. However, the kits used in the published studies have been discontinued. An alternative route would be to use an organic extraction to recover the fluorescent or chromoproteins and subject them to mass spectrophotometery. The resulting protein sequence could be used to design new primers specific for the targeted AFPs or ACPs to re-screen the cDNA libraries.

4. Conclusion

Addressing the origins and purpose of AFP and ACPs is difficult. It is clear they serve in a vast and variable capacity within the cell; however the suggested functions thus far are supported by weak and ambiguous data leaving the biological role of these proteins open to debate. The phylogenetics of both the proteins as well as the host organisms is unresolved (Shagin et al. 2004; Alieva et al. 2008; Daly et al. 2008). The poly and para-phyletic nature within the *Order* of sea anemones, in particular, are suggestive of homoplasy; i.e., multiple origins. If this is the case then it would not be surprising that the gene repertoire of these would show a similar trend. The possibilities of convergent evolution and atavism should be taken into consideration for future phylogenetic studies. Furthermore, linking host relation and environmental factors instead of looking at the proteins in isolation may provide more detailed insight as to biological significance and origin of AFP and ACP.

Is the pursuit of novel fluorescent and chromoproteins from Anthozoans (or in general) a worthy research interest? From a blue biotechnology viewpoint, recovery of novel AFPs or ACPs is not a high priority research project. Over 100 fluorescent and chromoproteins have been discovered and characterized (in part or fully) and the most commonly used remain *A. victoria* mutants. Furthermore, the lack of knowledge about the biological role has frustrated efforts in generating mutants of AFPs and ACPs with desired traits. Furthermore, fluorophore dyes and bioluminescent systems are well established alternatives to fluorescent proteins as markers or tags. In terms of biological research, new FPs and CPs need to be considered not only in context of their protein classification but also in perspective with the host and environmental circumstances. The lack of insight into the cellular biology of sea anemones, their associated symbionts and the role of FPs and CPs in those interactions would surely make for interesting scientific research.

Marine biotechnology in South Africa should be given serious consideration. The infrastructure for developing a sustainable biotechnology industry is currently under construction. The lack of a developed marine biotechnology industry is a reflection of limited attention in this field. Developments in the facilities, tools and necessary expertise to study marine biotechnology needs to

be made, but the basis for all of this already exists and just needs to be encouraged. The potential for a prosperous marine community in South Africa is realistic considering the unique coastal waters that host a wide biodiversity including many endemic species. The proximity of national research institutes, housed in Universities, to industrial areas allows the interaction between industry and the science community to be fostered.

The National Research foundation has put into place the funding for marine biotechnology projects. The Knowledge development fund hosts a program, SEAchange, specific for funding marine interests that enable technological development. In this regard, marine biotechnology may become an avenue used to fulfill future government mandates in terms of national biotechnology strategy. The great demand for novel pharmaceuticals, nutraceuticals, bioploymers and pesticides can be met if the scope for screening is widened to include marine sources. To accomplish this though the tools to exploit marine biotechnology as well as the expertise in this field needs to be developed at a more accelerated pace.

5. References

1. Adam V, Lelimousin M, Boehme S, Desfonds G, Nienhaus K, Field MJ, Wiedenmann J, McSweeney S, Nienhaus GU, and Bourgeois D. (2008) "Structural characterization of IrisFP, an optical highlighter undergoing multiple photo-induced transformations." *PNAS* 105, no. 47: 18343-8.
2. Alfandari D, and Darribere T. (1994) "A simple PCR method for screening cDNA libraries." In *Pcr Methods And Applications*, 4:46-49. Cold Spring Harbor Laboratory Press.
3. Alieva NO, Kozen KA, Field SF, Meleshkevitch EA, Hunt ME, Beltran-Ramirez V, Miller D, Wiedenmann J, Salih A, and Matz MV. (2008) "Diversity and evolution of coral fluorescent proteins." *PLoS ONE* 3, no. 7: 1-12.

4. Allen MJ, and Jaspars M. (2009) “Realizing the potential of marine biotechnology.” *Industrial Biotechnology* 5, no. 2: 77-83.
5. Andersen JB, Roldgaard BB, Lindner AB, Christensen BB, and Licht TR. (2006) “Construction of a multiple fluorescence labelling system for use in co-invasion studies of *Listeria monocytogenes*.” *BMC microbiology* 6, no. 86.
6. Ando R, Hama H, Yamamoto-Hino Mi, Mizuno H, and Miyawaki A. (2002) “An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein.” *PNAS* 99, no. 20: 12651-6.
7. Ando R, Mizuno H, and Miyawaki A. (2004) “Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting.” *Science* 306, no. 5700: 1370-3.
8. Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, and Weightman AJ. (2006) “New Screening Software Shows that Most Recent Large 16S rRNA Gene Clone Libraries Contain Chimeras.” *Applied and Environmental Microbiology* 72, no. 9: 5734-5741.
9. Aslanidis C, de Jong PJ, and Schmitz G. (1994) “Ligation-independent Cloning (LIC) of PCR Products Minimal Length Requirement of the Single-stranded Tails for.” *Genome Research* 4: 172-177.
10. Asselbergh I, Flors C, Ferrighi L, Botek E, Mizuno H, Ando R, Miyawaki A, Hofkens J, Van Der Auweraer M, and Clays K. (2008) “Second-Harmonic Generation in GFP-like Proteins.” *JACS* 130, no. 46: 15713-15719.
11. Aury JM, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM, Aiach N, Arnaiz O, Billaut A, Duharcourt S, Guigo R, Beisson J, Blanc I, Bouhouche K, Ca F, Gogendeau D, Katinka M, Keller A, Kissmehl R, Klotz C, Koll F, Malinsky S, Nowacki M, Nowak JK, Plattner H, Lepe G, Moue A, Serrano V, Zagulski M, Dessen P, Be M, Sperling L, Meyer E, Cohen J, Weissenbach J, Scarpelli C, Scha V, Wincker P. (2006) “Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*.” *Nature* 444: 171-178.

12. Bailly J, Fraissinet-tachet L, Verner M, Debaud JC, Lemaire M, Wésolowski-Lovel M, Marmeisse R. (2007) "Soil eukaryotic functional diversity , a metatranscriptomic approach." *The ISME Journal*: 632-642.
13. Baird GS, Zacharias DA, and Tsien RY. (2000) "Biochemistry , mutagenesis , and oligomerization of DsRed, a red fluorescent protein from coral." *PNAS* 97, no. 22: 11984-11989.
14. Barondeau DP, Kassmann CJ, Tainer J, and Getzoff ED. (2005) "Understanding GFP chromophore biosynthesis: controlling backbone cyclization and modifying post-translational chemistry." *Biochemistry* 44, no. 6: 1960-70.
15. Barondeau DP, Putnam CD, Kassmann CJ, Tainer J, and Getzoff ED. (2003) "Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures." *PNAS* 100, no. 21: 12111-6.
16. Baumann D, Cook M, Ma L, Mushegian A, Sanders E, Schwartz J, and Yu CR. (2008) "A family of GFP-like proteins with different spectral properties in lancelet *Branchiostoma floridae*." *Biology Direct* 3, no. 28.
17. Beddoe T, Ling M, Dove S, Hoegh-Guldberg O, Devenish RJ, Prescott M, and Rossjohn J. (2003) "The production, purification and crystallization of a pocilloporin pigment from a reef-forming coral." *Acta Cryst.* D59: 597-599.
18. Berrow NS, Alderton D, Sainsbury S, Nettleship J, Assenberg R, Rahman N, Stuart DI, and Owens RJ. (2007) "A versatile ligation-independent cloning method suitable for high-throughput expression screening applications." *Nucleic acids research* 35, no. 6.
19. Bogdanov AM, Mishin AS, Yampolsky IV, Belousov VV, Chudakov DM, Subach FV, Verkhusha VV, Lukyanov S, and Lukyanov KA. (2009) "Green fluorescent proteins are light-induced electron donors." *Nature Chemical Biology* 5, no. 7: 459-461.
20. Bogdanov AM, Mishin AS, Yampolsky IV, Belousov VV, Chudakov DM, Subach FV, Verkhusha VV, Lukyanov S, and Lukyanov KA. (2010) "Green fluorescent proteins are light-induced electron donors." *Nature Chemical Biology* 5, no. 7: 459-461.

21. Bomati EK, Manning G, and Deheyn DD. (2009) “Amphioxus encodes the largest known family of green fluorescent proteins, which have diversified into distinct functional classes.” *BMC Evolutionary Biology* 9, no. 77.
22. Bridge D, Cunnigham CW, DeSalle R, and Buss LW. (1995) “Class-level relationships in the phylum *Cnidaria* molecular and morphological.pdf.” *Mol. Biol. Evol.* 12, no. 4: 679-689.
23. Bulina ME, Chudakov DM, Mudrik NN, and Lukyanov KA. (2002) “Interconversion of Anthozoa GFP-like fluorescent and non-fluorescent proteins by mutagenesis.” *BMC Biochemistry* 3, no. 7.
24. Bussow K, Nordhoff E, Lubbert C, Lehrach H, and Walter G. (2000) “A Human cDNA Library for High-Throughput Protein Expression Screening.” *Genomics* 65: 1- 8.
25. Bystrova MF, Romanov RA, Rogachevskaja OA, Churbanov GD, and Kolesnikov SS. (2010) “Functional expression of the extracellular-Ca_2^+-sensing receptor in mouse taste cells.” *Journal of Cell Science* 123: 972-982.
26. Campbell TN, and Choy FYM. (2002) “Approaches to Library Screening.” *J. Mol. Microbiol. Biotechnol.* 4, no. 6: 551-554.
27. Chalfie M, Tu Y, Euskirchen G, Ward WW, and Prasher DC. (1994) “Green fluorescent protein as a marker for gene expression.” *Science* 263: 802–805.
28. Chan MCY, Karasawa S, Mizuno H, Bosanac I, Ho D, and Prive GG. (2006) “Structural Characterization of a Blue Chromoprotein and Its Yellow Mutant from the Sea Anemone *Cnidopus Japonicus*.” *Journal of Biological Chemistry* 281, no. 49: 37813-37819.
29. Chomczynski P, and Sacchi N. (1987) “Single-Step Method of RNA Isolation by Acid Guanidinium Thiocyanate-Phenol-Chlorofrom Extraction.” *Analytical biochemistry* 162: 156-159.
30. Chudakov DM, Chepurnykh TV, Belousov VV, Lukyanov S, and Lukyanov KA. (2006) “Fast and Precise Protein Tracking Using Repeated Reversible Photoactivation.” *Traffic* 7, no. 10: 1304-1310.

31. Chudakov DM, Lukyanov S, and Lukyanov KA. (2005) “Fluorescent proteins as a toolkit for in vivo imaging.” *Trends in biotechnology* 23, no. 12: 605-613.
32. Chudakov DM, Lukyanov S, Lukyanov KA. (2007) “Using photoactivatable fluorescent protein Dendra2 to track protein movement Tech Insight.” *Biotechniques* 42, no. 5: 553-563.
33. Cody CW, Prasher DC, Westler WM, Prendergast FG, and Ward WW. (1993) “Chemical structure of the hexapeptide chromophore of the *Aequorea* green-fluorescent protein.” *Biochemistry* 32, no. 5: 1212-8.
34. Cubitt A, Heim R, Adams SR, Boyd AE, Gross LA, and Tsien RY. (1995) “Understanding, improving using green fluorescent proteins.” *Molecular Biology*, no. November: 2967-2971.
35. Dahllöf, I. (2002) “Molecular community analysis of microbial diversity.” *Current Opinion in Biotechnology* 13, no. 3: 213-217.
36. Daly M, Chaudhuri A, Gusmão L, and Rodríguez E. (2008) “Molecular Phylogenetics and Evolution Phylogenetic relationships among sea anemones (*Cnidaria* : *Anthozoa* : *Actiniaria*).” *Molecular Phylogenetics and Evolution* 48: 292-301.
37. Daly M, Brugler MR, Cartwright P, Collins AG, Dawson MN, Fautin DG, France SC, Mcfadden CS, Opresko DM, Rodriguez E, Romano SL, Stake JL. (2007) “The phylum *Cnidaria*: A review of phylogenetic patterns and diversity 300 years after *Linnaeus*.” *Zootaxa* 182: 127 - 182.
38. Daly M, Fautin DG, and Cappola VA. (2003) “Systematics of the Hexacorallia (*Cnidaria* : *Anthozoa*).” *Zoological Journal of the Linnean Society* 139: 419-437.
39. Das D, and Georgiadis MM. (2004) “The Crystal Structure of the Monomeric Reverse Transcriptase from Moloney Murine Leukemia Virus.” *Structure* 12: 819-829.
40. Delong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard N, Martinez A, Sullivan MB, Edwards R, Brito BR, Chisholm SW, Karl DM. (2006) “Community Genomics among stratified microbial assemblages in the ocean’s interior.” *Science* 311: 496-503.
41. Ehrenberg, M. (2008) “The green fluorescent protein: discovery, expression and development.”
42. Faulkner, DJ. (2000) “Marine pharmacology.” *Antonie van Leeuwenhoek* 77: 135-145.

43. Field SF, Bulina MY, Kelmanson IV, Bielawski JP, and Matz MV. (2006) "Adaptive Evolution of Multicolored Fluorescent Proteins in Reef-Building Corals." *Journal of Molecular Evolution*: 332-339.
44. Fradkov AF, Chen Y, Ding L, Barsova EV, Matz MV, and Lukyanov SA. (2000) "Novel fluorescent protein from *Discosoma* coral and its mutants possesses a unique far-red fluorescence." *FEBS Letters* 479: 127-130.
45. Fradkov AF, Verkhusha VV, Staroverov DB, Bulina ME, Yanushevich YG, Martynov VI, Lukyanov S, and Lukyanov KA. (2002) "Far-red fluorescent tag for protein labelling." *Biochemical Journal* 398: 17-21.
46. Furman JL, Badran AH, Shen S, Stains CI, Hannallah J, Segal DJ, and Ghosh I. (2009) "Systematic evaluation of split-fluorescent proteins for the direct detection of native and methylated DNA." *Bioorg Med Chem Lett* 19, no. 14: 3748-3751.
47. Garrett G and Grisham G. (2005) *Biochemistry*, 5TH ED.
48. Georgiadis MM, Jessen SM, Ogata CM, Telesnitsky A, Goff SP, and Hendrickson WA. (1995) "Mechanistic implications from the structure of a catalytic fragment of Moloney murine leukemia virus reverse transcriptase." *Structure*, no. 3: 879-892.
49. Gilbert JA, Field D, Huang Y, Edwards R, Li W, Gilna P, and Joint I. (2008) "Detection of Large Numbers of Novel Sequences in the Metatranscriptomes of Complex Marine Microbial Communities." *PLoS ONE* 3, no. 8.
50. Gold V, and Bethell D. (1977) *Advances in physical organic chemistry*. 15th ED.
51. Grant S, Grant WD, Cowan DA, Jones BE, Ma Y, Ventosa A, Heaphy S. (2006) "Identification of Eukaryotic Open Reading Frames in Metagenomic cDNA Libraries Made from Environmental Samples." *Appl. Environ. Microbiol* 72, no. 1: 135-143.
52. Gruber DF, Kao H, Janoschka S, Tsai J, and Pieribone VA. (2008) "Patterns of Fluorescent Protein Expression in Scleractinian Corals." *Biol. Bull.* 215, no. October: 143-154.
53. Gu W, and Wang Y. (2011) *Gene discovery for disease models*, 1ST ED.

54. Gurskaya NG, Fradkov AF, Terskikh A, Matz MV, Labas YA, Martynov VI, Yanushevich YG, Lukyanov KA, and Lukyanov SA. (2001) "GFP-like chromoproteins as a source of far-red fluorescent proteins." *FEBS* 507: 16-20.
55. Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK, Desantis TZ, Sodergren E, Methe B, Petrosino JF, Knight R, Birren BW. (2011) "Chimeric 16S rRNA sequence formation and detection in Sanger Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons." *Genome Research*, no. 21: 494-504.
56. Haddock SHD, Mastroianni N, and Christianson LM. (2009) "A photoactivatable green-fluorescent protein from the *Phylum Ctenophora*." *Proceedings of the royal society. B*: 1-6.
57. Haine, ER. (2008) "Symbiont-mediated protection." *Proceedings of the royal society* 275: 353-361.
58. Hay ME, and Fenical W. (1996) "Chemical ecology and marine biodiversity: Insights and products from the sea." *Oceanography* 9: 10-20.
59. Head IM, Saunders JR, and Pickup RW. (1998) "Microbial Evolution, Diversity, and Ecology: A Decade of Ribosomal RNA Analysis of Uncultivated Microorganisms." *Microbial ecology* 35, no. 1: 1-21.
60. Heestand, NE. (2009) "Phylogeny and Evolution of *Anthopleura* (*Cnidaria*: *Anthozoa*: *Actinaria*)".
61. Heim RA, Cubitt B, and Tsien RY. (1995) "Improved green fluorescence." *Nature* 373: 663-664.
62. Heim R, and Tsien RY. (1996) "Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer." *Current Biology* 6: 178-182.
63. Heim R, Prasher DC, and Tsien RY. (1994) "Wavelength mutations and posttranslational autoxidation of green fluorescent protein." *Proc. Natl. Acad. Sci* 91, no. December: 12501-12504.

64. Henderson NJ, and Remington SJ. (2005) “Crystal structures and mutational analysis of amFP486, a cyan fluorescent protein from *Anemonia majano*.” *PNAS* 102, no. 36: 12712-12717.
65. Henderson NJ, and Remington SJ. “The Kindling Fluorescent Protein: A Transient Photoswitchable Marker.” *Physiology* 21 (2006): 162-170.
66. Hill M, Allenby A, Ramsby B, Schönberg C, and Hill A. (2011) “Molecular Phylogenetics and Evolution *Symbiodinium* diversity among host clionaid sponges from Caribbean and Pacific reefs: Evidence of heteroplasmy and putative host-specific symbiont lineages.” *Molecular Phylogenetics and Evolution* 59, no. 1: 81-88.
67. Hu Q, Li L, Wang Y, Zhao W, Qi H, and Zhuang G. (2010) “Construction of WCB-11 : A novel phiYFP arsenic-resistant whole-cell biosensor.” *Journal of Environmental Sciences* 22, no. 9: 1469-1474.
68. Inouye S, Ogawa H, Yasuda K, Umesono K, and Tsuji FI. (1997) “A bacterial cloning vector using a mutated *Aequorea* green fluorescent protein as an indicator.” *Gene* 189: 159-162.
69. Inouye S, and Tsuji FI. (1994) “*Aequorea* green fluorescent protein Expression of the gene and fluorescence characteristics of the recombinant protein.” *FEBS Letters* 341: 277-280.
70. Ip D, Wong K, and Wan DC. (2007) “Characterization of Novel Orange Fluorescent Protein Cloned from Cnidarian Tube Anemone *Cerianthus* sp.” *Marine Biotechnology* 9: 469-478.
71. Ivanova EP, and Mikhailov VV. (2001) “A New Family , *Alteromonadaceae* fam . nov ., Including Marine *Proteobacteria* of the Genera *Alteromonas* , *Pseudoalteromonas* , *Idiomarina* , and *Colwellia*.” *Microbiology* 70, no. 1: 10-17.
72. Izuno MM, Asukawa KY, and Nouye KI. (2010) “Insight into the Mechanism of the Stabilization of Moloney Murine Leukaemia Virus Reverse Transcriptase by Eliminating RNase H Activity.” *Biosci. Biotechnol. Biochem* 74, no. 2: 440-442.
73. Johansen SD, Karlsen O, Okkenhaug S, Hansen H, Moum T, Coucheron DH, and Seternes OM. (2010) “Approaching marine bioprospecting in hexacorals by RNA deep sequencing.” *New Biotechnology* 27, no. 3: 267-275.

74. Karasawa S, Araki T, Nagai T, Mizuno H, and Miyawaki A. (2004) "Cyan-emitting and orange-emitting fluorescent proteins as a donor/acceptor pair for fluorescence resonance energy transfer." *Biochemical Journal* 381, no. 1: 307-312.
75. Karasawa S, Araki T, Yamamoto-Hino M, and Miyawaki A. (2003) "A green-emitting fluorescent protein from *Galaxeidae* coral and its monomeric version for use in fluorescent labeling." *The Journal of biological chemistry* 278, no. 36: 34167-71.
76. Kornbluth S, Paulson KE, and Hanafusa H. (1988) "Novel tyrosine kinase identified by phosphotyrosine antibody screening of cDNA Novel Tyrosine Kinase Identified by Phosphotyrosine Antibody Screening of cDNA Libraries." *Mol. Cell. Biol.* 8, no. 12: 5441-5544.
77. Krishna M, and Ingole B. (2009) "Evolving trends in biosciences: multi-purpose proteins – GFP and GFP-like proteins." *Current Science* 97, no. 7: 1022-1032.
78. Labas YA, Gurskaya NG, Yanushevich YG, Fradkov AF, Lukyanov KA, Lukyanov SA, and Matz MV. (2002) "Diversity and evolution of the green fluorescent." *PNAS* 99, no. 7: 4256-4261.
79. Leutenegger A, D'Angelo C, Matz MV, Denzel A, Oswald F, Salih A, Nienhaus GU, and Wiedenmann J. (2007) "It's cheap to be colorful Anthozoans show a slow turnover of GFP-like proteins." *FEBS Journal* 274: 2496-2505.
80. Li Chuan, and Ronald M Evans. "Ligation independent cloning irrespective of restriction site compatibility." *Nucleic acids research* 25, no. 20 (1997): 4165-4166.
81. Loos DC, Habuchi S, Flors C, Hotta J, Nienhaus GU, and Hofkens J. (2006) "Photoconversion in the Red Fluorescent Protein from the Sea Anemone *Entacmaea quadricolor*: Is *Cis-Trans* Isomerization Involved?" *J. Phys. Chem. B.* 128: 6270-6271.
82. Lope JV, Ranzer LK, Ledger A, Schoch B, Duckworth A, Mccarthy PJ, and Kerr RG. (2008) "Comparison of Bacterial Diversity within the Coral Reef Sponge, *Axinella corrugata*, and the Encrusting Coral *Erythropodium caribaeorum*." In *Proceedings of the 11th Session number 26 International Coral Reef Symposium*, 7-11.

83. Lukyanov KA, Fradkov AF, Gurskaya NG, Matz MV, Labas YA, Savitsky AP, Markelov ML, Zaraisky AG, Zhao X, Fang Y, Tan W, Lukyanov SA. (2000) "Natural Animal Coloration Fluorescent Protein Homolog." *Journal of Bacteriology* 275, no. 34: 25879-25882.
84. March JC, Roa G, and Bentley WE. (2003) "Biotechnological applications of green fluorescent protein." *Applied Microbiology* 62: 303-315.
85. Martynov VI, Savitsky AP, Martynova NY, Savitsky PA, Lukyanov KA, and Lukyanov SA. (2001) "Alternative Cyclization in GFP-like Proteins Family." *Journal of Biological chemistry* 276, no. 24: 21012-21016.
86. Masuda H, Takenaka Y, Yamaguchi A, Nishikawa S, and Mizuno H. (2006) "A novel yellowish-green fluorescent protein from the marine copepod, *Chiridius poppei*, and its use as a reporter protein in HeLa cells." *Gene* 372: 18 - 25.
87. Matz MV, Fradkov AF, Labas YA, Savitsky AP, Zaraisky AG, Markelov ML, and Lukyanov SA. (1999) "Fluorescent proteins from nonbioluminescent *Anthozoa* species." *Nat. Biotechnol.* 17: 969-973.
88. Mishin AS, Subach FV, Yampolsky IV, King W, Lukyanov A, and Verkhusha VV. (2008) "The first mutant of the *Aequorea victoria* green fluorescent protein that forms a red chromophore." *Biochemistry* 47, no. 16: 4666-4673.
89. Miyawaki, A. (2002) "Green Fluorescent Protein-like Proteins in Reef *Anthozoa* Animals." *Cell Structure and Function* 27: 343-347.
90. Miyawaki A, Nagai T, and Mizuno H. (2003) "Mechanisms of protein fluorophore formation and engineering." *Current Opinion in Chemical Biology* 7: 557-562.
91. Munroe DJ, Loebbertt R, Bric E, Whirron T, Prawirrt D, Vu D, Buckler A, Winterpachtt A, Zabelt B, and Housman DE. (1995) "Systematic screening of an arrayed cDNA library by PCR." *Proc. Natl. Acad. Sci* 92: 2209-2213.
92. Murray AE, and Grzymski JJ. (2007) "Diversity and genomics of Antarctic marine micro-organisms." *Philosophical transactions of the royal society* 362: 2259-2271.

93. Müller-taubenberger A, and Anderson KI. (2007) “Recent advances using green and red fluorescent protein variants.” *Applied Microbiology* 77: 1-12.

94. Nagai T, Ibata K, Park ES, Kubota M, Mikoshiba K, and Miyawaki A. (2002) “A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications.” *Nat. Biotechnol.* 20: 87-90.

95. Nienhaus GU, Nienhaus K, Holzle A, Ivanchenko S, Renzi F, Oswald F, Wolff M, Rocker C, Vallone B, Weidemann W, Heilker R, Nar H, Wiedenmann J. (2006) “Symposium-in-Print : Green Fluorescent Protein and Homologs Photoconvertible Fluorescent Protein EosFP : Biophysical Properties and Cell Biology Applications.” *Photochemistry and Photobiology* 82: 351-358.

96. Nienhaus K, Nar H, Heilker R, and Nienhaus GU. (2008) “*Trans-Cis* Isomerization is Responsible for the Red-Shifted Fluorescence in Variants of the Red Fluorescent Protein eqFP611.” *J. Am. Chem. Soc.* 130: 12578-12579.

97. Niwa H, Inouye S, Hirano T, Matsuno T, Kojima S, Kubota M, Ohashi M, and Tsuji FI. (1996) “Chemical nature of the light emitter of the *Aequorea* green fluorescent protein.” *PNAS* 93, no. 24: 13617-22.

98. Olenych SG, Claxton NS, Ottenberg GK, and Davidson MW. (2005) *The Fluorescent Protein Color Palette. Advances.*

99. Onishi M, Shibuya A, Miyajima A, and Nolani GP. (1995) “Efficient screening of retroviral cDNA expression libraries.” *Biochemistry* 92: 9146-9150.

100. Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY, and Remington SJ. (1996) “Crystal structure of the *Aequorea victoria* green fluorescent protein.” *Science* 273: 1392–1395.

101. Palmer CV, Modi CK, and Mydlarz LD. (2009) “Coral Fluorescent Proteins as Antioxidants.” *PLoS ONE* 4, no. 10.

102. Patterson GH, Knobel SM, Sharif WD, Kain SR, and Piston DW. "Use of the Green Fluorescent Protein and Its Mutants in Quantitative Fluorescence Microscopy." *Biophysical Journal* 73 (1997): 2782-2790.

103. Patterson GH, and Lippincott-Schwartz J. (2002) "A photoactivatable GFP for selective photolabeling of proteins and cells." *Science* 297, no. 5588: 1873-7.

104. Pawlik, JR. (1993) "Marine invertebrate chemical defenses." *Chem. Rev.* 93: 1911-1922.

105. Pearson CVM, Rogers AD, and Sheader M. (2002) "The genetic structure of the rare lagoonal sea anemone , *Nematostella vectensis* Stephenson (*Cnidaria* ; *Anthozoa*) in the United Kingdom based on RAPD analysis." *Molecular Ecology* 11: 2285-2293.

106. Piatkevich KD, and Verkhusha VV. (2009) "Advances in engineering of fluorescent proteins and photoactivatable proteins with red emission." *Current Opinion in Chemical Biology* 14, : 1-7.

107. Pletneva N, Pletnev S, Tikhonova T, Popov V, Martynov VI, and Pletneva V. (2006) "Structure of a red fluorescent protein from *Zoanthus*, zRFP574, reveals a novel chromophore." *Biological Crystallography* 538: 527-532.

108. Pletneva N, Tikhonova T, Martynov VI, Pletneva V, Pakhomov A, Wlodawer A, Dauter Z, and Pletnev S. (2007) "Refined crystal structures of red and green fluorescent proteins from the button polyp *Zoanthus*." *Biological Crystallography* 63: 1082-1093.

109. Pollok BA, and Heim R. (1999) "Using GFP in FRET-based applications." *Cell Biology* 9: 57-60.

110. Pomponi, SA. (1998) "The potential for the marine biotechnology industry." *Journal of Organic Chemistry*: 101-104.

111. Poponov SY, Mudrik NN, Lukyanov S, and Lukyanov KA. (2005) "Far-red fluorescent proteins evolved from a blue chromoprotein from *Actinia equina*." *Biochemical Journal* 392: 649-654.

112. Prasher DC, Eckenrode VK, Ward WW, Prendergast FG, and Cormier MJ. (1992) "Primary structure of the *Aequorea victoria* green-fluorescent protein." *Gene* 111: 229-233.

113. Proksch P, Edrada RA, and Ebel R. (2002) “Drugs from the seas – current status and microbiological implications.” *Appl Microbiol Biotechnol* 59: 125-134.

114. Reid BG, and Flynn GC. (1997) “Chromophore formation in green fluorescent protein.” *Biochemistry* 36, no. 22: 6786-91.

115. Rohwer F, Seguritan V, Azam F, and Knowlton N. “Diversity and distribution of coral-associated bacteria.” *Marine Ecology Progress Series* 243 (2002): 1-10.

116. Rosa S, Milone A, Kujumgiev A, Stefanov K, Nechev I, and Popov S. (2000) “Metabolites from a marine bacterium *Pseudomonas* / *Alteromonas* , associated with the sponge *Dysidea fragilis*.” *Comparative Biochemistry and Physiology* 126: 391 - 396.

117. Russell NJ, and Nichols DS. (1999) “Polyunsaturated fatty acids in marine bacteria – a dogma rewritten.” *Microbiology* 145: 767-779.

118. Salih A, Larkum A, Cox G, Kühl M, and Hoegh-guldberg O. (2000) “Fluorescent pigments in corals are photoprotective.” *Nature* 48: 850-853.

119. Schuett C, and Doepke H. (2010) “Endobiotic bacteria and their pathogenic potential in Cnidarian tentacles.” *Helgol Mar Res* 64: 205-212.

120. Schuett C, Doepke H, Groepler W, and Wichels A. (2005) “Diversity of intratunical bacteria in the tunic matrix of the colonial ascidian *Diplosoma migrans*.” *Helgol Mar Res* 59: 136-140.

121. Shagin DA, Barsova EV, Yanushevich YG, Fradkov AF, Lukyanov KA, Labas YA, Semenova TN, Ugalde JA, Meyers A, Nunez JM, Widder EA, LukyanovSA, Matz MV. (2004) “GFP-like Proteins as Ubiquitous Metazoan Superfamily : Evolution of Functional Features and Structural Complexity.” *Molecular Biology and Evolution* 21, no. 5: 841-850.

122. Shaner NC, Patterson GH, and Davidson MW. (2007) “Advances in fluorescent protein technology.” *Journal of Cell Science* 120: 4247-4260.

123. Shkrob MA, Mishin AS, Chudakov DM, Labas YA, and Lukyanov KA. (2008) “Chromoproteins of the Green Fluorescent Protein Family: Properties and applications.” *Russian journal of Bioorganic Chemistry* 34, no. 5: 517-525.

124. Song M, Li Y, and Kiledjian M. (2010) “Multiple mRNA Decapping Enzymes in Mammalian Cells.” *Molecular Cell*, no. 40: 423-432.

125. Stakenborg T, Vicca J, Butaye P, Maes D, De Baere T, Verhelst R, Peeters J, de Kruif A, Haesebrouck F, and Vaneechoutte M. (2005) “Evaluation of amplified rDNA restriction analysis (ARDRA) for the identification of *Mycoplasma* species.” *BMC infectious diseases* 5: 46.

126. Stangegaard M, Høgh Dufva I, and Dufva M. (2006) “Reverse transcription using random pentadecamer primers increases yield and quality of resulting cDNA.” *Biotechniques* 40, no. 5: 649-656.

127. Stepanenko OV, Verkhusha VV, Kuznetsova IM, Uvershy VM, and Turoverov KK. (2008) “Fluorescent proteins as biomarkers and biosensors: Throwing color lights on molecular and cellular processes.” *Curr. Protein Pept Sci.* 9, no. 4: 338-369.

128. Stiel AC, Andresen M, Bock H, Hilbert M, Schilde J, Scho A, Eggeling C, Egner A, Hell SW, and Jakobs S. (2008) “Generation of Monomeric Reversibly Switchable Red Fluorescent Proteins for Far-Field Fluorescence Nanoscopy.” *Biophysical Journal* 95, no. September: 2989-2997.

129. Stiel, AC, Trowitzsch S, Weber G, Andresen M, Eggeling C, Hell SW, Jakobs S, and Wahl MC. (2007) “1.8A˚ bright-state structure of the reversibly switchable fluorescent protein Dronpa guides the generation of fast switching variants.” *Biochem J.* 42: 35-42.

130. Subach OM, Gundorov IS, Yoshimura M, Subach FV, Zhang J, Gruenwald D, Souslova E, Chudakov DM, and Verkhusha VV. (2008) “Conversion of Red Fluorescent Protein into a Bright Blue Probe.” *Cell Press Chemistry & Biology* 15: 1116-1124.

131. Tanaka H, Ohshima N, and Hidaka H. (1999) “Isolation of cDNAs Encoding Cellular Drug-Binding Proteins Using a Novel Expression Cloning Procedure: Drug-Western.” *Molecular Pharmacology* 55: 356 -363.

132. Tringe SG, Von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, Bork P, Hugenholtz P, Rubin EM. (2005) "Communities Comparative Metagenomics of Microbial Communities." *Science* 308: 554-557.

133. Tsien, RY. (1998) "The Green fluorescent protein." *Annu. Rev. Biochem* 67: 509-544.

134. Turcic K, Pettikiriarachchi A, Battad J, Wilmann PG, Rossjohn J, Dove SG, Devenish RJ, and Prescott M. (2006) "Amino acid substitutions around the chromophore of the chromoprotein Rtms5 influence polypeptide cleavage." *Biochemical and Biophysical Research Communications* 340: 1139-1143.

135. Uchiyama T, Abe T, Ikemura T, and Watanabe K. (2005) "Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes." *Nature Biotechnology* 23, no. 1: 88-93.

136. Verkhusha VV, and Lukyanov KA. (2004) "The molecular properties and applications of *Anthozoa* fluorescent proteins and chromoproteins." *Nature Biotechnology* 22, no. 3: 289-296.

137. Verlag F, Ivanova EP, Sawabe T, Zhukova NV, Gorshkova NM, Nedashkovskaya OI, Hayashi K, Frolova GM, Sergeev AF, Pavel KG, Mikhailov VV, Nicolau DV. (2003) "Occurrence and Diversity of Mesophilic *Shewanella* Strains Isolated from the North-West Pacific Ocean." *System. Appl. Microbiol.* 26: 293-301.

138. Vincentelli R, Cimino A, Geerlof A, Kubo A, Satou Y, and Cambillau C. (2011) "High-throughput protein expression screening and purification in *Escherichia coli*." *Methods* 55, no. 1: 65-72.

139. Vogt A, Angelo CD, Oswald F, Denzel A, Mazel CH, and Matz MV. (2008) "A Green Fluorescent Protein with Photoswitchable Emission from the Deep Sea." *PLoS ONE* 3, no. 11: 45-49.

140. Wachter RM, Elsliger MA, Kallio K, Hanson GT, and Remington SJ. (1998) "Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein." *Structure* 6: 1267-1277.

141. Wachter, RM. (2006) “Symposium-in-Print: Green Fluorescent Protein and Homologs The Family of GFP-Like Proteins: Structure, Function, Photophysics and Biosensor Applications. Introduction and Perspective.” *Photochemistry and Photobiology* 82: 339-344.

142. Wang F, Wang J, Jian H, Zhang B, Li S, Wang F, Zeng X, Zeng X, Gao L, Bartlett DH, Yu J, Hu S, Xiao X. (2008) “Environmental Adaptation : Genomic Analysis of the Piezotolerant and Psychrotolerant Deep-Sea Iron Reducing Bacterium *Shewanella piezotolerans* WP3.” *Genome* 3, no. 4: 1-12.

143. Wang J, Chou Y, Chou J, Chen CA, and Chen W. (2008) “*Tenacibaculum aiptasiae* sp. nov., isolated from a sea anemone *Aiptasia pulchella*.” *International Journal of Systematic and Evolutionary Microbiology* 58: 761-766.

144. Werren JH, and Neill SLO. (1879) *The evolution of heritable symbionts*,.

145. Wiedenmann J, Ivanchenko S, Oswald F, and Nienhaus GU. (2004) “Identification of GFP-like Proteins in Nonbioluminescent, *Azooxanthellate Anthozoa* Opens New Perspectives for Bioprospecting.” *Marine Biotechnology* 6: 270-277.

146. Wiedenmann J, Ivanchenko S, Oswald F, Schmitt F, Röcker C, Salih A, Spindler K, and Nienhaus GU. (2004) “EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion.” *Proceedings of the National Academy of Sciences of the United States of America* 101, no. 45: 15905-10.

147. Wiedenmann J, Schenk A, Röcker C, Girod A, Spindler K, and Nienhaus GU. (2002) “A far-red fluorescent protein with fast maturation and reduced oligomerization tendency from *Entacmaea quadricolor* (*Anthozoa, Actinaria*).” *PNAS* 99, no. 18: 11646-51.

148. Wilmann PG, Battad J, Petersen J, Wilce MCJ, Dove S, Devenish RJ, Prescott M, and Rossjohn J. (2006) “The 2.1A Crystal Structure of copGFP, a Representative Member of the Copepod Clade Within the Green Fluorescent Protein Superfamily.” *J. Mol. Biol* 359: 890-900.

149. Wilman PG, Petersen J, Pettikiriarachchi A, Buckle AM, Smith SC, Olsen S, Perugini MA, Devenish RJ, Prescott M, and Rossjohn J. (2005) “The 2.1A ˚ Crystal Structure of the Far-red

Fluorescent Protein HcRed: Inherent Conformational Flexibility of the Chromophore." *J. Mol. Biol* 349: 223-237.

150. Won JH, Rho BJ, and Song JI. (2001) "A phylogenetic study of the *Anthozoa* (*Phylum Cnidaria*) based on morphological and molecular characters." *Coral Reefs*, no. 20: 39-50.

151. Wood V, William RG, Rajandream M, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayles J, Baker S, Basham D, Bowman S, Brooks K, Brown D, Brown S, Chillingworth T, Churcher C, Collins M, Connor R, Cronin A, Davis P, Feltwell T, Fraser A, Gentles S, Goble A, Hamlin N, Harris D, Hidalgo J, Hodgson G, Holroyd S, Hornsby T, Howarth S, Huckle EJ, Hunt S, Jagels K, James K, Jones L, Jones M, Leather S, Mcdonald S, Mclean J, Mooney P, Moule S, Mungall K, Murphy L, Niblett D, Odell C, Oliver K, Neil SO, Pearson D, Quail MA, Rabbinowitsch E, Rutherford K, Rutter S, Saunders D, Seeger K, Sharp S, Skelton J, Simmonds M, Squares R, Squares S, Stevens K, Taylor K, Taylor RG, Tivey A, Walsh S, Warren T, Whitehead S, Woodward J, Volckaert G, Aert R, Robben J, Grymonprez B, Weltjens I. (2002) "The genome sequence of *Schizosaccharomyces pombe*." *Nature* 415: 871-881.

152. Yang TT, Sinai P, Green G, Kitts PA, Chen YT, and Lybarger L. (1998) "Improved fluorescence and dual color detection with enhanced blue and green variants of the green fluorescent protein." *Journal of Biological chemistry* 273: 8212-8216.

153. Yanushevich YG, Gurskaya NG, Staroverov DB, Lukyanov SA, and Lukyanov KA. (2003) "A Natural Fluorescent Protein That Changes Its Fluorescence Color during Maturation." *Russian journal of Bioorganic Chemistry* 29, no. 4: 325-329.

154. Yanushevich YG, Shagin DA, Fradkov F, Shakhbazov KS, Barsova EV, Gurskaya NG, Labas YA, Matz MV, Lukyanov KA, and Lukyanov SA. (2005) "Spectral diversity among members of the green fluorescent protein family in hydroid jellyfish (*Cnidaria, Hydrozoa*)." *Russian Journal of Bioorganic Chemistry* 31, no. 1: 43-47.

155. Yun J, and Ryu S. (2005) "Screening for novel enzymes from metagenome and SIGEX, as a way to improve it." *Microbial Cell Factories* 5: 1-5.

156. Zhu YY, Machleder EM, Chenchik A, Li R, and Siebert PD. (2001) “Reverse Transcriptase Template Switching: A SMART™ Approach for Full-Length cDNA Library Construction.” *Biotechniques* 30, no. 4: 892-897.

157. Zongjun D, Wany Z, Hongjie X, Guoqiang L, and Guanjun C. (2010) “Isolation and diversity analysis of heterotrophic bacteria associated with sea anemones.” *Acta Oceanol. Sin* 29, no. 2: 62-69.

Appendix A (general recipes)

Media

All media components were purchased from Sigma-Aldrich and Merck unless otherwise stated. Luria-Bertani (LB) agar was prepared using 1 % (w/v) tryptone, 0.5 % (w/v) yeast extract, 1 % (w/v) NaCl and 1.5 % (w/v) bacteriological agar. SOB medium (broth) was prepared from 2 % (w/v) tryptone, 0.05 % (w/v) NaCl, 0.5 % (w/v) yeast extract and 0.02 % (w/v) KCl. SOC medium (broth) was prepared by aseptically adding 2 M $MgCl_2$ to 0.5 % (w/v) and 1 M (filter sterilised) glucose to 2 % (w/v) to previously autoclaved SOB. Psi (ψ) broth was prepared containing 2 % (w/v) tryptone, 0.5 % (w/v) yeast extract, 0.5 % (w/v) $MgSO_4$ and the pH was set to 7.6 with KOH. $MgCl_2$ and glucose

The antibiotics kanamycin (kan) and ampicillin (amp) were added as required at final concentrations of 50 µg/ml and 100 µg/ml, respectively, unless otherwise stated. For blue white selection 80 µg/µL X-gal and 1 mM IPTG were added to LB agar.

Extraction and Lysis buffers

Solution D [4M Guanidine isothiocyanate, 25mM Sodium Citrate, 0.5 % Sarcosyl, 0.1M β-Mercaptoethanol]

Preparation of solution D:

25g of Guanidium isothiocyanate was dissolved in 29.3mL of mQ water and double autoclaved.

To the 4M Guanidium isothiocyanate solution 1.76mL of 75mM sodium citrate and 2.64mL of 10 % sarcosyl were added. This stock solution can be stored at room temperature for at least 3 months.

To prepare solution D 360µL β-mercaptoethanol/50mL of stock solution was added. Solution D can be kept for 1 month at room temperature. It is very important to use fresh solution D for extractions to achieve the best results.

To prepare the lysis buffer for gDNA extraction stock solutions 1 M Tris –HCL, 20% SDS, and 20 mg/mL Proteinase K were used to prepare 0.01M Tris-HCL, 0.1 % SDS and 300µg of Proteinase K lysis buffer fresh each time.

MOPS/formaldehyde gels 10 X MOPS running buffer containing 0.4 M MOPS, 0.1 M NaAc, 0.01 M EDTA pH 7.0 was prepared from stock solutions of filter sterilized 2 M MOPS, 3 M NaAC and 0.5 M EDTA.

Appendix B (strains, plasmid/vectors and Primer sequences)

Table 5 Strains, plasmid/vector genotypes and primers sequences used in this study.

Strains		
Name	Genotype	Reference
GeneHogs	F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G *fhu*A::IS2 (confers phage T1 resistance)	
BL21 (DE3) pLysS	F-, ompT, hsdSB(rB-,mB-), gal, dcm, (DE3), pLysS, CamR	

Vectors and Plasmids		
Name	Properties	Reference
pUC19	Ampicillin resistance marker; LacZα gene; colE1 origin; high copy number	
pSMART2IF	(identical to pUC19) contains In-Fusion linker molecules	
pEGFP-N3	Human cytomegalovirus (CMV) immediate early promoter; Enhanced green fluorescent protein gene; SV40 early mRNA polyadenylation signal; f1 single-strand DNA origin; Bacterial promoter for expression of Kanamycin resistance gene; SV40 origin of replication; SV40 early promoter; Kanamycin/neomycin resistance gene; Herpes simplex virus (HSV) thymidine kinase (TK) polyadenylation signal; colE1 plasmid replication origin, 500 copy number	

pET21a	T7 promoter and terminator; T7 and His•Tag coding sequences; *lacI* coding sequence; pBR322 origin; *bla* coding sequence; f1 origin	

Primers

	Name	Sequence	Reference
18S rRNA gene	EukA	5’-AACCTGGTTGATCCTGCCAGT-3’	Diez, et al. 2001
	EukB	5’-TGATCCTTCTGCAGGTTCACCTAC-3’	Diez, et al. 2001
cDNA synthesis oligonuclotides	SMARTer II A Oligonucleotide	5'-AAGCAGTGGTATCAACGCAGAGTACXXXXX-3'	Clontech
	3' SMART CDS Primer II A	5’-AAGCAGTGGTATCAACGCAGAGTACT(30)N-1N-3’	Clontech
16S rRNA gene	F1	5’-AGAGTTTGATCITGGCTCAG-3’	Modified fD1 Weisburg et al. 1991
	R5	5’-ACGGITACCTTGTTACGACTT-3’	Modified rP2 Weisburg et al. 1991
General vector primers	M13 For	5´-CGCCAGGGTTTTCCCAGTCACGAC-3´	Promega
	M13 Rev	5´-TCACACAGGAAACAGCTATGAC-3´	Promega
	pJET1.2 forward	5’-CGACTCACTATAGGGAGAGCGGC-3’	Fermentas Inqaba
	pJET1.2 reverse	5’-AAGAACATCGATTTTCCATGGCAG-3	Fermentas Inqaba

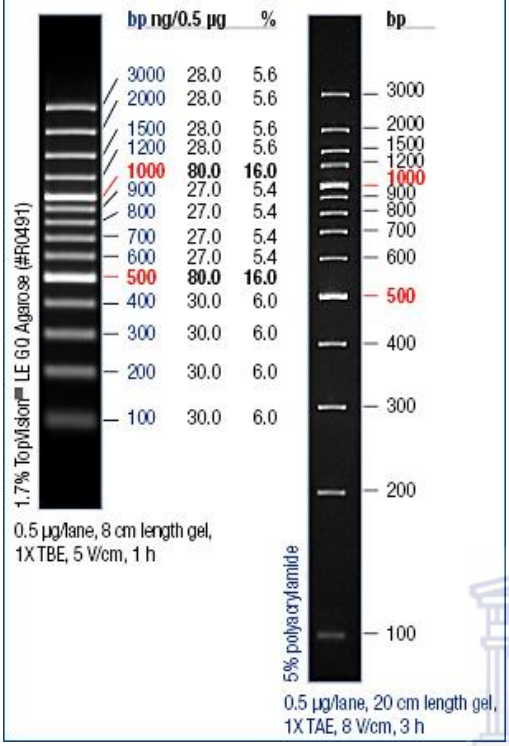
Degenerate chromoprotein primers	dCPfwd	5’-GGHGGWCCWYTGCCATTTGCYT-3’	Tanya Nyman
	dCPrv	5’-GCATMTGRGGKCCATYRGWAGG-3’	
Degenerate fluorescent protein primers	AfpF2	5’-CCAYTKCCWTTYKCYKYTGA-3’	Tanya Nyman
	AfpF3	5’-CCAYTKCCWTTYKCYKYCGA-3’	
	AfpF6	5’-CCTYTKCCWTTYKCYKYCGA-3’	
	AfpR	5’- GGWCCRTYDGSRGGAAARTT-3’	
Functional cDNA synthesis primers	FUNF	5’- AGTTCGCCAACGCCACCATGGTAGGAGGCGCAGAGTAC-3’	Tanya Nyman
	FUNR	5’- AAGTGCGGCCGC_T(30)_N-1N -3’	

Appendix C –DNA Markers and Ladders

A

GeneRuler™ 100 bp Plus DNA Ladder

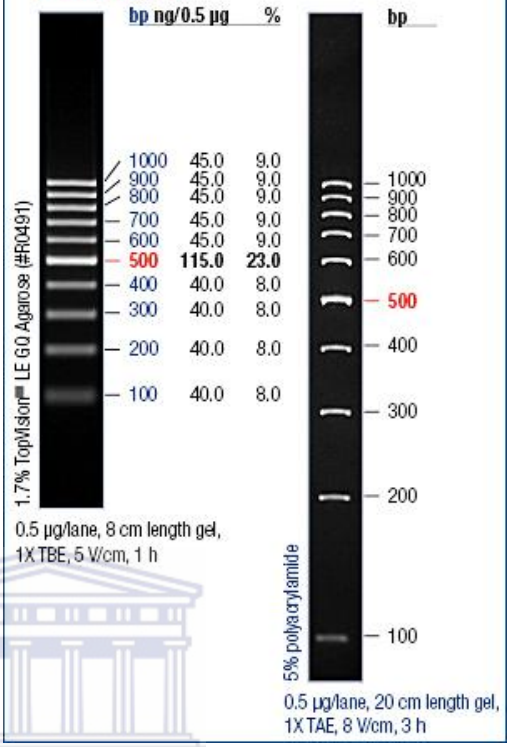
O'GeneRuler™ 100 bp Plus DNA Ladder, ready-to-use



B

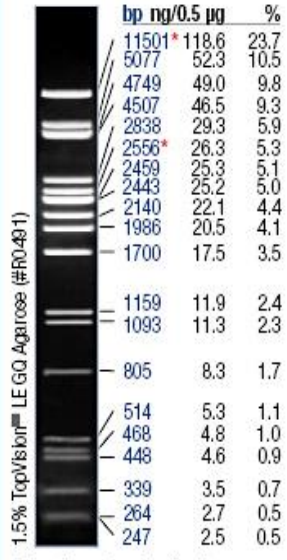
GeneRuler™ 100 bp DNA Ladder

O'GeneRuler™ 100 bp DNA Ladder, ready-to-use



C

Lambda DNA/PstI Marker, 24

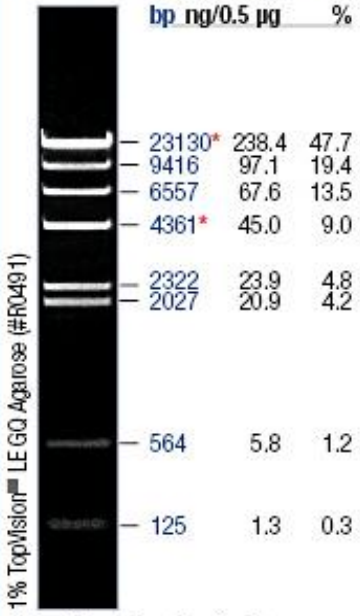


Range

29 fragments (in bp): 11501*, 5077, 4749, 4507, 2838, 2556*, 2459, 2443, 2140, 1986, 1700, 1159, 1093, 805, 514, 468, 448, 339, 264, 247, *216, 211, 200, 164, 150, 94, 87, 72, 15 (oblique fragments comprise 2.3%).*

D

Lambda DNA/HindIII Marker, 2



Range

8 fragments (in bp): 23130*, 9416, 6557, 4361*, 2322, 2027, 564, 125.

E

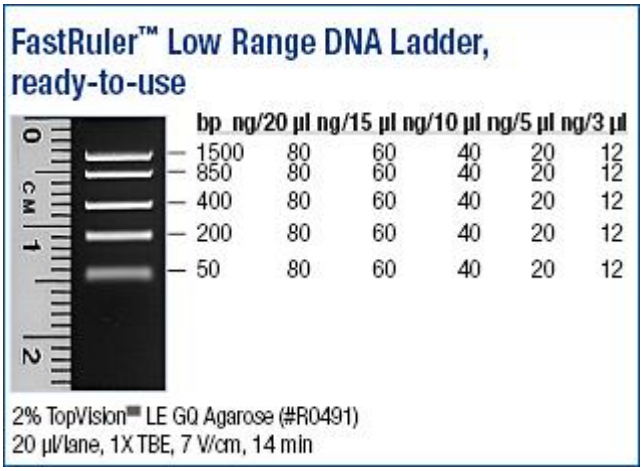
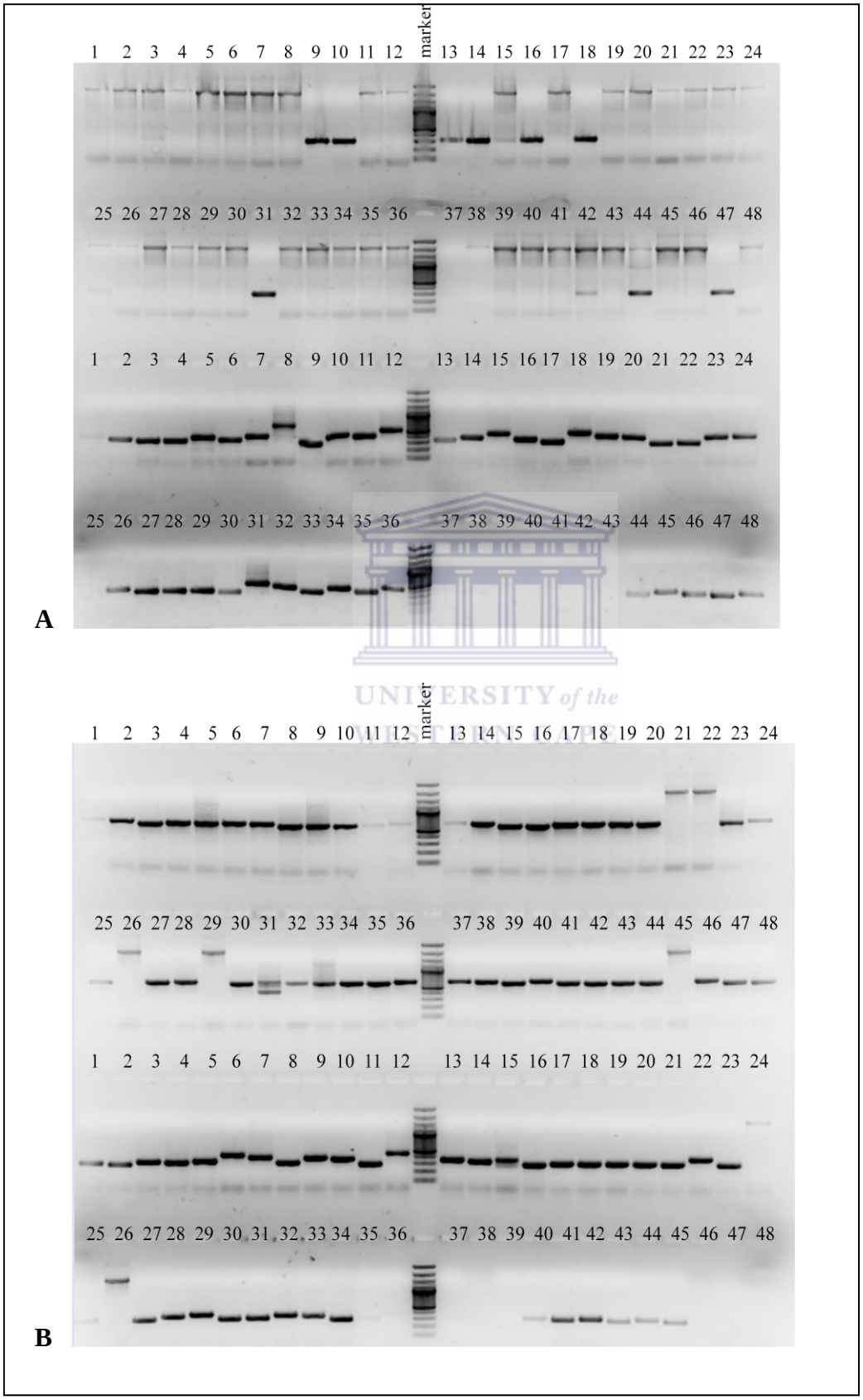


Figure 44. Commercially available DNA ladders and molecular weight markers used in this thesis.

Appendix D (additional Figures)

M13 PCR of rRNA clone libraries



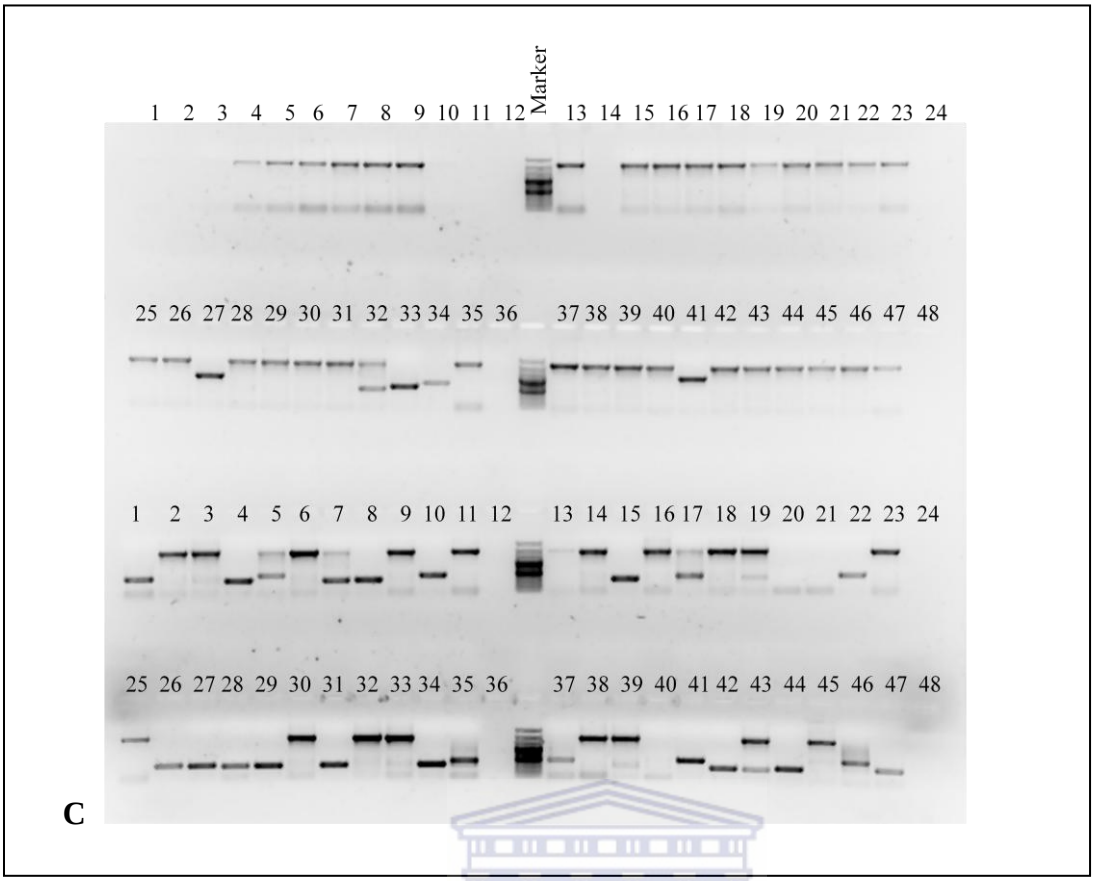


Figure 45. M13 colony PCR from 18S and 16S rRNA diversity libraries of A) *P. varia* B) *P. flagellifera* C) *B. capensis*. The top 48 (unblocked) are the 18S rRNA amplicons and the bottom (blocked) 48 are the 16S rRNA amplicons.

RNA checks

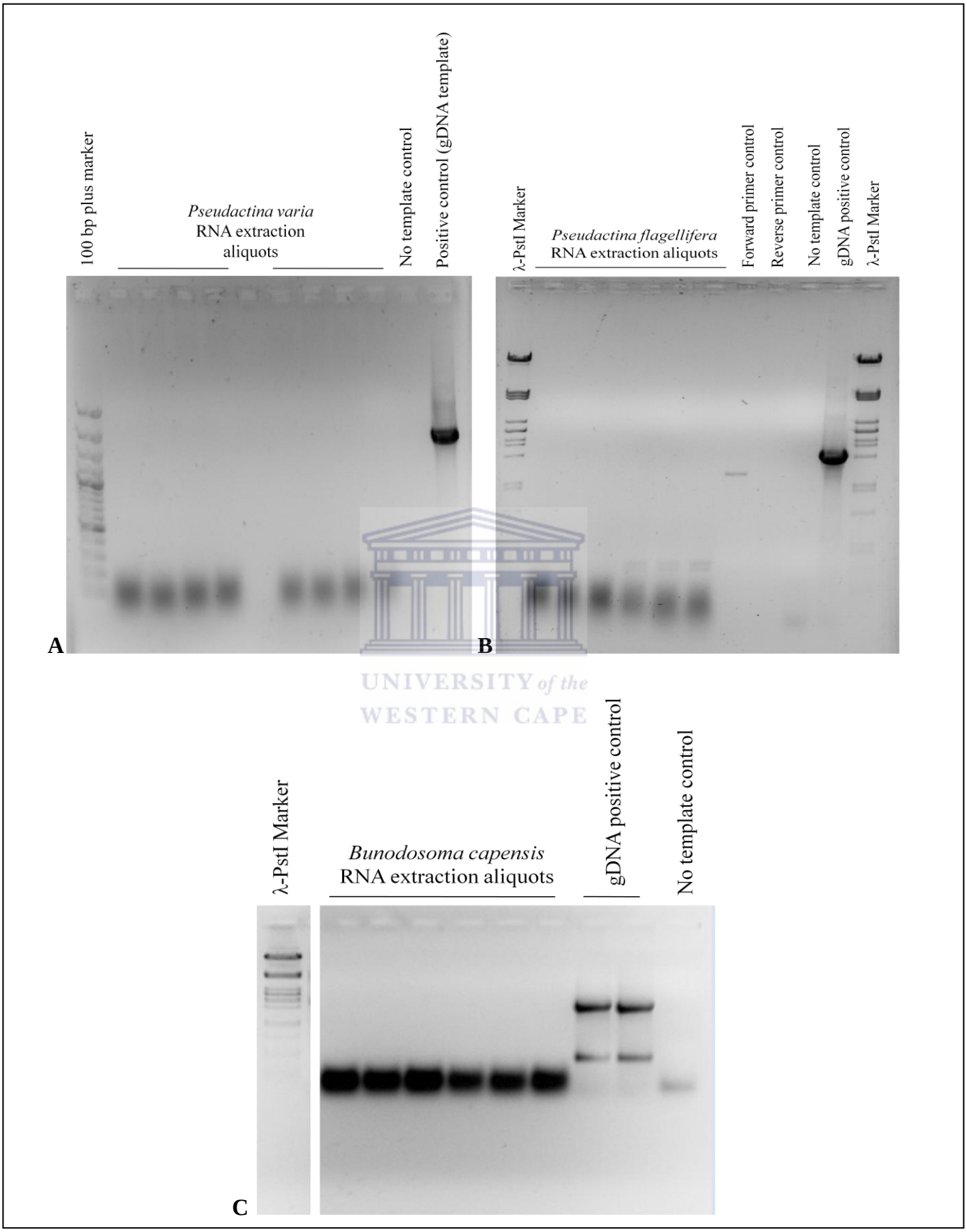


Figure 46. A) *Pseudactina varia* B) *Pseudactina flagellifera* and C) *Bunodosoma capensis* show 18S rRNA PCR to check RNA for gDNA contamination.

cDNA positive control optimizations

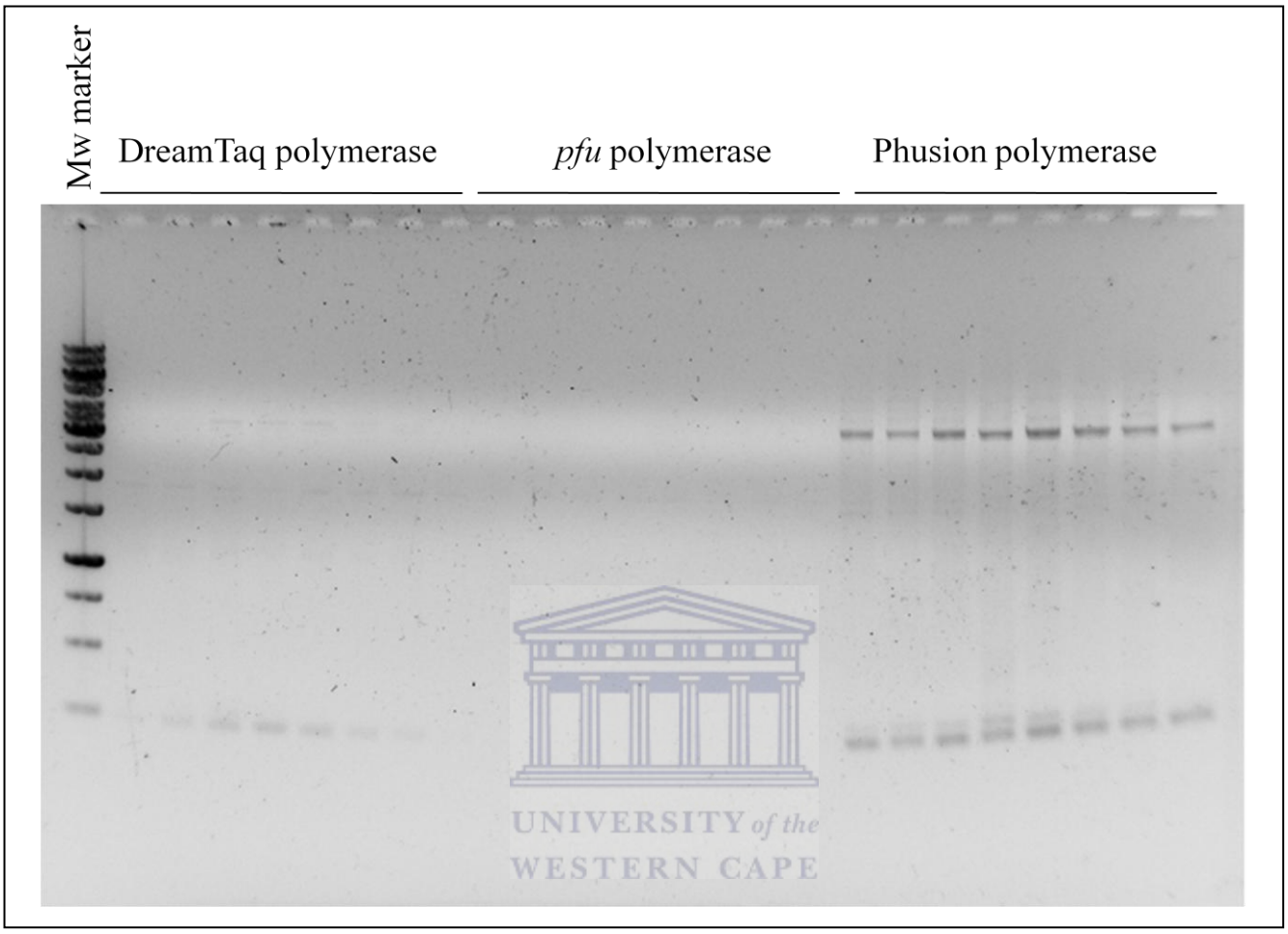


Figure 47. Effect of various polymerases on amplification. The high fidelity polymerase *pfu* was not capable of amplifying the target positive control. Phusion polymerase was only able to amplify the target in the HC buffer (which contains various denaturants which disrupt any secondary structures formations) the polymerases where tested across an annealing gradient from 52-60°C.

UNIVERSITY of the
WESTERN CAPE