Symbiodinium diversity and potential hybridisation on the highly biodiverse coral reefs of Timor-Leste

Joshua Ian Brian

A thesis submitted to the Victoria University of Wellington
in partial fulfilment of the requirements for the degree of

Master of Science

in Marine Biology

Victoria University of Wellington, New Zealand

2018

VICTORIA UNIVERSITY OF WELLINGTON
Te Whare Wānanga o te Ūpoko o te Ika a Māui

Atauro Island, Timor-Leste

Abstract

To persist in oligotrophic waters, reef-building corals rely on nutritional interactions with their intracellular symbionts: photosynthetic dinoflagellates of the genus *Symbiodinium*. This relationship is threatened by increasing environmental stress, which can stimulate loss of these symbionts from coral tissues ('coral bleaching'). Members of the genus *Symbiodinium* display high levels of genetic diversity, and demonstrate a corresponding diversity in physiological responses to environmental change. However, the true diversity and potential for genetic adaptation in this genus remain poorly characterised.

This thesis aimed to further the understanding of symbiont diversity and adaptive potential by conducting assessments of *Symbiodinium* at Atauro Island and the neighbouring Timor-Leste mainland. These sites have previously been shown to be of outstanding conservation value, with extremely high levels of coral diversity. Atauro Island also possibly hosts the highest diversity of reef fish in the world. However, the *Symbiodinium* communities at these sites have never been assessed. Two specific objectives were therefore addressed here. The first was to measure *Symbiodinium* diversity at Atauro Island (four sites) and Timor (three sites), using direct sequencing of three gene regions: *cob* gene, mitochondrion; ITS2 region, nucleus; and psbA^{ncr} region, chloroplast; in addition to Next Generation Sequencing of the ITS2 region. The second objective was to establish evidence for *Symbiodinium* hybridisation, a potentially rapid evolutionary mechanism that may facilitate adaptation to environmental stress, by looking for genetic incongruences between *Symbiodinium* organelles.

Coral symbionts of Timor-Leste were found to be similar to those from other coral reefs of the Indo-Pacific, with several host generalist and multiple host specific types observed. However, there were also several novel *Symbiodinium* types found (C15p, C15q, C1x, C1z). Despite their geographic proximity, there were strong differences observed between the symbiont communities of Atauro Island and mainland Timor. In particular, the coral genus *Pocillopora* hosted clade C symbionts exclusively at Atauro Island, while it hosted clade D exclusively at Timor sites. Other symbiont types also showed geographic partitioning, and diversity was 1.25 times higher at Atauro Island, a figure consistent for the *cob* and ITS2 regions. While Timor sites have comparable *Symbiodinium* diversity to other reefs globally, Atauro has noticeably elevated *Symbiodinium* diversity. Next Generation Sequencing affirmed these patterns, with Atauro Island sites having much more diverse cryptic populations of *Symbiodinium*, largely driven by symbionts in clade C. The exception was clade D symbionts, which were proportionally far more diverse at Timor, a pattern consistent in multiple coral genera.

There was strong evidence of genetic incongruence at two Atauro Island sites, with all testing procedures identifying genetic discordance between organellar and nuclear genomes, consistent with

theoretical predictions of hybridisation. This study therefore presents strong evidence for *Symbiodinium* hybridisation, and its corroboration by multiple loci is significant. Putative hybrid *Symbiodinium* always had a common type as one of the possible parents, with a rarer symbiont as the other. For example, one putative hybrid had organellar genes of the common generalist *Symbiodinium* C40, while it was identified as the rare type C3z with the ITS2 region. Both of these *Symbiodinium* types were also found in congruent relationships, which strongly supports the possibility that they sexually reproduced to produce the incongruent putative hybrid.

Environmental stressors, such as increased temperature, turbidity and sedimentation, are suggested reasons for lowered *Symbiodinium* diversity at Timor, as they may impose a selection pressure on corals to only keep highly beneficial symbionts. This reduction in diversity likely limits the potential for adaptive change through methods like hybridisation, and highlights the need to assess and conserve symbiont diversity to the same extent as coral diversity.

Acknowledgements

Table of Contents

List of Figures

List of Tables

Chapter 1: General Introduction

1.1. Coral reefs and symbiosis

1.1.1. Symbiosis

Symbiosis was originally defined as "the living together of unlike organisms" (de Bary 1879). Broadly, symbiosis has since been recognised as the persistent association between different species, a definition which encompasses a continuum from positive (mutualistic) to negative (parasitic) interactions (Douglas 2008). In addition, a symbiosis may switch from mutualism to parasitism, or *vice versa,* in changing conditions (Werren *et al.* 2008). These sorts of associations are both common and ecologically important, such as the symbiosis between mycorrhizal fungi and terrestrial plants, which is responsible for up to 75% of phosphorus and 80% of nitrogen taken up by land plants annually (van der Heijden *et al.* 2008). It is also apparent in most animal taxa, such as the association between the *Wolbachia* bacterium and most terrestrial arthropods (Werren *et al.* 2008).

Such associations can be ectosymbiotic, where one partner remains outside the other, or endosymbiotic, where one partner lives in the tissues of another, which can be considered a controlled infection (Hentschel *et al.* 2000). Aside from the ubiquitous case of bacteria living inside animal guts, one of the most common and ecologically relevant cases of endosymbiosis occurs between single-celled photosynthetic organisms and marine invertebrates (Yellowlees *et al.* 2008). Examples include the green alga *Chlorella* living inside hydra (Muscatine *et al.* 1974), and the cyanobacterium *Prochloron* inside ascidians (Seewaldt and Stackebrandt 1982). However, one of the most studied and globally recognised endosymbionts in this category are single-celled dinoflagellates of the class Dinophyceae (*sensu* Gomez 2012). These were identified as microalgae by the end of the mid-nineteenth century (Brandt 1881), though they were thought to be cryptomonads by authors of this time (Trench 1997). They were first successfully described as dinoflagellates by Hovasse (1923), though he considered them purely parasitic. It was not until the work of Kawaguti (1944) that they became widely recognised as potentially providing a benefit to their hosts. Today, there are eight recognised endosymbiotic genera of dinoflagellates (Gomez 2012; see also Trench 1997 for a review). The most ecologically prevalent of these genera was initially named *Gymnodinium* (Kawaguti 1944), but was seminally defined by Freudenthal (1962) as *Symbiodinium*.

1.1.2. *Symbiodinium* and coral reefs

Symbiodinium (Freudenthal 1962) are circular brown single-celled dinoflagellates (Trench and Blank 1987), known colloquially as 'zooxanthellae'. They are photosynthetic organisms, using sunlight to capture energy and transform it into organic compounds for metabolic use. In nature, *Symbiodinium* exist with a wide variety of organisms such as giant clams (*Tridacna* spp.), jellyfish, sponges and sea anemones (Baillie *et al.* 2000b; Pochon *et al.* 2006; Sachs and Wilcox 2006; Yellowlees *et al.* 2008). However, they are most commonly recognised for the association with cnidarians of the order Scleractinia, the reef-building corals (Figure 1.1a).
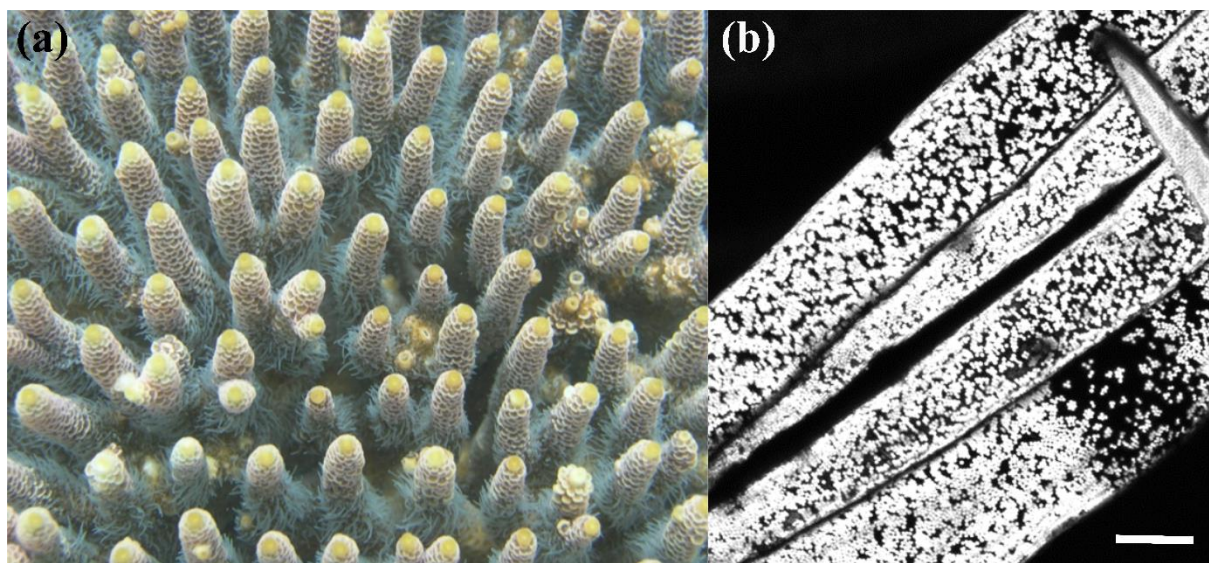
Figure 1.1: The cnidarian-dinoflagellate symbiosis. (a) *Symbiodinium* contributes nutritional support in addition to the often bright colouration of reef-building corals (*Acropora millepora*, Lamsana Inlet, Timor-Leste). (b) Confocal image of *Symbiodinium* cells (fluorescent white) inside the tentacles of sea anemone *Exaiptasia pallida*, a good model organism for coral reefs. Each symbiont is sequestered in a host cell. Scale bar is 50 μm. Both images by the author.

The association between hard corals and dinoflagellate symbionts has ancient roots: it is believed to have originated ~240 million years ago (MYA) in the Triassic (Glynn 1996; Stat *et al.* 2006) and explains the establishment of productive reefs at this time (Muscatine *et al.* 2005). It has also been hypothesised that coral-algal symbioses may have established as early as the Middle Silurian, 430 MYA (Zapalski 2014). The current *Symbiodinium* species complex only evolved ~65 MYA (Tchernov *et al.* 2004), and perhaps as recently as 50 MYA (Pochon *et al.* 2006; 2014), indicating that hard corals have had a variety of symbiotic partners through evolutionary time.

It has been suggested that cnidarians are well suited to host a photosymbiont thanks to their simplistic body plan and high surface to volume ratio, which facilitates the efficient capture of light (Venn *et al.* 2008), and allows large densities of *Symbiodinium* to be present (Figure 1.1b). It was initially thought

that these symbionts existed intercellularly between the cells of the coral polyps (Kawaguti 1964). While some bivalves do harbour *Symbiodinium* intercellularly in a tubular system (Coffroth and Santos 2005), it has since been shown that coral gastrodermal cells host one or more symbionts intracellularly, which are contained in an animal-derived vacuole known as the symbiosome (Davy *et al.* 2012). These symbionts can either be obtained through vertical transmission (being passed from parent to offspring *via* gametes), or horizontal transmission (taken up from the surrounding seawater) (Douglas 1998; Schwarz *et al.* 1999; Byler *et al.* 2013).

The cnidarian-dinoflagellate symbiosis is important as it facilitates the growth of coral reefs, and hence underpins their critical ecosystem provisioning in oligotrophic tropical waters (Odum and Odum 1955; Muscatine and Porter 1977). This is achieved through the symbionts photosynthesising and providing a large majority of the resulting materials to their coral hosts (Muscatine *et al.* 1984). It has been estimated that symbionts provide between 80% and 130% of a coral's daily metabolic needs (Muscatine and Porter 1977), with the accepted average around 95% (Falkowski *et al.* 1984; Hoegh-Guldberg *et al.* 2007).
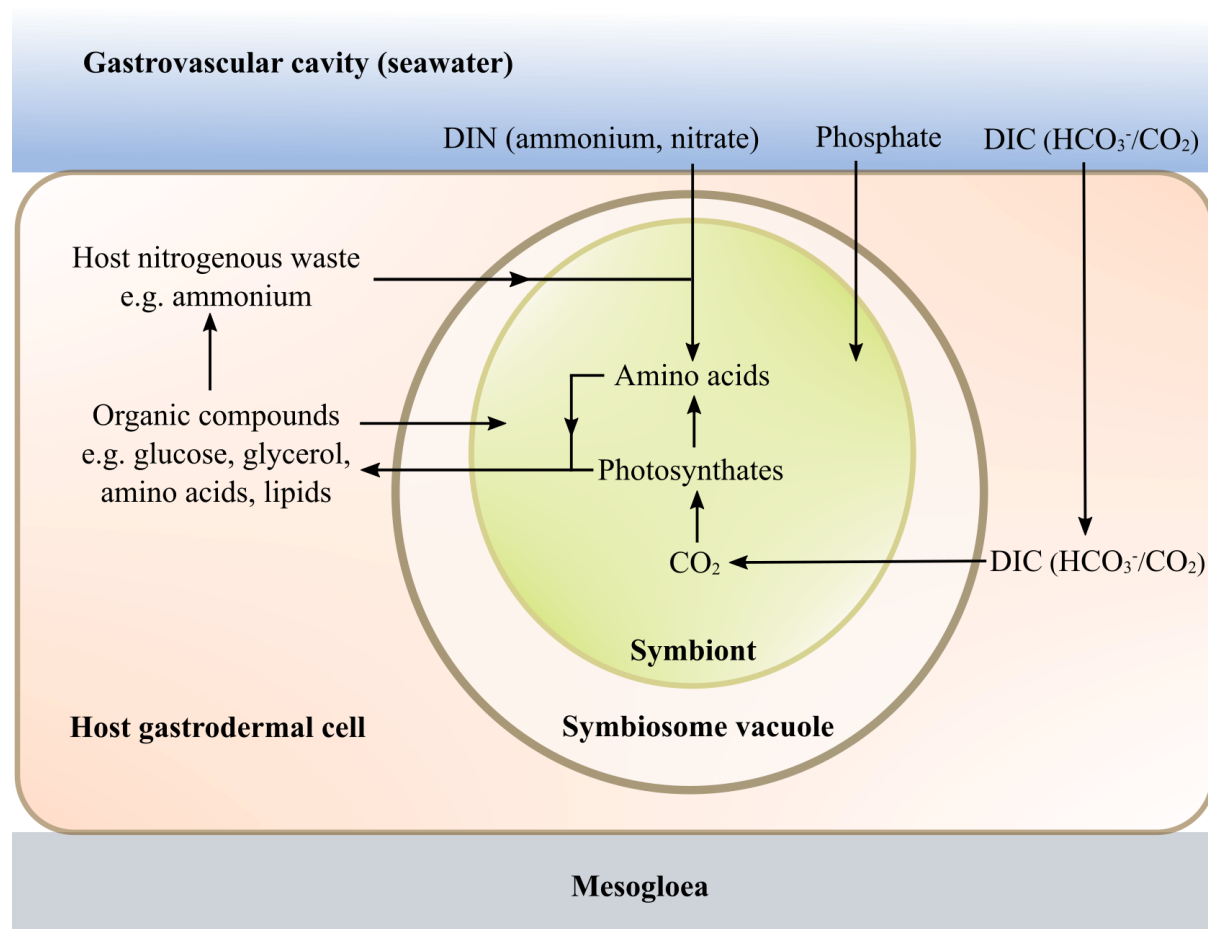
Figure 1.2: Partner location and nutritional exchange in the coral-algal symbiosis. DIN=Dissolved Inorganic Nitrogen; DIC=Dissolved Inorganic Carbon. Adapted from Davy *et al.* (2012).

This is accomplished through the transfer of high-energy carbon-based compounds from symbiont to host (Muscatine and Porter 1977; Venn *et al.* 2008; Figure 1.2), as demonstrated by the pioneering ^{14}C-radiotracer study of Muscatine and Hand (1958). Since then, it has been demonstrated that compounds transferred include glucose (Burriesci *et al.* 2012), glycerol (Trench 1979) and lipids (Kopp *et al.* 2015), as well as a wide variety of organic acids and amino acids (see Davy *et al.* 2012 for a review). These organic materials encourage host tissue growth as well as calcium carbonate accretion, ensuring that corals maintain their ecological function (Cantin *et al.* 2009). In addition, the symbiosis allows for the efficient recycling of limiting nutrients such as nitrogen by taking up host nitrogenous waste and using it in the production of compounds such as amino acids which it transfers back to the host (Muscatine and Porter 1977; Yellowlees *et al.* 2008; Figure 1.2). Further, corals in symbiosis display reduced net ammonium production, also allowing for nitrogen conservation (Wang and Douglas 1998). In return, the symbiont receives multiple benefits, such as protection from grazers and a stable position near the surface to capture downwelling light (Wilcox 1998; Davy *et al.* 2012). As such, this productive association has turned coral reefs into a socially and ecologically important ecosystem.

1.1.3. Coral importance and threats

Coral reefs are extremely important for tropical coastal communities. They provide an accessible source of protein in terms of abundant fish communities, both for local consumption and export, and protect shorelines from erosion (Moberg and Folke 1999; Cinner 2014). Coral mining provides many raw materials for building and construction, especially in nations which lack developed infrastructure (Moberg and Folke 1999). Corals are also a crucial source of tourism for many tropical nations that are still growing economically (Hughes *et al.* 2003; Pratchett *et al.* 2008), and hold a cultural and spiritual significance for many communities whose ancestry and mythology is intrinsically linked with the nearshore marine environment (Hicks *et al.* 2009). Overall, coral reefs have been valued at US$375 billion annually (Costanza *et al.* 1998; Cinner 2014).

Corals are also hugely ecologically important. Despite encompassing <0.5% of the ocean's surface (Moberg and Folke 1999), they host between one quarter and one third of the world's fish species (McAllister 1991; Allen 2008). Their topographical complexity provides a wide variety of feeding opportunities and niches (Pratchett *et al.* 2008; Graham and Nash 2013), though it is important to note that a topographically complex environment alone is not enough to support this level of biodiversity: live coral cover is crucial to maintain ecosystem function (Coker *et al.* 2014). In addition, coral reefs also provide nursery areas for juveniles of many species that do not associate with corals as adults (Coker *et al.* 2014), further emphasising their vital ecological function.

Therefore, the effects of coral loss are widely felt. Coral degradation is associated with a severe economic impact for nations which conduct a majority of their fishing around reefs such as Indonesia (Cheung *et al.* 2010), thanks to a breakdown of trophic structure and associated fish communities. There may also be significant economic losses due to a reduction in tourism (Pratchett *et al.* 2008). On a more local scale, ecological failure of reefs has led to extreme hardship for many vulnerable communities, such as those on the eastern coast of Africa (Cinner *et al.* 2012), and a corresponding degradation of cultural values (Hicks *et al.* 2009).

A significant number of the world's coral species are at an elevated risk of extinction (Carpenter *et al.* 2008). This failure of reef ecosystems is attributable to a variety of factors. Destructive fishing practices, such as dynamite fishing or cyanide fishing, break or poison coral reefs, with their continued use preventing recovery (Edinger *et al.* 1998). Ocean acidification also contributes to the breakdown of reefs, by both reducing the concentration of carbonate ions that corals use to build their skeletons and increasing skeletal dissolution into the surrounding seawater (Anthony *et al.* 2008; Edmunds *et al.* 2016). A wide range of coral diseases affect both Caribbean and Indo-Pacific populations, and are positively correlated with increased temperatures (Bruno *et al.* 2007). Further, increased runoff leads to eutrophication, which can create an imbalance in host-symbiont exchange and promotes macroalgal growth, which competes with corals for space and light (Dubinsky and Stambler 1996). Invasive organisms such as *Terpios hoshinota* ("black disease sponge") also have the potential to smother corals, leading to significant impairment or death (Madduppa *et al.* 2017). In addition, the interaction of multiple factors can have synergistic effects and accelerate coral degradation (Higuchi *et al.* 2015). However, the most well-studied and pervasive factor threatening coral reefs is extreme temperature (Iglesias-Prieto *et al.* 1992; Hoegh-Guldberg 1999).

1.1.4. Coral bleaching

Heat stress is the most common cause of coral bleaching (Carilli *et al.* 2012), a phenomenon so-called because it causes a pale white colour in corals, as opposed to the typical rich brown. This is due to either the symbionts losing their photosynthetic pigments, or a complete loss of the symbionts themselves from coral tissues (Hoegh-Guldberg and Smith 1989; Bieri *et al.* 2016). Their loss is a major contributor to coral mortality, as corals cannot maintain their metabolic needs through heterotrophy alone, leading to a cessation of calcium carbonate deposition and susceptibility to other impacts such as disease (Glynn 1996). The first verified report of bleaching associated with heat stress came from corals on the Pacific coast of Panama, which experienced a loss of symbionts during abnormally high sea-surface temperatures (Glynn 1983). Given that corals live close to their thermal tolerance anyway (Fitt *et al.* 2001), increased temperatures associated with climate change have led to catastrophic bleaching (Hoegh-Guldberg *et al.* 2007). There have been three pan-tropical bleaching events associated with El

Niño events, all in recent memory (1998, 2010, 2015-16), which led to 16% global coral mortality during the 1998 event (Hoegh-Guldberg 1999) and over 60% of corals bleaching on the Great Barrier Reef in 2016 (Hughes *et al.* 2017).

Early work implicitly linked bleaching incidence with symbiont health, concluding that bleaching occurred due to damage to photosystem II (PSII) of the symbionts (Jones *et al.* 1998; Warner *et al.* 1999). More recently, the main mechanism has been identified as oxidative stress (Lesser 2011), through the generation of reactive oxygen species (ROS) (Levin *et al.* 2016). Several proteins in the thylakoid membrane of the symbiont chloroplast are easily destabilised, and consistently repaired (Weis 2008). However, under heat stress the damage outpaces the repair mechanism, leading to dysfunction of the photosystem and a back-up of excitation energy (Weis 2008). These free electrons, along with oxygen produced by photosynthesis, go on to form ROS, such as superoxide (O_2^-) and hydrogen peroxide (H_2O_2) (Lesser 2006). These ROS can damage cellular constituents such as membranes directly, which may lead to *in situ* symbiont degradation or expulsion from the host (Tchernov *et al.* 2004). There is also the potential for high levels of ROS to stimulate the production of nitric oxide (NO) *via* an immune response, which can be produced by both hosts (Weis *et al.* 2008) and symbionts (Bouchard and Yamasaki 2008). NO has been shown to increase under heat stress and stimulate caspase-like enzyme activity, associated with apoptosis (Hawkins *et al.* 2013). Interestingly, despite the primary site of ROS production typically thought to be the symbiont, there is evidence of a host antioxidant response before the symbiont is visibly impacted (Hawkins *et al.* 2013; Krueger *et al.* 2015). In addition, heat stress was demonstrated to significantly alter the metabolic profile of compounds transferred from host to symbiont, as well as the other way around (Hillyer *et al.* 2016). Collectively, these results suggest that bleaching is due to impacts on both the hosts and the symbionts. The observation that different host-symbiont combinations have variable susceptibility to bleaching (*e.g.* Baker 2001; Bay *et al.* 2016) makes understanding symbiont diversity a high priority for the study of coral reef biology.

1.2. Symbiont diversity

1.2.1. Recognition of diversity within *Symbiodinium*

The genus *Symbiodinium* was originally thought to be monospecific, with a single representative *Symbiodinium microadriaticum* isolated from the upside-down jellyfish *Cassiopeia xamachana* (Freudenthal 1962). However, early taxonomic work began to reveal diversity that challenged this view. Gel electrophoresis of protein composition of symbionts isolated from different hosts revealed significant polymorphism, and led the authors to suggest that assigning all *Symbiodinium* to the same species was erroneous (Schoenberg and Trench 1980a). Further work on these symbiont isolates

revealed divergent morphologies and varied potential to colonise hosts (Schoenberg and Trench 1980b, c). Investigation of chromosome volumes between symbionts also suggested that there are multiple species in the genus (Blank and Trench 1985). However, it was not clear whether the differences presented just indicated different strains of the same species, due to slightly different environments inside coral hosts (*e.g.* Fitt *et al.* 1981).

The uncertainty can be ascribed to a lack of resolution in early techniques: characteristics like cell size, cell shape or protein composition do not exhibit significant variation and are poorly correlated with true phylogenies, particularly in *Symbiodinium* (Wilcox 1998; LaJeunesse *et al.* 2012). While a combination of morphological, behavioural and physiological methods was used to definitively split the genus into multiple species (Trench and Blank 1987), it was not until the advent of sophisticated genetic techniques that the true diversity within *Symbiodinium* came to be appreciated (see Section 2.2).

Figure 1.3: Phylogeny of clades (A-I) within *Symbiodinium*. Adapted from a tree constructed by S. Wilkinson for inclusion in Sheppard *et al.* (2017). Species in inverted commas lack formal description.

The seminal work of Rowan and Powers (1991a, b) identified three different 'groups' of *Symbiodinium*: A, B, and C. These are potentially as different to each other as the whole genus is to other dinoflagellates (Rowan and Powers 1992). These different groups became known as clades, a term in common usage by the turn of this century (*e.g.* Carlos *et al.* 1999; LaJeunesse 2001). There are now nine known clades,

designated A-I (Pochon and Gates 2010; Figure 1.3). Of the nine, clades A-D most commonly associate with corals (Pochon *et al.* 2014), though F and G also persist naturally with corals (Pochon *et al.* 2006). Of the remaining clades, H and I associate with foraminiferans (Pochon *et al.* 2014) and E is generally exclusively free-living (Jeong *et al.* 2014), though it has been reported inside an anemone (LaJeunesse 2001). Members of all clades have also been found in the free-living state (Pochon *et al.* 2014).

Recently, fine-scale research has also revealed marked diversity within clades, with intra-clade genetic groups generally referred to as 'types' (Finney *et al.* 2010; Silverstein *et al.* 2011). Clade C is the most diverse, with potentially hundreds of types (LaJeunesse 2005; Thornhill *et al.* 2014). However, even clades originally thought not to be that diverse have been shown to contain high variety, such as clade B (Parkinson *et al.* 2016) and clade D (LaJeunesse *et al.* 2014). The possible reasons for such diversification are numerous. Symbionts show a large number of somatic mutations, which could play a role in adaptive evolution (van Oppen *et al.* 2011a). This could be combined with a lack of gene flow which further encourages differentiation, particularly in clade C (Howells *et al.* 2009). Indeed, it is estimated that a massive radiation of *Symbiodinium* occurred 15-5 MYA, thanks to the closing of the Tethys Sea and Central American Isthmus which led to isolated populations (Pochon and Pawlowski 2006). In addition, individual coral hosts can be thought of as a distinct habitat, with subtle differences in intracellular pH and temperature (LaJeunesse 2005). This encourages niche diversification and could promote high diversity, even in those corals where the symbionts are transmitted horizontally (LaJeunesse *et al.* 2004a; Thornhill *et al.* 2014).

1.2.2. Summary of genetic tools used in *Symbiodinium* systematics

The use of genetic markers was facilitated through the use of the polymerase chain reaction (PCR), which allowed for the amplification of DNA reasonably cheaply and quickly, meaning that genes could be studied in greater detail than before. See Table 1.1 for studies that utilise the genes discussed below.

Initially, *Symbiodinium* taxonomy was primarily determined *via* the analysis of rDNA – nuclear genes in the ribosomal operon (Coffroth and Santos 2005; Barbrook *et al.* 2006). The first of these utilised was the ribosomal small subunit gene, known as the SSU or 18S gene. This was analysed with restriction fragment length polymorphisms (RFLPs) and direct sequencing (Rowan and Powers 1991a, b). However, as it is a coding region it is relatively conserved and evolves slowly (Wilcox 1998; Hansen and Daugbjerg 2009). Even at the time it was recognised as inappropriate for exploring diversity (Rowan and Powers 1991a), and has since been discredited as a useful molecular marker for *Symbiodinium* (LaJeunesse 2001). The large subunit gene (LSU or 28S) was judged to offer higher resolution and able to answer specific ecological questions (Wilcox 1998). It is useful for differentiating between different clades, and is still used to explore broad-scale diversity within the genus. It is

considered by some to be the clade-level benchmark (Pochon *et al.* 2014). However, while it has also been used to define new species (Hansen and Daugbjerg 2009), its within-clade resolution is poor and it cannot be used to differentiate closely-related types (LaJeunesse 2001). This is in contrast to the internal transcribed spacer region (ITS), consisting of the ITS1, 5.8S and ITS2 regions. While the 5.8S region is also coding and adds nothing to the information provided by the SSU and LSU, the non-coding ITS1 and ITS2 exhibit much greater phylogenetic resolution (LaJeunesse 2001). In particular, the ITS2 region was judged to be superior to all previously tested gene regions (LaJeunesse 2002). The use of the ITS2 is discussed further in Section 1.2.3.

Markers using symbiont organelles were subsequently explored. In the chloroplast, the cp23S gene was found to support the phylogenies produced by the ITS region, providing the first non-nuclear verification of *Symbiodinium* taxonomy (Santos *et al.* 2002). It evolves an order of magnitude faster on average than either the LSU or SSU (Santos *et al.* 2002), and remains in use today for fine-scale within-clade genetic studies. The psbA-D1 protein-coding region of the chloroplast minicircle has also been investigated, and shown to evolve 1.2-18 times faster than the SSU (Takishita *et al.* 2003). However, it has rarely appeared in studies since. In contrast, markers in the mitochondrial genome (*cob*, *co1*) are reasonably common, despite their protein-coding nature contributing to a slow evolutionary rate (LaJeunesse *et al.* 2012). They have largely been used to ground-truth findings from other genes.

Microsatellites, small repeated elements in the genome, are sporadically used in *Symbiodinium* taxonomy. They are excellent markers for distinguishing between similar types within clades, but remain uncommonly used due to their low transferability: new markers often have to be designed for each specific study. In terms of resolution, they complement rather than add to current markers (Thornhill *et al.* 2014). This issue regarding development of markers (especially trying to make them useful for all clades), is one of the main reasons why less than 15 gene loci have been utilised for *Symbiodinium* systematics (Pochon *et al.* 2012, 2014; Table 1.1). Of these, three loci have been considered benchmarks (Pochon *et al.* 2012): the 28S/LSU region for clade-level studies, and the ITS2 and cp23S markers for within-clade diversity.

Table 1.1: Genes or gene regions that have been used in *Symbiodinium* genetic studies. Synonyms are separated by slashes (where applicable). Genes that have been explored but have not been utilised in an actual phylogenetic study are excluded.

GENE	SELECTED REFERENCES
Nuclear genome	
nr18S / 18S / SSU	Rowan and Powers 1991a, b; Rowan and Powers 1992; Carlos *et al.* 1999; Takishita *et al.* 2003
nr28S / 28S / LSU	Wilcox 1998; van Oppen *et al.* 2001a; LaJeunesse *et al.* 2003; Pochon *et al.* 2006; Sampayo *et al.* 2009; Pochon *et al.* 2014; Rouzé *et al.* 2017
ITS1	Baillie *et al.* 2000a; van Oppen *et al.* 2001a, 2005; Rodriguez-Lanetty 2003; Sampayo *et al.* 2009; Cumbo *et al.* 2013
5.8S	Sampayo *et al.* 2009
ITS2	LaJeunesse *et al.* 2003; LaJeunesse 2005; Sampayo *et al.* 2009; Finney *et al.* 2010; Pochon and Gates 2010; Kemp *et al.* 2015; Wilkinson *et al.* 2015; Ziegler *et al.* 2017a; this study
elf2	Pochon *et al.* 2014; Hume *et al.* 2016
Mitochondrial genome	
cytochrome b / cob / cyt b	Zhang *et al.* 2005; Sampayo *et al.* 2009; LaJeunesse *et al.* 2012; Pochon *et al.* 2014; Hume *et al.* 2015; Parkinson *et al.* 2015b; this study
cytochrome oxidase 1 / co1 / cox1	Takabayashi *et al.* 2004; Pochon *et al.* 2014; Bongaerts *et al.* 2015b
Chloroplast genome	
cp23S	Santos *et al.* 2002; Pochon *et al.* 2006; Sampayo *et al.* 2009; Pochon and Gates 2010; Pochon *et al.* 2014; Hume *et al.* 2015
psbA minicircle non-coding region / psbA[ncr]	LaJeunesse and Thornhill 2011; Pinzon and LaJeunesse 2011; Jeong *et al.* 2014; Thornhill *et al.* 2014; Hume *et al.* 2015; this study
psbA minicircle D1 protein-coding gene / psbA-D1	Takishita *et al.* 2003; Pochon *et al.* 2014; Hume *et al.* 2016
Microsatellites	
Various tandem repeat regions	Santos and Coffroth 2003; Pettay and LaJeunesse 2007; Finney *et al.* 2010; Wham *et al.* 2011, 2014; LaJeunesse *et al.* 2012, 2014; Thornhill *et al.* 2014

1.2.3. Assessing diversity with the ITS2

Of the three benchmarks, the ITS2 region is by far the most commonly used (Stat *et al.* 2006, 2012; Wilkinson *et al.* 2015; Ziegler *et al.* 2017a). Thanks to the tandem repeat of rDNA genes in the nucleus, it is multiple copy (Arif *et al.* 2014). It is also reasonably short, ranging from 180-320 base pairs, though some may be shorter or longer (LaJeunesse 2001; Ziegler *et al.* 2017a). This combination means that it is easy to detect and amplify, even if the DNA is slightly degraded (Yao *et al.* 2010). In addition, it is biparentally inherited (unlike organellar genes), which means ITS2 variation within the genome can signify recombination (Baldwin *et al.* 1995; Rybalka *et al.* 2013). It is considered a universal barcode marker for several large taxonomic groups, including plants (Yao *et al.* 2010) and fungi (Schoch *et al.* 2012).

1.2.3.1. Issues with the ITS2 region

There are, however, issues associated with the ITS2 region. Despite its non-coding nature, it has recently been posited to have quite a slow evolutionary rate (Thornhill *et al.* 2014), and may not be able to fully resolve *Symbiodinium* species, especially in the highly diverse clade C (LaJeunesse 2005). Another problem is the method of assessment of the ITS2. Traditionally, this has been done *via* PCR and then denaturing gradient gel electrophoresis (DGGE), a method that can separate PCR products of the same length but different base composition (Muyzer and Smalla 1998). The dominant bands, representing the most common PCR products, are then excised and potentially sequenced (LaJeunesse 2002). This became a common way of assessing diversity in communities, particularly for symbionts inside a coral host (Arif *et al.* 2014; *e.g.* see LaJeunesse *et al.* 2003, 2004a, 2005). This can be problematic: weaker bands can be masked by background staining (Muyzer and Smalla 1998), potentially missing less common ITS2 variants. Indeed, it has been shown that PCR-DGGE is unable to detect symbiont populations at less than 5-10% abundance inside a coral (Pochon *et al.* 2006; Thornhill *et al.* 2006; LaJeunesse *et al.* 2008; Quigley *et al.* 2014). There is also a lack of reproducibility: it is down to the individual researcher as to what bands to sequence, and machine differences may mean that band strength and migration distance are inconsistent between studies.

This is exacerbated by intragenomic variation found within the ITS2. Because it is multiple-copy, a single cell can exhibit different ITS2 sequences due to a range of processes. In some cases, ITS2 sequences can vary more between cells of the same type than between different types (Stat and Gates 2010). One of the most common causes of this is incomplete lineage sorting (ILS) (Som 2014; Figure 1.4). In ILS, polymorphisms in a single ancestral lineage are not fully separated in descendant lineages, leading to different copies of the gene in a single organism. Many studies report such intragenomic variation in *Symbiodinium* from both the ITS1 and ITS2 region (*e.g.* Baillie *et al.* 2000a; van Oppen *et al.* 2005; Thornhill *et al.* 2007; Arif *et al.* 2014), and this is often attributed to ILS (LaJeunesse *et al.* 2004a; Arif *et al.* 2014). However, other evolutionary mechanisms such as hybridisation can also generate intragenomic variation (Wilkinson *et al.* 2015). Despite the prevalence of intragenomic variation in *Symbiodinium*, it remains largely unquantified (Thornhill *et al.* 2007; Arif *et al.* 2014). Using the insensitive DGGE method of assessment, such subtleties are not considered.

1.2.3.2. Advances for the ITS2 region

A recent advance has been quantitative PCR (qPCR). This is a more sensitive technique and allows the visualisation of low-frequency ITS2 variants (Zhou *et al.* 2011; Bay *et al.* 2016). Using qPCR,

11
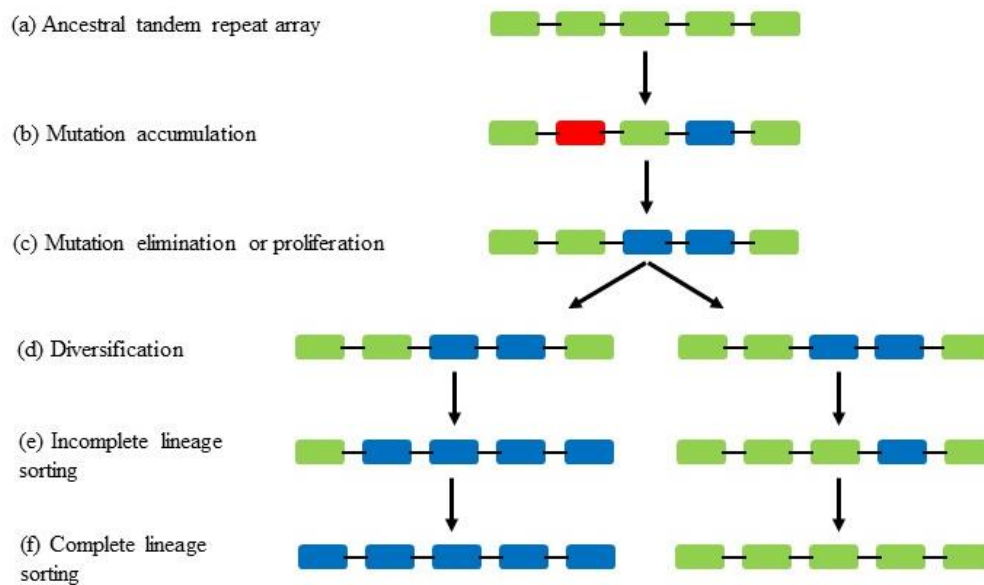
Figure 1.4: Lineage sorting in a multi-copy gene. (a) An ancestral lineage with multiple copies (green squares) of a single gene. (b) Gene copies mutate (red and blue squares) during duplication or repair, and persist due to redundancy in the genetic code or through their presence in a non-coding region. (c) Through processes like unequal crossing over during meiosis, mutations can either proliferate (blue mutation) or be eliminated (red mutation). (d) Through routes like allopatry, the ancestral lineage may diversify into multiple descendant lineages. (e) The processes in (c) continue, leaving lineages with different copies of the gene region within the same genome. Lineages characterised at this stage display intragenomic variation. (f) If enough generations have passed, the gene regions will completely homogenise and give a consistent phylogenetic signal (subject to new mutations arising). Adapted from Wilkinson (2015).

researchers have detected significant background populations of different symbionts in corals thought previously to exclusively host one specific type (*e.g.* Mieog *et al.* 2007; Correa *et al.* 2009; Silverstein *et al.* 2012; Kennedy *et al.* 2015). While this approach has increased resolution in the ITS2, it is still impossible to tell if rare divergent sequences represent a separate symbiont population or intragenomic variation. Bacterial cloning has been suggested as a way of separating out all the sequences in a sample, but this can overinflate diversity estimates by increasing the frequency of rare sequences (Thornhill *et al.* 2007; LaJeunesse and Thornhill 2011).

Because of these issues, an operational taxonomic unit (OTU) has been proposed, where everything less than 0.03 genetic units is collapsed into a single OTU (which is treated as a 'species'); the variation within this OTU is believed to result from meaningless intragenomic differences (Arif *et al.* 2014). This has allowed the partitioning of intra- and inter-genomic variation (Quigley *et al.* 2014; Thomas *et al.* 2014). This approach has been combined with advances in Next Generation Sequencing (NGS), which can return readable sequences for all variants in a sample as opposed to just the most common (as *per* normal sequencing of PCR products). Thanks to lowered costs, this method has become increasingly frequent in the last four years for studying symbiont diversity, and has led to a 10,000-fold increase in sensitivity (*e.g.* Arif *et al.* 2014; Quigley *et al.* 2014; Thomas *et al.* 2014; Cunning *et al.* 2015b; Bay *et*

al. 2016; Ziegler *et al.* 2017a). However, the OTU approach also has limitations. For example, the ecologically distinct *Symbiodinium* types C1 and C3 are separated by just one base pair in the ITS2 region, meaning they would likely be clustered in a single OTU if the researcher was not aware of this.

1.2.4. Beyond the ITS2

Analysing diversity within *Symbiodinium* is made more difficult by the difference in copy number between different clades (Arif *et al.* 2014). For example, clade D has three times as many copies of the ITS1 region as clade C does (Mieog *et al.* 2007). This can lead to incorrect comparisons of diversity, as a higher copy number may lead to higher levels of intragenomic variation (Thornhill *et al.* 2007; LaJeunesse and Thornhill 2011). Recently, other markers have been developed that may delineate species in a more convincing manner. The chloroplast psbA minicircle non-coding region (psbA^{ncr}) was first suggested in the mid-2000s (Moore *et al.* 2003; Barbrook *et al.* 2006), but was not utilised in a diversity study until 2011 (LaJeunesse and Thornhill 2011; Pinzon and LaJeunesse 2011). Since then, it has been used to define new symbiont species (*e.g.* Hume *et al.* 2015; LaJeunesse *et al.* 2015), and has been shown to be very fine-scale, evolving 10-20 times faster than the ITS2 (Thornhill *et al.* 2014). In addition, intragenomic variation appears to be less of an issue (LaJeunesse and Thornhill 2011). However, its high variability makes it difficult to align: in one study even sequences from two types of the same clade could not be aligned (Thornhill *et al.* 2014).

Other markers have also been determined to have potential, such as the *actin* gene (Watanabe *et al.* 2006; Pochon *et al.* 2012) and the rad24 gene (Pochon *et al.* 2012). However, these are yet to be used extensively in *Symbiodinium* systematics.

The *Symbiodinium* taxonomy field has reached the realisation that a single marker by itself is not appropriate for symbiont taxonomy (Pochon *et al.* 2012; Arif *et al.* 2014; Ziegler *et al.* 2017a). As well as the advantages and drawbacks they all possess, a single marker cannot be used to fully understand issues of hybridisation, horizontal gene transfer or indeed any issue impacting the evolutionary history of diversity (Som 2014; Govindarajulu *et al.* 2015). Increasingly, multiple markers are being used. Initially this took the form of concatenated datasets used to verify previous observations (Pochon *et al.* 2006). It has now become common to simultaneously use genes from the nucleus, chloroplasts and mitochondria in diversity studies, as this approach produces defendable results (*e.g.* Sampayo *et al.* 2009; LaJeunesse *et al.* 2012, 2014).

Despite this, there is still a lack of consensus on what constitutes a species within the genus *Symbiodinium* (Coffroth and Santos 2005), due to a continuing lack in our understanding of the observed variation (Parkinson and Baums 2014). It has been suggested that studies on gene *expression*

are required, as this is another way in which closely-related taxa can diverge (Parkinson *et al.* 2016). Nevertheless, while issues remain, it has been convincingly established that there is significant diversity within the genus *Symbiodinium*. Further, this diversity is functionally relevant.

1.3. Functional and ecological diversity of *Symbiodinium*

1.3.1. Symbionts and temperature

There is significant variation exhibited by *Symbiodinium* in response to a range of environmental factors. The most widely-studied of these is temperature, particularly with respect to the differences between clades C and D. Generally, clade D is better at dealing with thermal stress, both in terms of abnormally cold (McGinley *et al.* 2012) and hot (Oliver and Palumbi 2011) sea surface temperatures. For example, clade D symbionts maintain coral growth rates at temperatures that would otherwise lead to coral degradation (*e.g.* Baker *et al.* 2013; Cunning *et al.* 2015a). However, it is erroneous to make broad-scale generalisations based on clades, as these wide phylogenetic groups do not correlate with response to temperature (Warner *et al.* 1999; Tchernov *et al.* 2004). For example, one of the most heat-resistant symbionts known, *Symbiodinium thermophilum* from the Persian/Arabian Gulf, is a variant of the C3 subgroup (Hume *et al.* 2015; Howells *et al.* 2016b), showing how members of even a supposed heat-susceptible clade can confer high tolerance to their coral hosts. This shows how detailed fine-scale understanding of the diversity in symbiont communities is crucial to understanding how they may react to climate change.

It may be questioned whether these differences are just due to acclimation at various locations. However, evidence has shown a definite genetic basis to these different thermal tolerances (*e.g.* Grégoire *et al.* 2017). For instance, clade D symbionts display a high heritability of photoprotective pigments, which allows for rapid thermal adaptation (Császár *et al.* 2010). They also have highly specific genes involved in photosynthesis and the thylakoid membrane (Ladner *et al.* 2012), and significantly differ in gene expression in periods of thermal stress (Barshis *et al.* 2014). These differences allow thermally tolerant symbionts to scavenge (and hence reduce) ROS better (Tchernov *et al.* 2004; McGinty *et al.* 2012; Levin *et al.* 2016). They can even reduce electron flow, which generates ROS in the first place (Aihara *et al.* 2016). However, there can also be thermal acclimation within genetically indistinguishable types, depending on the temperature at which their host lives. This has been shown to be applicable to both clade A (Takahashi *et al.* 2013) and clade C (Howells *et al.* 2012; 2016b). In general, this demonstrates that there is functional variance at all levels of symbiont diversity.

1.3.2. Symbionts and other environmental factors

Symbiodinium genetic diversity also correlates to different functional responses to a range of other environmental factors. The functional diversity in response to disparate light regimes was one of the first responses to be studied in *Symbiodinium*: as photosynthetic organisms, light is the most important resource to respond to (Hennige *et al.* 2009; Parkinson *et al.* 2016). Like temperature, there are broad clade-level differences in response to light regime (Rowan and Knowlton 1995; Rowan *et al.* 1997; Iglesias-Prieto *et al.* 2004). Indeed, certain symbiont types may occupy specific parts of a coral host based on available light (Kemp *et al.* 2015). In addition, the same coral species may host different symbiont populations at different depths due to a variable light regime (Bongaerts *et al.* 2015a). This also has a genetic basis, with highly differential gene expression among types in clade B in response to light availability (Parkinson *et al.* 2016).

Symbionts can also influence host health with reference to a range of other factors. The resident symbiont community can impact host susceptibility to disease; for example, clade D increases resistance relative to clade A (Rouzé *et al.* 2016). There are also differences in UV protection, with clade A producing mycosporine-like amino acids (MAA) that act like a natural sunscreen (Banaszak *et al.* 2000). This was independent of actual UV regime, which again reveals a genetic basis to these differences, as it is not just acclimation to the local conditions (Banaszak *et al.* 2000). However, it must be recognised that all of these examples represent gross generalisations across clades, and further study is likely to reveal diverse patterns both among and within clades.

Further, multiple environmental influences can act over even small spatial scales (Cunning *et al.* 2015b). This will affect host health, and govern niche differentiation in *Symbiodinium*. As such, an understanding of host-symbiont associations is critical for predicting a response to future environmental impacts.

1.3.3. Specific vs. general associations between corals and symbionts

Initial work suggested that a coral hosted one specific symbiont *i.e.* a one-to-one association (Rowan and Powers 1991a). This hypothesis makes two independent claims: (a) there is only a single *Symbiodinium* type within a coral colony, and (b) the symbiont type found in a certain species differs from types found in other species. Recently, both these assertions have been challenged.

1.3.3.1. A single resident population in a coral host?

Background populations were thought not to exist, as it was hypothesised that genetic heterogeneity would cause significant instability (Baums *et al.* 2014). In particular, it was thought that this would encourage parasitism: co-existence would lead to competition between symbiont types, and the ones that took most advantage of their host would be most successful (Schoenberg and Trench 1980c; Douglas 2008). This appeared to be confirmed by early diversity studies (see review in Trench 1993). However, in many cases this was due to symbiont consortia being studied through culturing rather than *in hospite*: this had the effect of reducing diversity down to the dominant type, which would outgrow potential background populations in laboratory conditions (Santos *et al.* 2001; Baker 2003). While it is true that there is generally one highly dominant type inside a colony (Thornhill *et al.* 2014), it has been shown that a single colony can host a diverse range of background symbionts (Santos *et al.* 2001; Silverstein *et al.* 2012; Kennedy *et al.* 2015; but see Lee *et al.* 2016). In fact, there can be up to four clades contained within a single host colony (Silverstein *et al.* 2012; Kemp *et al.* 2015). The threat of encouraging parasitic types appears to be ameliorated through a variety of mechanisms such as zonation within a host, using variables like light (Rowan *et al.* 1997; Kemp *et al.* 2015). Therefore, it has been convincingly established that a single host can contain multiple diverse symbiont types, which can persist through time.

1.3.3.2. Species-specific associations?

There are a wide range of host-symbiont associations on coral reefs, which can best be described as a continuum. On every reef there appear to be a few generalist symbionts associating with lots of coral species, as well as generalist hosts that can associate with different symbiont types, and then many specific host-symbiont relationships, with some falling in between these extremes (LaJeunesse *et al.* 2004a; Stat *et al.* 2006). Generally, a single coral colony associates with one dominant symbiont type in a relationship that remains stable through time (*e.g.* Baums *et al.* 2014; Thornhill *et al.* 2014; Stat *et al.* 2015). However, a meta-analysis showed that there are no obligate one-to-one relationships present in the coral-algal symbiosis (Fabina *et al.* 2012).

These associations are influenced by the host's mode of symbiont acquisition. Corals that gain their symbionts horizontally (*i.e.* from the surrounding environment) are more likely to host generalist symbionts, while corals that gain their symbionts *via* vertical transmission (*i.e.* from their parent) host symbionts that are often genus- or species-specific (Stat *et al.* 2008; Fabina *et al.* 2012). The generalist pattern observed in horizontally transmitting corals is explained by the nature of the acquisition: it has been shown that some juvenile coral settlers take up many different symbiont types, and a dominant type is then 'chosen' from among the highly variable initial consortium (Coffroth *et al.* 2001; van Oppen

2001; Poland *et al.* 2013; Yamashita *et al.* 2013). It is therefore intuitive that from these consortia, generalist symbionts that have the capacity to associate with diverse groups of corals would come to dominate. This also explains why the same coral species can have different dominant symbionts in different biogeographical regions (*e.g.* Finney *et al.* 2010; Silva-Lima *et al.* 2015), that is, the host will associate with a symbiont that may be better suited to the local conditions. However, horizontally transmitting corals can also have highly conserved and specific associations with symbionts (Santos *et al.* 2004), revealing the host cellular environment to be a major driver of ecological diversification in *Symbiodinium*.

There are also broad-scale geographic patterns in host-generalist/host-specific symbionts. The Caribbean is noted for having a much higher proportion of host-specific symbionts than the Indo-Pacific, where generalists often dominate reef communities (LaJeunesse *et al.* 2004a, 2008; Finney *et al.* 2010). This has been hypothesised to be due to more geographic isolation between coral populations in the Caribbean (Finney *et al.* 2010) or greater environmental stability in the Indo-Pacific (LaJeunesse *et al.* 2008). However, there are also several places, such as Hawaii, where this pattern is reversed (*e.g.* LaJeunesse *et al.* 2004b; Stat *et al.* 2015). This shows that different locations need to be studied in detail in order to gain a complete understanding of these relationships.

Overall, evidence suggests that there are consistent associations in a single coral *colony*, with symbionts that may be general or specific. However, these associations can vary within a coral *species*, depending on location and environmental regime. In most cases, there also appear to be significant background symbiont populations. This potential for a range of associations leads to the possibility of adaptive change.

1.4. Mechanisms of adaptive change in the coral-algal symbiosis

1.4.1. 'Switching' and 'shuffling'

With the recognition that corals could potentially associate with multiple symbiont types, ideas concerning the natural adaptation of coral reefs to climate change were rapidly promoted. One of the first hypotheses proposed was the adaptive bleaching hypothesis (ABH; Buddemeier and Fautin 1993). The ABH states that bleaching is actually an adaptive mechanism: the resident symbionts are removed, allowing more thermally resistant types to be taken up and proliferate inside the host, thus maintaining ecological function in stressful conditions. This gave rise to the ideas of 'switching' and 'shuffling' (*sensu* Baker 2003). In switching, corals take up novel symbiont types, while in shuffling a better-adapted background population of symbionts proliferates inside the host (Figure 1.5).
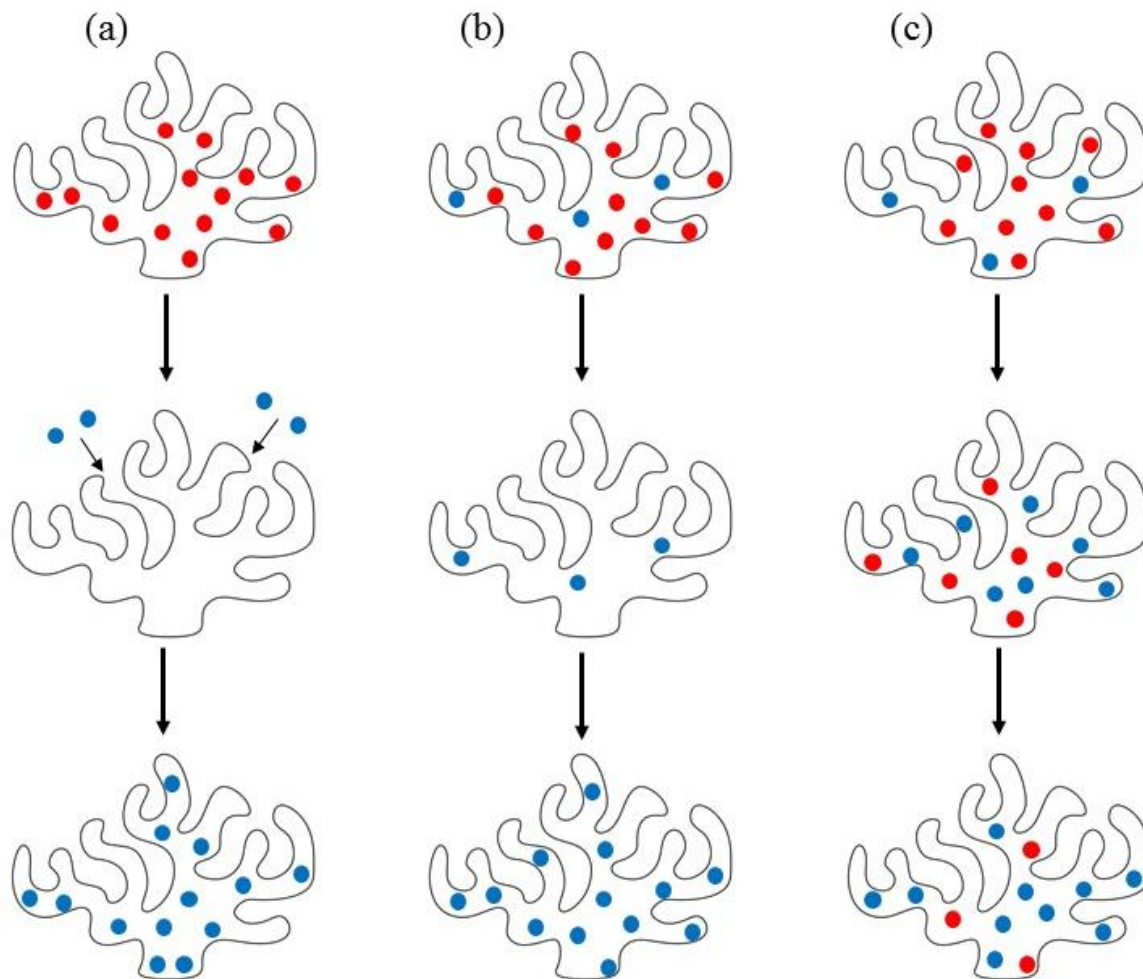
Figure 1.5: The adaptive bleaching hypothesis. (a) Symbiont switching *via* bleaching and subsequent uptake of a novel symbiont partner. (b) Symbiont shuffling *via* death or loss of the dominant symbiont type and proliferation of a thermally tolerant background type. (c) Symbiont shuffling *via* a subtler shift in the dominant symbiont type. Red and blue circles represent two different symbiont types.

1.4.1.1. Symbiont switching

Initial work focused on symbiont switching, where an exogenous symbiont type is taken up by a bleached coral (Figure 1.5a). There was early evidence that this mechanism occurred. Lewis and Coffroth (2004) experimentally induced bleaching in the octocoral *Briareum* sp., and showed that it took up a novel type of symbiont, providing proof-of-principle for switching as a mechanism that may help scleractinian corals. A similar study was conducted by Baker (2001), who stimulated bleaching in a range of Caribbean corals *via* depth transplantation, and also showed them taking up novel types at their new depths. From this, he concluded that coral reefs will be a lot more resilient than is given credit for. However, there were issues with these studies. In both, insensitive detection methods were used to characterise symbiont communities, meaning that it cannot be ruled out that the observed changes in symbiont communities were through proliferation of a cryptic background population. In the octocoral study, it has since been noted that the 'novel' symbiont taken up actually occurs naturally in *Briareum*

(Stat *et al.* 2006), and so this does not provide adequate evidence of true switching. The Caribbean study was also highly criticised in the literature: there were several major issues with experimental design, such as inconsistent recovery conditions for the corals (Hoegh-Guldberg *et al.* 2002).

Perhaps unsurprisingly, studies since have been unsuccessful in replicating these early results. It has been found that while technically an exogenous novel symbiont can be taken up by a bleached coral, this association is highly ephemeral and soon collapses (Coffroth *et al.* 2010). Recent evidence from Boulotte and colleagues (2016) used sensitive NGS techniques, and also documented reef-building corals taking up novel symbiont types after bleaching events. However, these novel types largely remained below 1% abundance of the symbiont populations (with one exception reaching 33%), and the uptake was not accompanied by a wholesale loss of the resident symbiont populations. Therefore, this is severely unlikely to have a functional impact on the resilience of coral populations to thermal stress.

Because of the highly equivocal nature of switching evidence, the emphasis has subsequently been placed on symbiont shuffling (*e.g.* see the changing emphasis in the studies of Baker 2001, 2002, 2003; Baker *et al.* 2004).

1.4.1.2. Symbiont shuffling

Early studies provided implicit evidence for symbiont shuffling. Rowan *et al.* (1997) and Baker *et al.* (2004) observed that there was a higher incidence of thermally resistant types in the years following bleaching events, which had previously been rare. However, these were only reef-scale patterns and were not assessed on a colony-by-colony basis. Therefore, the pattern could also have been due to mortality of colonies that hosted the more common vulnerable type, leaving only the few colonies that hosted a resistant type to subsequently dominate the communities (*e.g.* Sampayo *et al.* 2008). The first definitive experimental evidence for shuffling showed that a change in the dominant symbiont from clade C to clade D inside a single host increased thermal tolerance by up to 1.5°C and prevented bleaching (Berkelmans and van Oppen 2006). Further experimentation in the symbiotic sea anemone *Entacmaea quadricolor* also showed the potential for symbiont shuffling and subsequent adaptation to a stressful temperature regime (Hill *et al.* 2014; Pontasch *et al.* 2014). However, these two studies found opposite results regarding which symbiont type increased in density, something Hill *et al.* (2014) attribute to different experimental designs. As such, this questions the ecological relevance of such studies.

Observational evidence from field studies (*i.e.* natural conditions) has provided weak support for shuffling. While it has been in observed in individual colonies, it was either highly atypical (Thornhill

et al. 2006), or resulted from death or loss of the main symbiont population (Figure 1.5b), but still led to mortality of most colonies as there was no subsequent increase in the densities of background populations (Jones *et al.* 2008). Generally, there is little evidence for shuffling outside of a manipulated setting (Coffroth and Santos 2005). Given the range of factors that *Symbiodinium* are adapted to over small spatial scales, highly controlled experiments may simplify conditions and allow artificial, unsustainable associations to persist. A significant body of evidence now suggests that the underlying principle behind the ABH, host-symbiont flexibility, is misguided.

1.4.1.3. Limitations of shuffling and switching

While shuffling might occur, this may not be a long-term solution to warming conditions. Most evidence of shuffling has observed a change to clade D symbionts, which are generally considered more thermally tolerant. However, clade D symbionts in general provide less organic material to the host (Baker *et al.* 2013), and PSII functions at a lower capacity than other clades (Cantin *et al.* 2009). A study by Bay *et al.* (2016) showed that after a bleaching event, corals shuffled to more resistant clade D symbionts, but these colonies then died a few months later, potentially due to a less beneficial symbiosis. Integrated metabolite and transcriptome data confirmed clade D symbionts may exhibit sub-optimal nutritional interactions with cnidarian hosts (Matthews *et al.* 2017). Therefore, such shuffling events may actually represent more opportunistic, parasitic symbionts taking advantage of the weakened resident symbionts (LaJeunesse *et al.* 2015). It has been proposed that shuffling will lead to long-term reductions in calcification rates, and may even threaten the future of coral reefs, especially in the Caribbean (Ortiz *et al.* 2013; Kennedy *et al.* 2015). While an ephemeral association may be all that is required to see a coral through a period of stress (Correa and Baker 2011), questions remain as to what happens when this stress becomes continuous, as increased temperature is predicted to do in the coming decades.

It should be noted that clade D symbionts are hosted naturally by many corals that persist quite happily, especially in the genera *Acropora* and *Pocillopora* (*e.g.* LaJeunesse *et al.* 2008; Czászár *et al.* 2010, Stat and Gates 2010; Rouzé *et al.* 2016; Chapter 2).This again highlights the risk of making clade-based generalisations, and draws attention to the fact that it may be an inherent *change* that is harmful, rather than any one symbiont type. Long term studies have shown that coral colonies maintain their dominant type, even in the face of environmental stress. For example, some *Pocillopora* colonies in the eastern Pacific host both *Symbiodinium* D1 (heat-resistant) while others host C1b-c (heat-sensitive), with the other symbiont type then found at background levels. After a thermal stress event, those colonies hosting C1b-c as their dominant type maintained this population and subsequently died, even though D1 was present at background levels and could have been shuffled (LaJeunesse *et al.* 2010; McGinley *et al.*

2012). Other studies report similar trends of long-term stability, even when shuffling may have been thought to be beneficial (*e.g.* Thornhill *et al.* 2006; Palumbi *et al.* 2014; Rouzé *et al.* 2016).

This suggests that these associations may be constrained in some way, by factors such as host genetics (Pandolfi *et al.* 2011) or differences in symbiont size between clades (Biquand *et al.* 2017). A functional symbiosis requires a suite of recognition molecules and processes for establishment and maintenance, which remain poorly understood (Davy *et al.* 2012). This is likely to be particularly limiting for symbiont switching. However, even shuffling may be challenging if corals have adapted to hosting a certain *Symbiodinium* type as its dominant symbiont. For instance, some regions show significant evidence for co-adaptation between host and symbiont, implying a symbiosis that has been consistent over evolutionary time-scales (Stat *et al.* 2015). There may be a subset of flexible coral taxa, like *Acropora* and *Pocillopora* (Putnam *et al.* 2012). However, these flexible taxa are often the most vulnerable and the first to bleach under environmental stress (*e.g.* Hoegh-Guldberg and Salvat 1995). This suggests the ability to switch or shuffle is not associated with greater stress tolerance, and challenges the very premise of the ABH as a mechanism for long-term coral persistence.

1.4.2. Host-symbiont co-adaptation

It has recently been proposed that interactions between host and symbiont genotypes provide significant scope for physiological variation (Baums *et al.* 2014; Palumbi *et al.* 2014; Howells *et al.* 2016b). This postulates that adaptation to temperature and other stressors takes place by both the host and the symbiont, leading to superior combinations (Rodriguez *et al.* 2009; Parkinson and Baums 2014). In fact, host gene expression may be able to positively influence symbiont performance (Parkinson *et al.* 2015a), a logical but largely ignored possibility in terms of environmental stress. This has been extended to suggest that past exposure to bleaching conditions can make the community more resistant, through adaptation of both symbiotic partners (*e.g.* Williams *et al.* 2010; Carilli *et al.* 2012; Guest *et al.* 2012).

However, evidence suggests that this requires a highly specific set of circumstances, such as a sub-bleaching increase in temperature, followed by a recovery period, before being exposed to temperatures exceeding the bleaching threshold (Ainsworth *et al.* 2016). On most reefs that showed 'increased resilience' following past bleaching conditions, this was due to death of susceptible colonies and survival of resistant ones that then proliferated (*e.g.* Guest *et al.* 2012; Pratchett *et al.* 2013). While this supports the notion that some combinations are better-adapted, it does not provide for any means of adaptive change at the pace required – if extreme temperatures bleach even the resistant ones, there will be no corals left. With the level of predicted warming to come, there are grave doubts that well-adapted host-symbiont combinations will be enough for corals to survive (Silverstein *et al.* 2015; Hughes *et al.* 2017).

1.4.3. Symbiont sex and hybridisation

It has long been held that sex should be greatly reduced or eliminated in endosymbiotic organisms, to prevent the host losing control of its symbionts (Law and Lewis 1983). However, there is evidence that endosymbionts can engage in frequent sex (*e.g.* Chesnick and Cox 1987), and it was thought likely that *Symbiodinium* does engage in sex at some frequency (Schoenberg and Trench 1980a; Santos and Coffroth 2003; Santos *et al.* 2004). This has since been verified, through multiple lines of evidence such as: (a) recombinant genotypes (Baillie *et al.* 2000b; Pettay *et al.* 2011; LaJeunesse *et al.* 2014; Thornhill *et al.* 2014); (b) linkage equilibrium, implying homogenisation of gene regions *via* sex (Baums *et al.* 2014; Thornhill *et al.* 2014, 2017); and (c) evidence of functional (*i.e.* not residual) meiosis genes, which are only used in sex (Chi *et al.* 2014; Levin *et al.* 2016). These studies have led to the conclusion that sex often occurs in coral symbionts, though the location and frequency remain unknown.

The adaptive potential of sex is widely acknowledged. However, the inference of sex in *Symbiodinium* also raises the potential for hybridisation: recombination between genetically divergent lineages. This can lead to rapid exchange of genetic material between lineages, and facilitate the adaptation of new traits at much greater speed than intra-species sex (Willis *et al.* 2006). This clearly has potential for possibly producing thermally resistant symbionts at a pace demanded by the progress of climate change. In addition, this could allow for the consistent establishment of beneficial associations. Given the possible limitations on flexibility (see Section 1.4.1.3), corals may be more likely to take up thermally resistant types if they are produced *via* hybridisation with one of their own symbionts, as many of the recognition genes and molecules will be the same. A similar biological principle has been observed in plants (Moran and Alexander 2014).

The observation that corals frequently hybridise (*e.g.* Willis *et al.* 2006; Combosch and Vollmer 2015), in addition to the high prevalence of background symbiont populations, allows for diverse symbiont consortia to come into contact, so the potential for symbiont hybridisation certainly exists (Santos 2014). There is indirect evidence of hybridisation in *Symbiodinium*, such as incongruence between different molecular markers (LaJeunesse 2001). There are also several examples from field studies where symbiont populations were observed with two sets of co-dominant ITS2 repeats, which corresponded to two already well-defined types (*e.g.* LaJeunesse 2005; LaJeunesse *et al.* 2004a, 2005). Incomplete lineage sorting (ILS) or hybridisation are two parsimonious explanations. The most sustained example comes from Wilkinson *et al.* (2015), who showed two symbiont types but three distinct symbiont populations inside a single *Pocillopora* colony: C100 symbionts, C109 symbionts, and symbionts having co-dominant C100 and C109 repeats in the same cell. The extensive presence of the two 'pure' populations renders ILS far less likely, and points towards symbiont hybridisation. In addition, the distribution of these putative hybrids inside host colonies was related to a temperature

gradient (Wilkinson *et al.* 2016), suggesting temperature-dependent differences between the hybrids and parent populations. However, these studies took place at Lord Howe Island, the world's most southern coral reef system, and it is hence not clear whether hybridisation occurs outside of this atypical site.

Given the potential for rapid adaptation and this nascent evidence of existence, further study should be devoted to the potential of symbiont hybridisation.

1.5. Aims and objectives of the study

Given the importance of understanding symbiont diversity and potential mechanisms of how it may facilitate successful reefs, the aim of this study is to assess patterns of symbiont diversity and rapid adaptive potential at a prospering coral reef. The study site chosen was Atauro Island, Timor-Leste. Sitting in the Coral Triangle, conclusions drawn from this reef system will be transferable to other reefs in the Indo-Pacific and Caribbean. However, it is also unique for its diversity, being recently named by Conservation International as the most diverse reef in terms of fish in the world, as well as demonstrating remarkable coral diversity (Turak and DeVantier 2013). Therefore, this study may bring new understanding to the field of coral-algal symbiosis, as well as highlight what the symbiont communities of a successful reef look like. This aim of the study will be met by addressing the following two objectives:

1) Conduct a diversity assessment of coral symbionts at the reef systems of Atauro Island and mainland Timor-Leste.

H_0: All sites at Atauro Island and on nearby mainland Timor will demonstrate similar symbiont diversity and composition to each other, and to other reef sites globally.

H_1: The study sites will have higher symbiotic diversity than previously studied coral reef sites.

2) Establish evidence, or lack thereof, for symbiont hybridisation at Atauro Island.

H_0: Phylogenetic patterns of evolution will be congruent between different gene regions.

H_1: Symbionts will show evidence of symbiont hybridisation *via* incongruences between different gene regions.

Objective 1 will be addressed in Chapter 2, while Objective 2 will be addressed in Chapter 3. These two chapters are structured as independent manuscripts, therefore there may be some overlap of content in certain sections. A final chapter (Chapter 4) integrates the results of the two data chapters to draw conclusions on the adaptive potential of *Symbiodinium*.

Chapter 2: High symbiont diversity and differentiation at Atauro Island and mainland Timor-Leste

2.1. Introduction

Biodiversity is recognised as a crucial aspect of the global biota. It is explicitly linked to a variety of success measures, such as ecosystem resilience (Fischer *et al.* 2006), and provides a direct benefit to humans *via* the multitude of ecosystem services it facilitates (Benayas *et al.* 2009; Díaz *et al.* 2015). A fundamental aspect of biodiversity is genetic diversity, recognised by the IUCN as one of the three types of diversity deserving of conservation (Reed and Frankham 2003). Genetic diversity is positively correlated with fitness of populations and species, because it provides the raw material on which selection can act and hence facilitate adaptation in the face of changing conditions (Reed and Frankham 2003).

This is exemplified by reef-building corals, which provide benefits to a wide range of taxa, including humans (Pratchett *et al.* 2008; Cinner 2014; Coker *et al.* 2014). This is largely thanks to their symbionts, single-celled dinoflagellates of the genus *Symbiodinium*, which provide photosynthate to support coral metabolism in tropical waters lacking nutrients. Early genetic studies on these symbionts revealed remarkable diversity (*e.g.* Baillie *et al.* 1998; LaJeunesse 2001). Now, there are nine major clades recognised, designated A-I (Pochon and Gates 2010). Further, these genetic differences have been convincingly shown to convey differential benefits to their coral hosts in the face of variable environmental factors, such as temperature, light, and pH level (Rowan and Knowlton 1995; Yamashita *et al.* 2013; Cunning *et al.* 2015a; Kemp *et al.* 2015; Rouzé *et al.* 2016). This highlights the potential value of symbiont diversity, something further accentuated by coral transmission mode. Generally, juveniles of horizontally transmitting corals (*i.e.* those that obtain symbionts from the environment in each successive generation) host a diverse range of symbionts, before intracellular sorting processes result in adult colonies hosting one major type, with others at background levels (Coffroth *et al.* 2001; Poland *et al.* 2013; Yamashita *et al.* 2013). This potentially allows the coral to select the most advantageous symbiont for the ambient conditions from a diverse candidate pool. In addition, some corals have the ability to gain symbionts both from the environment *and* their parent, in contrast to most corals which only use one of these methods (Byler *et al.* 2013). This gives an opportunity to associate with symbionts that have clearly been successful (inherited from the parents) while still facilitating potential adjustment to changes in the environmental regime (by taking up novel symbionts). Taken together, these results indicate that coral fitness and environmental adaptation are inextricably linked with symbiont diversity.

In the face of climate change, therefore, symbiont diversity has been proposed as a mechanism for continued coral survival into the future (Baker 2001; Hume *et al.* 2015). If this is to be the case, we require first a high-level understanding of symbiont diversity, and then action to protect that diversity (Baskett *et al.* 2009; Kemp *et al.* 2015; Hume *et al.* 2016). However, there is still a long way to go before even the first step in this process is achieved. In general, a thorough understanding of symbiont diversity, and therefore its hypothesised contribution to coral persistence, faces two broad issues.

The first issue is procedural. Since symbiont diversity began to be recognised, an ecological viewpoint dominated, with fundamental cellular questions left behind (Davy *et al.* 2012), including a thorough understanding of this diversity. Theories such as the 'adaptive bleaching hypothesis' (Buddemeier and Fautin 1993) stimulated over a decade of intensive studies on 'switching' and 'shuffling', whereby corals can take up new beneficial symbionts or stimulate a latent background type to increase in density, respectively. Such studies displayed an incomplete understanding of symbiont diversity, and generated misunderstandings in the field. For example, see Stat and colleague's (2006) questioning of Lewis and Coffroth (2004), or the communications between Baker (2001, 2002) and Hoegh-Guldberg and colleagues (2002). Updates to symbiont taxonomy also call into question some findings from previous studies. Gibbin and Davy (2013), for example, draw conclusions based on the separation of clade E symbionts into subclades E1 and E2, when it has since been argued they are very likely the same thing (Jeong *et al.* 2014). In contrast, other studies have suggested that populations previously interpreted as one species (*e.g.* Howells *et al.* 2013) were actually multiple co-existing types (Wham and LaJeunesse 2016), though note that this was then challenged by Howells *et al.* (2016a).

Issues may have arisen thanks to an excessively rigorous application of the phylogenetic species concept. Recent taxonomy of *Symbiodinium* has been largely based on the internal transcribed spacer region (ITS2) of the nuclear ribosome-encoding DNA, which occurs in repeated arrays in the genome. This marker, along with other recently-developed markers such as the chloroplast minicircle psbA[ncr] region, is considered to give 'species-level' resolution (LaJeunesse and Thornhill 2011; Pochon *et al.* 2012). In contrast, less variable markers like the mitochondrial *cob* gene may provide sub-clade resolution only (LaJeunesse *et al.* 2012). *Symbiodinium* with variable ITS2 sequences have generally been considered as different species, even though it is likely that much of this represents intraspecific variation in a population, or intragenomic variation in a single symbiont cell (Correa and Baker 2009; Figure 1.4). It has been suggested that some sort of genetic clustering method should be applied to reduce spurious allocation of species (Correa and Baker 2009; Arif *et al.* 2014). In addition, there has been a recent effort to include ecological considerations in concordance with genetic methods (*e.g.* Jeong *et al.* 2014; Hume *et al.* 2015; Parkinson *et al.* 2015), in an attempt to standardise *Symbiodinium* systematics, or at least increase comparability across studies. However, the field is still plagued by the fact that what constitutes a species is still debated and largely undefined (Parkinson and Baums 2014;

Howells *et al.* 2016a). Therefore, further studies on intra-clade diversity, using a variety of genetic markers, are still a necessity for understanding coral-algal interactions.

The second issue is ecological. Unfortunately, understanding symbiont diversity at a global level may not be helpful, as many coral-algal relationships appear to be specific to certain regions. Symbionts show limited long-distance dispersal ability (Grupstra *et al.* 2017), with evidence of highly restricted gene flow within both the Caribbean and Pacific for example (Howells *et al.* 2009; Baums *et al.* 2014). The same corals can host different symbionts in different locations (Baird *et al.* 2007; Oliver and Palumbi 2011), indicating that coral-symbiont relationships need to be understood at a local level. Indeed, many coral communities host symbiont consortia distinct to that particular region (Fabina *et al.* 2012; Bongaerts *et al.* 2015b; Stat *et al.* 2015); this can be due to factors such as specialised symbiont types requiring certain conditions (D'Angelo *et al.* 2015). Therefore, conclusions drawn from these relationships often cannot be applied to other ecoregions. As such, simple diversity studies from different regions are continually finding novel relationships (*e.g.* Rouzé *et al.* 2017), many of which could have functional relevance in certain conditions (Howells *et al.* 2016b). This ultimately aids our understanding of how reef communities may survive into the future. This suggests diversity studies should be conducted on a location-by-location basis, but such studies remain sparse (Santos 2014).

This study aims to address the above two problems, by using multiple genetic methods consistent with previous studies to explore the symbiont diversity of a previously uncharacterised reef system. The reefs of Timor-Leste are exceptionally diverse, with ~400 coral species identified in a recent survey (Turak and DeVantier 2013). Atauro Island, to the north of mainland Timor across the Ombai Strait, was recently named as the most piscine-diverse reef system in the world (Conservation International, August 2016). As a deep and swiftly-moving body of water (Sprintall *et al.* 2004), the Ombai Strait may generate ecological distinctions between the two regions, meaning that they should not be treated as homogeneous for the purposes of conservation management. Such considerations are increasingly relevant, with the designation of a broad marine park in the waters of Timor-Leste (Nino Konis Santana National Park), as well as many locally managed marine protected areas, including at least four on Atauro Island. Effective management relies on an understanding of the diversity housed by the various protected areas. However, the waters of Timor-Leste are severely understudied, thanks to the violent Indonesian occupation of 1975-1999 (Silove *et al.* 2014) and social crisis of 2006-2007 (Scambary 2009). Despite its highly successful reefs, only two published studies of coral diversity have been conducted there (Ayling *et al.* 2009; Turak and DeVantier 2013), and the symbiont communities have never been assessed. Given the rich coral communities of Atauro Island and Timor, and the unprecedented fish diversity at Atauro Island, a study of the symbionts may bring new light to the links between *Symbiodinium* diversity and reef health, and whether symbiont diversity reflects the diversity in other reef taxa.

This study tested two hypotheses. The first hypothesis was that Atauro Island and mainland Timor have a similar symbiont community diversity and composition. The second was that Atauro Island and/or Timor have a greater diversity of *Symbiodinium* than other previously studied coral reef sites. These hypotheses were tested using three gene regions: the *cob* gene of the symbiont mitochondrial DNA, the ITS2 region of the symbiont nuclear ribosomal DNA and the psbA^{ncr} region of the symbiont chloroplast DNA. In addition, Next Generation Sequencing (NGS) of the ITS2 region was used to assess the presence and identity of background symbiont populations. This multi-gene approach will be useful for future studies, as they will be able to compare with this study even if they choose one gene region to analyse. Overall, this chapter contributes to an understanding of global coral-algal symbiosis patterns. In addition, it may potentially provide further information on the future fate of coral reefs, by quantifying the symbiont communities at a highly biodiverse reef site.

2.2. Methods

2.2.1. Study Location

Four sites were sampled at Atauro Island (Timor-Leste), 25 kilometres north of the island of Timor (Figure 2.1). Three sites were sampled along the northern coast of Timor-Leste, to compare symbiont diversity at these reefs with the reefs at Atauro Island. The seven sites chosen were based on sites sampled in a previous study of coral diversity (Turak and DeVantier 2013). This allowed a comparison of symbiont diversity with coral diversity. In addition, it facilitated ease of locating study sites, as it had previously been determined that there were reefs present at those sites. For details of all sites, see Table 2.1.

Table 2.1: Details of sampling locations, including names and abbreviations that will be used throughout the thesis.

Site name[a]	Geographic Coordinates	Region	Province/village	Abbreviation
Beloi Lagoon South	8° 14.547S, 125° 36.555E	Atauro	Beloi	BLS
Beloi Harbour	8° 13.282S, 125° 36.885E	Atauro	Beloi	BHB
Beloi Saddle Patch (Right)	8° 12.933S, 125° 37.58E	Atauro	Beloi	BSP
Beloi Barrier Reef	8° 13.603S, 125° 36.802E	Atauro	Beloi	BBR
Hera West	8° 31.183S, 125° 38.664E	Timor	Hera	HEW
Lamsana Inlet West	8° 30.826S, 126° 4.157E	Timor	Manatuto	LIW
Lamsana Inlet East	8° 30.977S, 126° 4.388E	Timor	Manatuto	LIE

[a]Following Turak and DeVantier (2013), though note they incorrectly refer to 'Beloi' as 'Belio.'
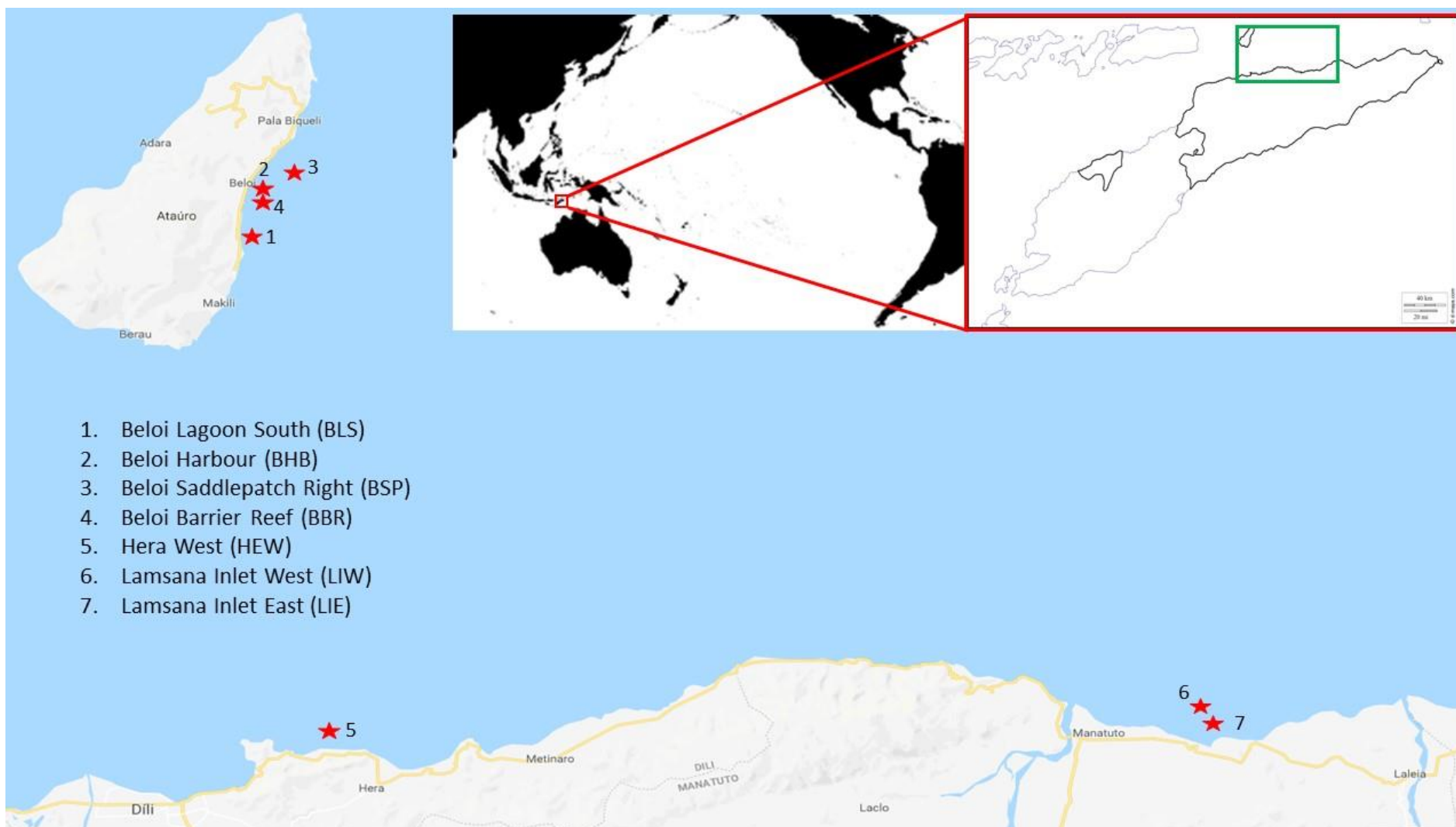
Figure 2.1: Map of northern Timor and Atauro Island, showing sampling locations (red stars).

1. Beloi Lagoon South (BLS)
2. Beloi Harbour (BHB)
3. Beloi Saddlepatch Right (BSP)
4. Beloi Barrier Reef (BBR)
5. Hera West (HEW)
6. Lamsana Inlet West (LIW)
7. Lamsana Inlet East (LIE)

Sampling took place in late April 2017, except Hera West, which was sampled in December 2016. Given that there have been no major stress events observed at these locations, and sea surface temperatures are nearly identical in the Banda Sea in December and May (Gordon and Susanto 2001), it is considered unlikely that this temporal difference in sampling time would have significantly affected the results.

Explicit environmental measures were not taken at the sites (as each site could only be visited once, on separate days, comparisons in any environmental differences found would be apocryphal). However, it was noted that the three Timor sites were highly turbid, in contrast to the four Atauro sites. This correlates strongly with the findings of Turak and DeVantier (2013), who recorded visibility readings of 5 m at Hera West and 4-8 m at Lamsana Inlet, in contrast to 20-25 m at all Beloi sites. This indicates that such conditions are likely the norm for these sites. Atauro Island also has no rivers, while chronic river plumes are evident off the coast of Timor. In addition, these authors found minor differences in water temperatures between the sites (which varied from 25ºC - 28ºC).

2.2.2. Sampling procedure

The reef flat of each site was sampled *via* snorkel at depths of 1-4 m. Samples were collected under permit number LNC-PC0012.VI.16 (Ministerio da Agricultura e Pescas, Timor-Leste). As the goal was to sample from a wide range of coral species, a roving search technique was utilised. This involved a surface search to identify suitable coral colonies. Once one was identified, a small piece (~1 cm^3) was removed using hand pliers (occasionally for massive corals, a small chisel was also utilised). For later identification purposes, coral photos were taken with an Olympus C5060 camera at a distance of ~20 cm, to facilitate identification *via* polyp characteristics in addition to colony shape. This procedure was repeated 100 times at each site, and a final search was conducted to ensure that no major coral species had been missed in the sampling. The exception was Hera West, where highly turbid conditions limited collection to 80 samples. Coral samples were placed in 1.7 mL Eppendorf tubes, preserved in DMSO buffer (20% DMSO, 250 mM EDTA, NaCl saturated, pH 8.0) and stored at -20ºC until DNA extraction. 650 samples from different coral colonies were collected across the seven sites. In total, 43 scleractinian genera were represented in the final analyses (see Table A1.4 in Appendix A1.7). In order to understand symbiont diversity of all reef-building cnidarians, six non-scleractinian genera were also sampled (one to seven samples *per* genus).

2.2.3. DNA extraction

The tubes containing samples were vortexed for 5 seconds to agitate the sample and suspend the *Symbiodinium* cells. A 100 µL aliquot of supernatant was added to a new tube, and centrifuged for 6

min (all centrifugation occurred at 12800 x g). The supernatant was removed and discarded, and 100 µL of guanidine solution (4 M Guanidinium isothiocyanate, 50 mM Tris-HCl, 10 mM EDTA, Sarkosyl 2% w/v, β-mercaptoethanol 1% v/v) were added to the algal pellet, which was then vortexed briefly and left at room temperature for three days for cell lysis. Following this incubation, samples were vortexed briefly and centrifuged for 5 min. A 50 µL aliquot of supernatant was removed and added to a new tube containing 50 µL cold isopropanol. These tubes were centrifuged for 10 min and the supernatant removed, 200 µL of 70% EtOH were added, and the samples were vortexed briefly and centrifuged for 5 min. The supernatant was removed and the DNA pellets dried at room temperature with the tube lids open for ~1 h. Samples were stored at room temperature, and transported to New Zealand as dried DNA.

Upon arrival in New Zealand, 50 µL of TE buffer (10 mM Tris-HCl, 1 mM EDTA) were added and left at room temperature for 10 min to re-hydrate the DNA. Samples were vortexed and centrifuged briefly at 1500 x g, and then stored at -20ºC until DNA amplification *via* PCR.

2.2.4. Amplification and sequencing of DNA

For all *cob* and psbA^{ncr} amplifications, PCR reactions consisted of 10 µL MyTaq HS Red Mix 2x (BioLine), ~20 ng sample DNA, 10 µg BSA, 0.25 µM of each primer and 6 µL PCRH$_2$O for a total volume of 20 µL. Amplicons were purified with MagNA solution (Rohland and Reich 2012, Appendix A1.2) and sequenced in the forward direction by Macrogen Inc. (South Korea).

2.2.4.1. Amplification of mitochondrial cob *gene*

Amplification was first carried out on the mitochondrial *cob* gene. Due to issues with the generality of primers and sample quality, two sets of primers were used to amplify the *cob* gene. Optimisation was carried out throughout the amplifications, so the cycling conditions reported are the final optimised conditions, but a number of samples were successfully amplified and sequenced using other cycling conditions (see Appendix A1.1.1).

A total of 252 samples were successfully amplified with the primer pair Dinocob1F (5'-ATG AAA TCT CAT TTA CAW WCA TAT CCT TGT CC-3') and Dinocob1R (5'-TCT CTT GAG GKA ATT GWK MAC CTA TCC A-3') (Zhang *et al.* 2005). The final optimised cycling conditions were a denaturation for 1 min at 95ºC, followed by 40 cycles of 15 s at 95ºC, 20 s at 55ºC, 30 s at 72ºC, and a final extension for 7 min at 72ºC.

A total of 304 samples were successfully amplified with the primer pair Cob_f1 (5'-TGA AAT CTC ATT TAC AAT CAT ATC CTT-3') and Cob_r1 (5'-CTA CAG GAA ATT GAC CAC CTA TCC-3')

(Pochon *et al.* 2012). The final optimised cycling conditions were a denaturation for 2 min at 95ºC, followed by 40 cycles of 20 s at 95ºC, 20 s at 58ºC, 30 s at 72ºC, and a final extension for 7 min at 72ºC. This primer pair amplified the same region as the primer pair above minus 3 bp at the start and 4 bp at the end of the sequence; these differences were trimmed during sequence processing.

All PCR runs included a positive and negative control. PCRs were carried out using an Applied Biosystems Veriti 96-well thermal cycler (Applied Biosystems). Successful amplifications were identified by running 4 µL of the samples on a 1.5% agarose gel at 100 V for 35 min using an Enduro 250 V Electrophoresis PowerPack (Labnet), before being visualised using a UVP Benchtop UV Transilluminator (302 nm UV light, Biostrategy). Samples were run against HyperLadder 50 bp (BioLine) to confirm the correct amplicon length. Absence of a clear band or excessive streaking of the band led to rejection of the amplification. In all cases, the negative control lacked a DNA band indicating that the PCR was free of contamination. The remaining 16 µL of successful PCRs were stored at -20ºC until purification.

In total, 556 of the 650 samples were successfully amplified (86%), with sites individually also averaging around this success rate. A failure rate of 14% could be explained by the lack of a thorough washing step of the original coral nubbins; this meant there may have been coral mucus still present, which can inhibit PCR reactions (Fredricks *et al.* 2005; Qvarnstrom *et al.* 2007). The exception was Beloi Harbour (59/95, 62%); it is unclear why the success rate was much lower for this site. Because of this, and due to financial limitations, not all successful amplifications were sequenced. Between 59 and 75 samples were chosen from each site, based on coral ID as assessed by the photographs. Samples were chosen with the aim of being comprehensive and reflective of the diversity at the sites; excluded samples came from highly common species or genera that were already included in the sequencing analysis, and so were unlikely to add additional diversity information. In total, 441 samples across the seven sites were chosen for sequencing. The resulting sequences underwent preliminary analysis, as the psbA^{ncr} region requires clade-specific primers (LaJeunesse and Thornhill 2011). Sequences were trimmed and identified to clade level using an automated classification tree (`insect` R package, github.com/shaunpwilkinson/insect).

2.2.4.2. Amplification of chloroplast minicircle psbA^{ncr}

Samples belonging to clade C were amplified with the primer pair 7.4-Forw (5'-GCA TGA AAG AAA TGC ACA CAA CTT CCC-3') and 7.8-Rev (5'-GGT TCT CTT ATT CCA TCA ATA TCT ACT G-3') (Moore *et al.* 2003). The final optimised cycling conditions were denaturation for 3 min at 95ºC, followed by 38 cycles of 20 s at 95ºC, 30 s at 57ºC, 30 s at 72ºC, and a final extension for 7 min at 72ºC (see Appendix A1.1.2 for details of other trial cycling conditions). PCRs were carried out and assessed

as *per* the *cob* gene. No amplification was evident for any of the negative controls, indicating that the runs were free of contaminant DNA. A total of 379 clade C samples were amplified in this fashion. In addition, 11 samples were observed to have multiple bands when run on the gel, potentially indicating co-dominant symbiont populations or intragenomic variants. An aliquot (8 µL) of these samples was re-run on a 1% gel at 90 V for 1.5 h. Individual bands were excised under UV light using a pipette tip, and re-amplified using 20 cycles of the above conditions.

Difficulties were encountered amplifying clade D psbA^{ncr} samples, often due to multiple bands present in a single sample (see Appendix A1.3 for details). The final approach involved amplification using the primers psbAFor_1 (5'-GCA GCT CAT GGT TAT TTT GGT AGA C-3') and psbARev_1 (5'-AAT TCC CAT TCT CTA CCC ATC C-3') (LaJeunesse and Thornhill 2011), with cycling conditions of 3 min at 95ºC, followed by 38 cycles of 20 s at 95ºC, 30 s at 56ºC, 1 min at 72ºC, and a final extension for 7 min at 72ºC. Samples were assessed and individual bands excised using the method described for the clade C samples. Individual bands were re-amplified using the novel primers SYMPSBANCRF (5'-AAT CGT GCT GAT CTA GGW ATG G-3') and SYMPSBANCRR (5'-GAG ACG ATT TGT TGT GGA TAG-3'), using 25 cycles of the above conditions. A 4 µL subsample of the resulting amplicons was re-run on a 1.5% agarose gel at 100 V for 35 min, to confirm the presence of a single band that could be sequenced. In this fashion, 33 of the 45 clade D samples had at least one band amplified for sequencing.

2.2.4.3. Next Generation Sequencing (NGS) of the nuclear ITS2 region

DNA concentration of samples was obtained using a NanoPhotometer (Implen), and samples were diluted down to a standard 5 ng/µL with TE buffer. A total of 383 samples plus one negative control were chosen for amplification, as this was considered the optimal trade-off between sequencing depth and the number of samples. Samples chosen were a subset of the successful *cob* gene samples, with the aim of again incorporating as much diversity as possible from the original sampling effort.

Samples were initially amplified with the primers ITSD (Pochon *et al.* 2001) and ITS2Rev2 (Stat *et al.* 2009), with Illumina adapters attached. The full primer sequences were: 5'-TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG GTG AAT TGC AGA ACT CCG TG-3' (forward); and 5'-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACAG CCT CCG CTT ACT TAT ATG CTT-3' (reverse). Underlines indicate the original primers. Reactions consisted of 12.5 µL SuperFi High-Fidelity MasterMix 2x (Invitrogen), 10 ng sample DNA, 10 µg BSA, 0.2 µM of each primer and 8.5 µL PCRH$_2$O for a total volume of 25 µL. Cycling conditions were a denaturation for 3 min at 95ºC, followed by 25 cycles of 30 s at 95ºC, 30 s at 55ºC, 30 s at 72ºC, and a final extension for 5 min at 72ºC. Amplicons were purified and stored at -20ºC.

A second PCR run was conducted to add unique indexing primers to each amplicon. Sixteen forward and 24 reverse indexing primers (IDT) were used, allowing for 384 unique combinations (see Appendix A1.4 for full primer list). Reactions consisted of 10 µL SuperFi MasterMix, 2 µL of PCR product from the initial amplification, 0.1 µM of each primer and 4 µL PCRH$_2$O for a total volume of 20 µL. The indexing runs used 8 cycles of the above conditions, and the resulting amplicons were purified.

Amplifications were quantified and checked for quality using qPCR (Applied Biosystems StepOne instrument). Reactions consisted of 5 µL SYBR Green Mix (BioLine), 0.5 µM of each of the primers ITSD/ITS2Rev2, 20 µg BSA, 2 µL of the indexed samples (diluted 1:1000) and 1.5 µL PCRH$_2$O for a total volume of 10 µL. Samples were run in duplicate. Cycling conditions were a denaturation for 10 min at 95ºC, followed by 40 cycles of 15 s at 95ºC and 1 min at 60ºC. C_T (cycle threshold) values were generated *via* a calculation (using an in-built machine algorithm) of when the DNA signal differed significantly from background fluorescence, using a baseline of 0.1. A melt curve was run (temperature elevation from 60°C to 95°C in 0.3°C increments each of 15 s duration) to ensure that only the target sequences were amplified. Samples were considered successful if duplicates had a standard deviation in C_T values <0.5 and melted within 1ºC of the average melting temperature of the target region (~82ºC). In total, 23 samples initially failed these requirements, generally due to the extensive presence of primer-dimers (evidenced by large peaks on the melt curve at <75ºC). These samples were re-done from the beginning, and 20 were subsequently confirmed to be successful, leaving 381 samples. Importantly, the negative control (which had been carried through both amplification and purification procedures) only showed the presence of primer-dimers, indicating that contamination had been avoided in the preparation of samples.

DNA concentrations for all samples were calculated from C_T values using a standard curve which was run using previously prepared clonal serial dilutions of known concentration (Wilkinson *et al.* 2015). Following quantification, different volumes of the final purified samples were pooled in a single tube to achieve equal concentration of all samples, with the concentration of the final pooled library being 4 nM DNA. (Note that due to lower amplification success of some samples, in occasional cases samples were present in lower relative concentrations in the final library). The pooled library was sequenced on a single lane of the Illumina MiSeq platform by the Centre for Genomics and Proteomics, University of Auckland, New Zealand.

2.2.4.4. Amplification of coral host ITS2 region

The coral host ITS2 region was sequenced, excluding samples belonging to the genus *Acropora* (as identified by the photographs). This was for two reasons: (a) The ITS2 region of *Acropora* is

problematic to amplify and sequence, due to an abnormally high amount of intragenomic heterogeneity, potentially caused by hybridisation, incomplete lineage sorting or pseudogene development (Wei *et al.* 2006). (b) *Acropora* is a highly distinctive genus and could be unambiguously identified *via* photographs alone, in contrast to other genera where molecular methods are required to confidently confirm identity. The host ITS2 region is reasonably data-deficient in GenBank, leading to some ambiguities in the identification of host species; however, typical animal markers (mainly mitochondrial ones such as the *cob* and cox1 genes) evolve unusually slowly in anthozoans and are therefore inappropriate to use (van Oppen *et al.* 2002). The ITS2 region provides a greater level of resolution (van Oppen *et al.* 2002; Forsman *et al.* 2006), hence its use here.

In total, 322 samples were successfully amplified with the primer pair SCLERforw (5'-GAR TCT TTG AAC GCA AAT GGC -3') and SCLERrev (5'-GCT TAT TAA TAT GCT TAA ATT CAG CG-3') (S. Wilkinson, unpublished), for a 100% success rate. All samples had the cycling conditions of 3 min at 95ºC, followed by 40 cycles of 15 s at 95ºC, 15 s at 56ºC, and 20 s at 72ºC. PCRs were carried out and assessed as *per* the clade C psbA^{ncr}. In addition, six samples were observed to have two bands when run on a gel, potentially indicating hybrid coral hosts, or accidental contamination (though all negative controls had no bands). An 8 µL sub-sample of these samples was re-run on a 1% gel at 90 V for 1.5 h. Individual bands were excised under UV light using a pipette tip, and re-amplified using 20 cycles of the above conditions. Initial inspection revealed that 111 sequences were of dubious quality, so these samples were also sequenced in the reverse direction.

2.2.5 Data processing

All direct sequencing products (*cob*, psbA^{ncr} and host ITS2) had low-quality ends trimmed and were aligned using the `aphid` R package (S. Wilkinson, cran.r-project.org/web/packages/aphid/index.html). All further data manipulation of direct sequencing products used Geneious v8.0.5 (Biomatters). For the coral host samples that were sequenced in both directions, forward and reverse sequences were aligned using free-end gaps, with the forward primer and a conserved part of the 28S region near the end of the sequence. Bases were edited manually, and the consensus sequence was extracted. In all cases, sequences were aligned using the Geneious global alignment tool with free end gaps, using a 65% similarity cost matrix. Sequences were identified by using the BLAST function in the NCBI database (*cob* gene and host ITS2 region). To verify host IDs, photos were also analysed and the resulting identifications cross-referenced with the genetic results to gain a consensus ID. The major text consulted was Veron and Stafford-Smith (2000). Because the psbA^{ncr} region is highly variable (over 120 unique reads from ~220 samples that has successful sequences), it was impractical – and likely ecologically spurious – to treat them all individually, so sequences were manually clustered according to a 95% similarity criterion in Geneious. This left 22 distinct groups, as well as several unique sequences that

could not be clustered and were left as individual sequences. The strong correspondence with distinct ITS2 types (see Results) supported this approach. The consensus sequence for each group was extracted, and identified largely using the repository of psbA[ncr] sequences found in Thornhill *et al.* (2014). The taxonomic state of these groups is uncertain; they may represent distinct species or even genera, given the uncertain nature of taxonomy within *Symbiodinium*. For all gene regions, rare sequences were checked manually in Geneious to verify that the divergent nucleotide sites had high-quality base calls.

The NGS ITS2 sequences were processed in R (R Core Team, 2017). Forward and reverse reads were assembled and trimmed for each sample (primer sequences removed), and only sequences that were found at an abundance of >1% in at least one sample were kept. Sequences <150 bp were discarded. Following this, an Operational Taxonomic Unit (OTU) approach was taken, to avoid analysing intragenomic variants (Arif *et al.* 2014). A genetic cut-off distance of 0.03 was used, and all sequences that differed by less than this amount were treated as intragenomic variation and collapsed into a single OTU. The most common sequence in each OTU was used to identify the OTU to clade level (following Arif *et al.* 2014, Ziegler *et al.* 2017a). In addition, the most dominant sequence overall in each sample was identified as far as possible using a specific ITS2 database (GeoSymbio, Franklin *et al.* 2012). Four OTUs were identified as non-*Symbiodinium* and removed. The negative control showed very low presence of one OTU at a frequency more than 3000 times less than the average sequence number of other samples. Two other samples were indistinguishable from the negative control, and subsequently removed.

In addition, a selection of the NGS samples were analysed using the program `mothur` (Schloss *et al.* 2009), following the MiSeq Standard Operating Procedure (Kozich *et al.* 2013), using the `chimera.vsearch` command to identify any PCR chimeras.

2.2.6. Verification of symbiont identity

Six samples had their dominant sequences identified as different clades by the different gene regions, a highly unexpected result. Five of these were clade C for the *cob* and psbA[ncr] gene regions but clade D for the ITS2 region, while one was clade B for the *cob* region but clade C for the psbA[ncr] and ITS2 gene regions. To verify these results, further qPCRs were run using the clade-specific primers SymB28S-1F (5′-CAC ATG TCG TGC TGA GAT TGC-3′) and SymB28S-1R (5′-CTC GCA TGC TGA GAA ACA CTG-3); SymC28S-1F (5′-TTG CTG AGA TTG CTG TAG GCT-3′) and SymC28S-1R (5′-TCC TCA AAC AGG TGT GGC-3′); and SymD28S-1F (5′-AAT GCT TGT GAG CCC TGG TC-3′) and SymD28S-1R (5′-AAG GCA ATC CTC ATG CGT ATG-3′) (Yamashita *et al.* 2011). The C/D samples were run in independent reactions with the SymC and SymD primers, while the B/C sample was run in independent reactions with the SymB and SymC primers. Reactions consisted of 5 µL SYBR Green

Mix (BioLine), 0.5 µM of the relevant primers, 20 µg BSA, 2 µL of the samples and 1.5 µL PCRH$_2$O for a total volume of 10 µL. Samples were run in triplicate. Cycling conditions were a denaturation for 10 min at 95ºC, followed by 40 cycles of 15 s at 95ºC and 1 min at 60ºC.

Results for the C/D samples revealed nearly identical amplification curves using both the SymC and SymD primers, indicating that both clades were present at similar levels. Hence, the incongruence between gene regions was attributed to preferential amplification of a certain clade, and these samples were maintained in the analysis. In contrast, the B/C sample showed amplification using the SymC but not the SymB primers. This suggests that the initial clade B result from the *cob* gene was contamination, and this sample was removed from all further analyses in this chapter.

2.2.7. Data analysis

Unless otherwise specified, all data analyses used R v.3.4.3 (R Core Team, 2017). To test the first hypothesis (*i.e.* Atauro and Timor sites have similar symbiont diversity and composition), a rarefaction analysis was conducted using the `vegan` package (Oksanen *et al.* 2017) to estimate the number of unique sequences *per* 50 samples. Using these results, overall richness of each gene region was compared between Atauro Island and Timor using one-way ANOVAs. Assumptions were checked with Levene's test and QQ plots. Also using these rarefied results, the effect of Atauro *vs.* Timor in terms of sequence type composition was analysed with permutational multivariate analysis of variance (PERMANOVA), implemented using the `adonis` function in the `vegan` package using 999 permutations. Replication was achieved *via* five random rarefied subsamples from each site using the `rrarefy` function. Differences between Atauro and Timor sites were also identified genetically using analysis of molecular variance (AMOVA), implemented using 1000 permutations in the program PopArt (Leigh and Bryant 2015). These tests were implemented independently for the *cob* and ITS2 gene regions, but not for the psbA^{ncr} sequences due to the artificial nature of the groups they were placed in. OTU richness was also explored between Atauro and Timor sites. The number of OTUs were compared between Atauro and Timor sites overall, as well as for the five most common coral genera (*Acropora, Porites, Pocillopora, Platygyra* and *Montipora*), using one-way ANOVAs. Assumptions were checked with Levene's test and QQ plots, and non-parametric Kruskall-Wallis tests implemented instead where necessary. The above comparisons were also executed in terms of OTU composition (the number of OTUs in each clade) using χ^2 tests. To account for differing numbers of samples from the different sites, ANOVA calculations used the mean number of OTUs *per* sample (this was not necessary for the χ^2 tests, where the computation of expected values accounts for different numbers of samples among sites).

To test the second hypothesis (*i.e.* Atauro Island and Timor have more diverse symbiont communities than other reefs around the world), ITS2 data from previous studies were collected that reported data in a way that allowed comparison (*i.e.* they reported total number of symbiont types and how many coral genera were sampled). Results from all studies were standardised to symbiont types *per* genera, and the resulting numbers compared. A comparison was also done with psbA^{ncr} sequences, which were also able to be standardised to the number of samples collected (not reported for the ITS2 studies).

2.3. Results

Sections 2.3.1 – 2.3.5 concern the direct sequencing results (including the dominant sequence in each sample from the Next Generation Sequencing), while subsection 2.3.6 describes the OTU-based results for the NGS. Statistical results reported are summaries only; full statistical tables can be found in Appendix A1.6.

2.3.1. Overall results

In total, 436 samples had at least one *Symbiodinium* gene region successfully amplified (*cob*: 422 samples; psbA^{ncr}: 224 samples; ITS2: 377 samples). There was a marked difference in the resolution of the three gene regions (quantified using rarefied numbers of samples for each unique sequence found), with a trend of increasing resolution from the *cob* to ITS2 to psbA^{ncr} gene regions (Figure 2.2). For the *cob* and ITS2 gene regions there were also geographic differences, with sequence diversity 1.25 times higher at Atauro Island than Timor (one-way ANOVA, $p = 0.033$ and $p = 0.049$ respectively).

Generally, the sequences found were well-established types that could be identified with confidence *via* the NCBI or GeoSymbio databases. Table 2.2 summarises the broad groupings found, and affirms Figure 2.2 regarding the general pattern of increasing diversity from the *cob* to psbA^{ncr} regions, with one exception: the ITS2 symbiont type C15 was further split into two subtypes, C15p and C15q, by the *cob* gene. It also shows near-total agreement between the three gene regions as to the identity of the dominant symbiont type in a given sample (with varying degrees of resolution). The one major exception occurred in the C1 radiation, where ITS2 symbiont type C1 commonly occurred with either *cob* type C42a or C1. This was not observed for the psbA^{ncr} region, which conformed to having distinct groups that matched the *Symbiodinium* C42a (C1x, C1v) or C1 (C1u) distinctions according to the *cob* gene. There were also several additional rare examples of incongruence, which were identified and explored by the analyses in Chapter 3. Note also that symbiont type C3 for the ITS2 region was actually found in samples that had the C1/3 sequence for the *cob* gene.
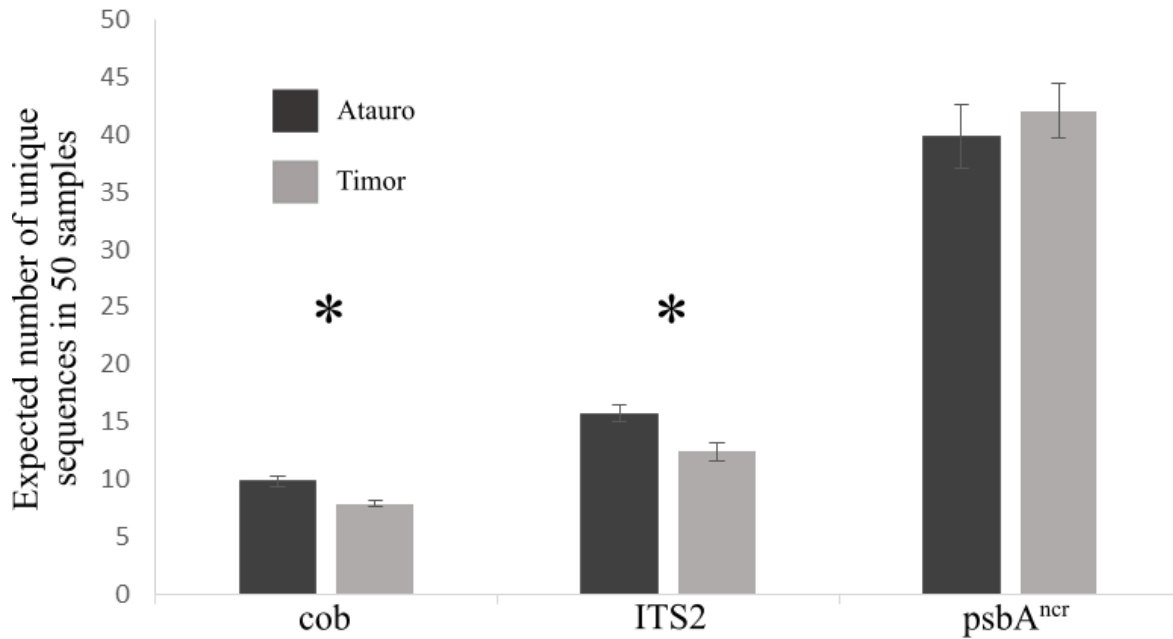
Figure 2.2: Rarefied sequence diversity between gene and geographic regions (mean $\pm$ SE). '*' above the bars indicates a significant difference between Atauro and Timor sites at $\alpha = 0.05$. psbA^{ncr} results are the original unique sequences, not the clustered units.

In addition to the types listed in Table 2.2, there were also unique sequences that were observed as the dominant type in only one or two samples. The *cob* gene had 15 such examples of unique symbiont types, none of which matched with any previously defined sequences. The ITS2 region had 32 examples, of which 10 were previously defined types (see Table A1.3). The psbA^{ncr} region had 24 examples which were <95% similar to any of the broad groupings applied, and did not match any previously defined types. Note that, in many cases, a sample had a unique sequence for more than one gene region.

Table 2.2: Summary of the main symbiont types found in sampling. See Appendix A1.5 for full details and justification behind naming. psbA^{ncr} groups represent collections of sequences with >95% similarity (see Methods). Italics indicate a novel sequence (see Appendix A1.5). This table does not include sequences that only appeared once or twice (see also Table A1.3). '-' in the psbA^{ncr} column indicates that there were no successful sequences for that type. Differing sizes of table cells are only to accommodate the diversity of other gene regions and are not indicative of the number of sequences recovered *per* type.

Broad Group	*cob*	ITS2	psbA^{ncr}
C3 radiation	C1/3	C66	-
		C3	Unique sequences
	C3	C40	*C40c*
			C40d
			C40e
			C40f
			C40g
			C40h
			C40i
			C40j
			C40k
			C40l
			C40m
			C40n
			C40o
		C3d/C21	*C3d/C21p*
			C3d/C21q
		C3z	C3z
	C15p	C15	Unique sequences
	C15q		
C1 radiation	C42a	C1d	C1d
		C1x	*C1x*
		C1 (*Symbiodinium goreaui*)	*C1v*
	C1		*C1u*
	C42p	*C1z*	Unique sequences
Clade D	D1	D1 (*Symbiodinium glynii*)	*D1p*
			D1q
		D5	-
		Unique sequences	*Dva*
			Dvb
			Dvc

2.3.2. *cob* gene analysis

Figure 2.3 shows that *Symbiodinium* C3 was the most common type, with 48% of the sequences, while type C42a dominated the sequences from the C1 radiation. There was a reasonably homogenous distribution of each sequence type among the seven sites, with a few exceptions. Type C1 was found disproportionately at Beloi Saddle Patch, while Timor sites (Hera West and Lamsana Inlet East and West) housed 61% of the clade D samples, despite only representing 44% of samples. In particular, nearly a quarter of all Hera West samples were clade D.

This was reflected statistically. The PERMANOVA analysis revealed that sequence types at the different regions' sites deviated significantly from those expected at random (p = 0.002, Figure A1.2). In addition, the AMOVA results also found a difference between the two geographic regions in terms of genetic distances (p = 0.014).
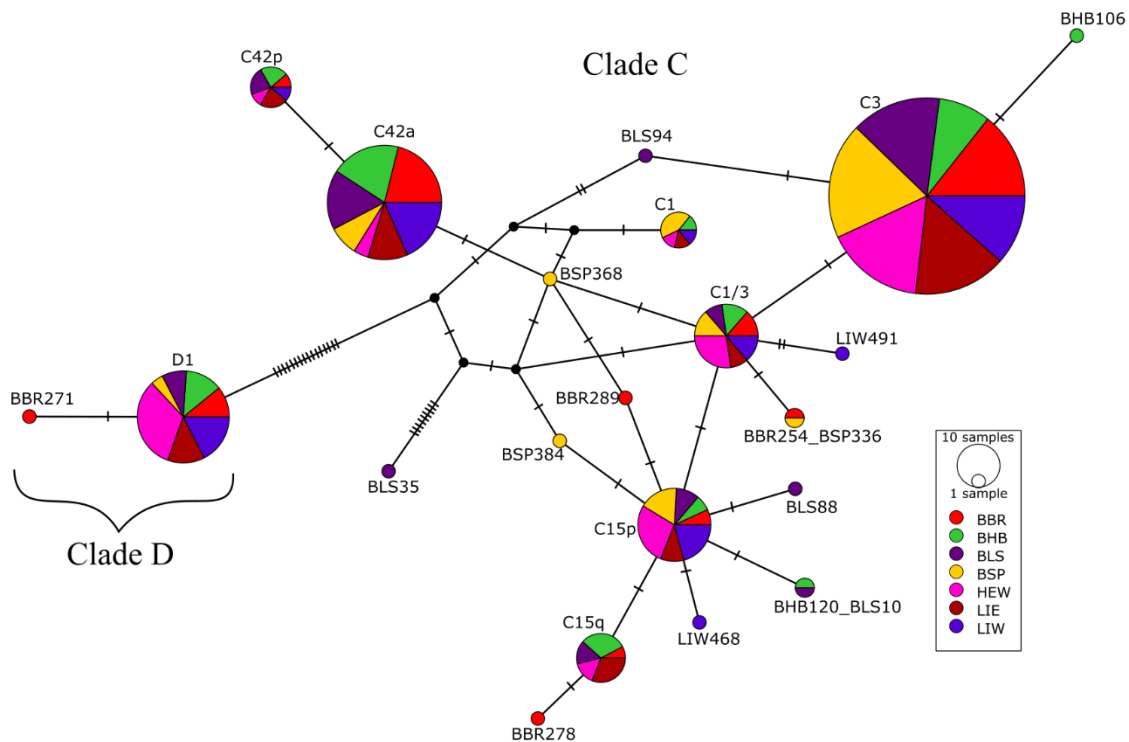
Figure 2.3: TCS Minimum Haplotype Network for the *cob* gene. Hatch marks on the lines represent a single base pair change, while black circles represent hypothesised intermediate haplotypes. The size of the circles is proportional to the number of samples with that sequence type. Line lengths have been optimised for ease of visualisation and are irrelevant. Unique sequences that did not match any previously defined type are labelled with their original sequence label (e.g. BBR278 = site BBR, sample 278)

2.3.3. ITS2 region analysis

As *per* the *cob* gene, the ITS2 network (Figure 2.4) shows a homogenous distribution across sites for most sequence types, with a few exceptions. As before, Hera West is overrepresented in the clade D sequences, both in terms of *Symbiodinium glynii* (D1) and also unique clade D samples. The increased resolution of this region also reveals further interesting patterns. Type C3z is only found at Beloi Saddle Patch and Beloi Harbour, while Beloi Saddle Patch also has a high number of unique sequences from both the C1 and C3 radiations. While it is not evident from the network, there was also a noticeable structure within symbiont types C1 and C1x (depicted as a single circle, see figure legend). Atauro sites accounted for 74% of type C1, while Lamsana Inlet accounted for 57% of type C1x, despite only representing 28% of all sequences. In addition, with the exception of clade D, nearly all unique sequences come from Atauro Island sites.

These results are also supported statistically. The PERMANOVA analysis revealed that sequence type distributions at Atauro Island and Timor deviated significantly from those expected at random (p = 0.001, Figure A1.3). In addition, just like the *cob* gene, the AMOVA results also found a significant difference between the geographic regions (p = 0.037).
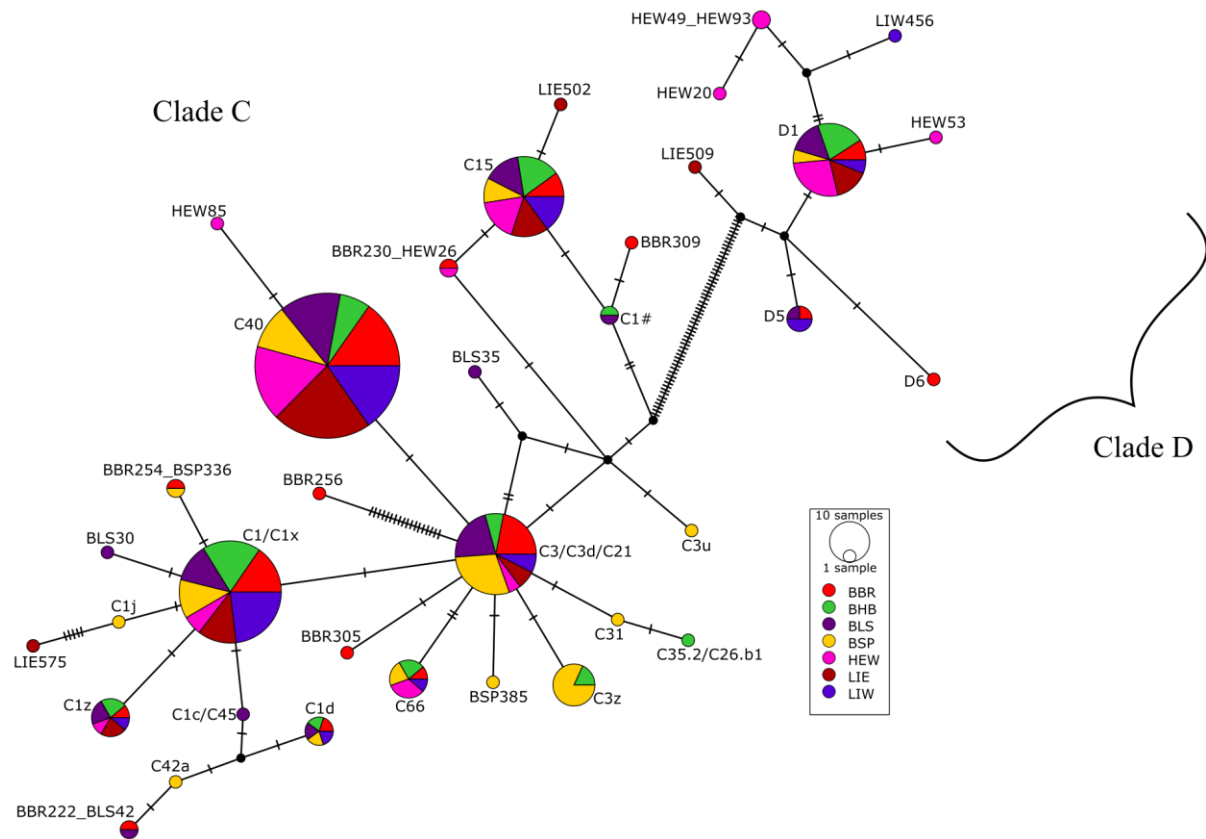
Figure 2.4: TCS Minimum Haplotype Network for the ITS2 region, with interpretation as *per* the legend of Figure 2.3. Note that the software could not distinguish types if they were separated by only an indel, so types C1 and C1x, as well as types C3d/C21 and C3 are rendered in a single circle.

2.3.4. psbA^{ncr} analysis

While most of the samples attempted were successfully amplified, sequencing results revealed low-quality sequences for many of the samples, characterised by either a low signal or multiple bases being detected at a single nucleotide site. This limited the number of samples analysed with the psbA^{ncr} region. This problem was particularly evident for clade D samples, where only 14 samples were acceptable for analysis. Interestingly, the samples with multiple bands (for both clade C and D) did not yield different sequences: once the low-quality regions of the sequence were trimmed off, the sequences from the same sample were identical. This would indicate that the different bands observed in the gel were caused by some PCR reactions in the sample failing to 'stop' at the correct part of the sequence, or the primers occasionally hybridising to another part of the sequence downstream from the target sequence, producing longer fragments with low-quality reads beyond a certain point (Chou *et al.* 1992).
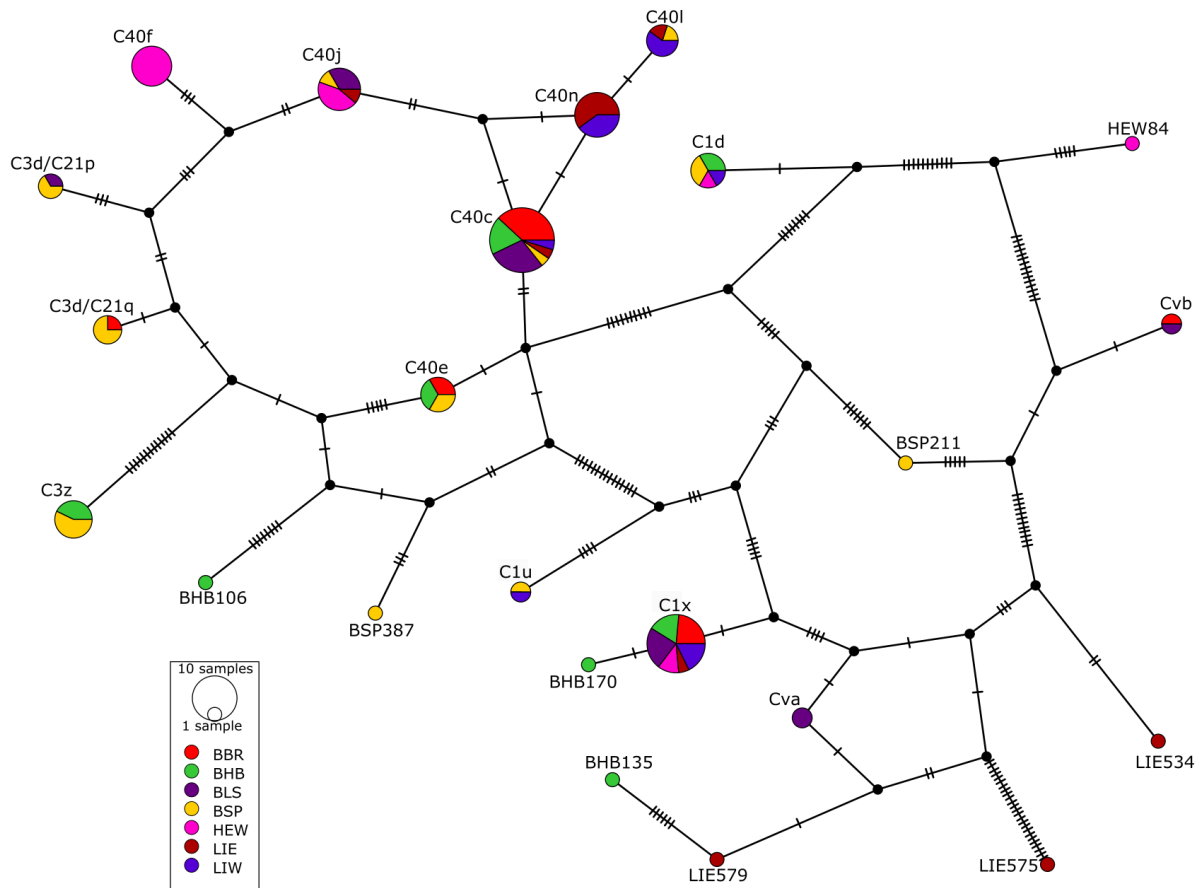
Figure 2.5: TCS Minimum Haplotype Network for the psbA^{ncr} region, with interpretation as *per* the legend of Figure 2.3. Note that the software could not distinguish types if they were separated by only an indel. Given the high number of these in the psbA^{ncr} region, only the groups that could be successfully separated by the program are displayed here, in order to demonstrate the higher level of partitioning that occurred with site.

Because the circles in Figure 2.5 only represent the consensus sequence from a range of closely-related sequences, it should be interpreted with caution. However, it is clear that there is very high variability at this locus, demonstrated by the number of differences between different groups. In addition, the increased resolution reveals much greater partitioning between geographic regions, especially within the C3 radiation. While the *cob* and ITS2 regions show that *Symbiodinium* C3/C40 sequences are distributed reasonably evenly across the seven sites, the specific groups within C40 displayed here are clearly region-specific. For example, symbiont type C40n is only found at Lamsana Inlet, while type C40f is found exclusively at Hera West. In contrast, types C40e and C40l are found only at Atauro sites, in addition to 90% of C40c. While statistical tests were not attempted on the psbA^{ncr} region, it follows and enhances the pattern of significant differentiation between Atauro and Timor sites.

2.3.5. Linking symbiont diversity with host diversity

There were several strong relationships between host genus and symbiont type, which are described below, as well as in Table A1.5. Unless otherwise specified, the symbiont type refers to the ITS2

designation (see Table 2.2). In all but one case, there was no geographic influence (*i.e.* the relationships described were consistent at Atauro and Timor sites). The significant exception is *Pocillopora*. At Timor sites, it exclusively associated with clade D types, apart from one sample at Lamsana Inlet West where it hosted type C1d (though note that *Pocillopora* was absent from Lamsana Inlet East). In contrast, at Atauro sites it exclusively associated with types from the C1 radiation, including all other samples of C1d.

There were several coral genera which showed high specificity for a single symbiont type, relationships which appeared particularly common in the C1 radiation. *Sandalolitha* exclusively hosted type C1x, while *Stylophora* exclusively hosted type C1z. However, while C1z was *only* found in *Stylophora*, C1x was hosted by a range of other genera including *Fungia, Leptastrea* and *Echinopora*. In addition, 6/7 of type C1 (as identified by the *cob* gene) was hosted by the reasonably rare genus *Astreopora*. However, there were also specific relationships evident for the C3 radiation. Symbiont type C40 was clearly the preferred dominant type for *Platygyra* (19/21), *Favites* (14/16) and *Diploastrea* (6/6), but its presence in a wide array of genera indicates that it is a host-generalist. The zoanthid *Millepora* only hosted C66. In addition, aside from one sample which hosted type D1, every other *Acropora* hosted symbiont types from the C3 radiation. One of the most striking examples of reciprocal specificity came from *Porites*: all but one sample of C15p (*cob* designation) was found in *Porites*, making up 34/37 of dominant symbiont types hosted by this genus. Further, C15q (*cob* designation) was almost exclusively hosted by *Montipora*, aside from one sample where it was hosted by *Porites*.

Aside from Hera West and Lamsana Inlet West, where clade D was exclusively hosted by *Pocillopora*, there was a reasonably even distribution of clade D types among genera at the other sites (where it was hosted by nine different genera). However, two reasonably uncommon genera showed an affinity for type D1: 7/7 *Seriatopora* hosted D1, while 6/13 *Galaxea* hosted D1, and the other 7/13 hosted sequences exclusively from the C1 radiation.

There was no specific relationship between unique sequences and host genus. In total there were 39 samples that had one or more unique sequences in at least one gene region, which were drawn from 17 host genera. However, it is worth noting that 6/7 of the non-scleractinian *Heliopora* hosted unique types, as well as 7/19 of *Montipora*. All other host genera and symbiont types showed reasonably general relationships, or were not present frequently enough to identify patterns. Interestingly, the highly specific psbA^{ncr} C40 types (C40c - C40o), which were often site-specific, showed no patterns in host genus. For example, type C40n was found in eight different coral genera.

Table 2.3 shows that, compared with previous studies, Atauro Island displays reasonably high symbiont diversity for the ITS2 region, with only Hawaii having a higher richness. In contrast, the Timor sites

Table 2.3: Diversity studies that report dominant symbiont types and the number of host genera that were sampled. More recent studies are absent, as they generally also report background symbiont types that are not separable from the dominant type in a sample. Notably high numbers of ITS2 types *per* host genus are bolded. For the psbA^{ncr}, the total number of samples was also able to be included; this could not be done for the ITS2 region due to this number not being reported by the other studies.

Study	Location	Host genera sampled	ITS2 types found	ITS2 types *per* genus	psbA^{ncr} unique sequences (total number of samples)	psbA^{ncr} types *per* genera *per* 100 samples
LaJeunesse *et al.* 2003	Southern GBR	40	23	0.58	-	-
LaJeunesse *et al.* 2003	Caribbean	50	35	0.70	-	-
LaJeunesse *et al.* 2004a	Southern/Central GBR	74	32	0.43	-	-
LaJeunesse *et al.* 2004a	Okinawa, Japan	31	20	0.65	-	-
LaJeunesse *et al.* 2004b	Hawaii	18	20	**1.11**	**-**	**-**
Thornhill *et al.* 2014	Caribbean, Pacific	38	25	0.66	244 (309)	2.08
This study	Atauro Island	40	34	**0.85**	79 (129)	1.53
This study	Timor	39	21	0.54	66 (97)	1.74

Table 2.4: Top five ITS2 OTUs in terms of the percentage of all sequences, the number of samples they appear in and their identity

OTU	% of all sequences	Total number of samples	ITS2 Identity
278	33.8 (1st)	152 (2nd)	C40
24	8.0 (2nd)	49	C15
154	7.1 (3rd)	52	C3d/C21
199	6.1 (4th)	91 (5th)	C1
320	4.6 (5th)	46	D1
119	3.2	195 (1st)	C3
271	1.0	125 (3rd)	C40p
145	0.9	112 (4th)	C115

display an ITS2 diversity comparable to other regions around the world. However, when looking at the psbA^{ncr} region, this trend is reversed, with Timor sites showing higher diversity than Atauro.

2.3.6. OTU-based analysis

Overall the NGS results were highly successful. The two most important quality statistics for Illumina MiSeq results are cluster density and base quality; a cluster density of 700-1000 k/mm^3 is optimal, while 75% of bases having Phred scores >30 is considered good. For these results, cluster density was 831 k/mm^3, and quality scores were 87% and 77% for the forward and reverse reads respectively, emphasising the high-quality nature of the data obtained. The confirmatory analysis run in `mothur` found spurious chimeras at a frequency of 0.04%, well below the 1% threshold used in these analyses.

After all processing, there were 6,163,296 sequence reads representing 852 unique sequences, which clustered into 366 OTUs over the 377 successful samples. Of those 366 OTUs, 295 were clade C, 66 were clade D, 4 were clade A and 1 was clade G. While this represents a high level of diversity, there was a distinct drop-off in the abundance of OTUs: the five most common OTUs accounted for 59% of all sequences recovered. Interestingly, the five largest OTUs in terms of number of sequences were not the same as the five most common OTUs in terms of the numbers of samples in which they appeared (Table 2.4).

The identity of the most common OTUs generally accorded with the dominant sequence identities found (Section 2.3.3). However, there were a few notable exceptions. ITS2 type C3, which was the dominant sequence in only four of the samples, was the top OTU in terms of number of samples it was found in. This indicates that it was present at background levels in a high number of samples. In addition, the third most common OTU in terms of the number of samples it appeared in was symbiont type C40p, a unique type that was not found as the dominant sequence in any of the samples. A similar pattern was observed with type C115; both appear to be extensively present at background levels at low frequencies, without ever being the most common type.

Strikingly, all samples had at least two distinct OTUs present, with an average of 7.1 *per* sample (minimum: 2; maximum: 19). While there was a high level of variability in the number of raw sequence counts recovered *per* sample, there was no relationship between this and the number of OTUs *per* sample (Figure A1.5), indicating that this is unlikely to influence results. In addition, there was no particular relationship between OTU richness and host genus. For example, 52 samples (14%) had more than 10 OTUs; those samples were evenly distributed among 23 different coral genera, indicating that there was not one particular genus prone to hosting a significantly higher number of OTUs. Further, 32 samples (8.5%) hosted more than one clade, which were spread over 20 different host genera. However,

there were differences evident in OTU richness between sites and broad geographic regions (Figure 2.6).
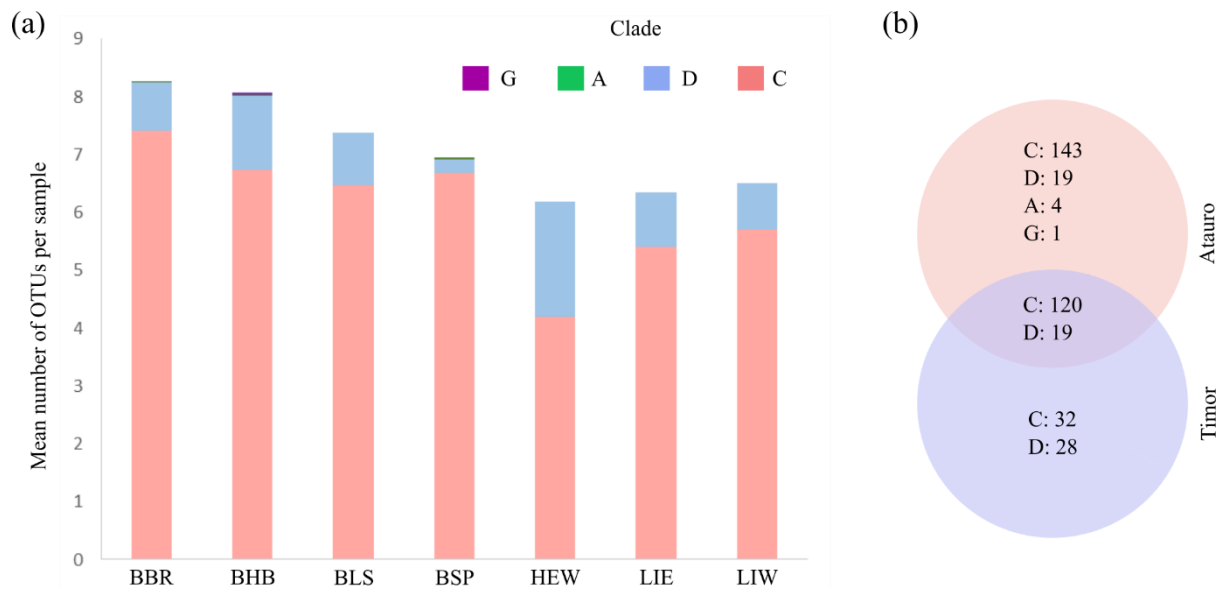
Figure 2.6: Comparison of OTU richness between sites and regions. (a) Average number of OTUs *per* sample belonging to the four clades found at each site. (b) Venn diagram of the number of OTUs unique to Atauro Island and Timor sites, separated by clade.

Figure 2.6 demonstrates that, generally, OTU richness was higher at Atauro sites than Timor sites, though this difference was not statistically substantiated (Kruskall-Wallis test, p = 0.383). However, there was a significant difference in composition between the two geographic regions (χ^2 test, p < 0.0001). Figure 2.6a shows that, while the number of clade D OTUs was similar among Atauro and Timor sites, the lower total number of OTUs means that, proportionally, clade D OTUs were much more prevalent at Timor sites (31% *vs.* 14%). This is supported by Figure 2.6b; the OTUs unique to Timor were evenly distributed among clade D and C, while unique Atauro OTUs were nearly all clade C.

To further explore regional differences without any potential confounding effect of host specificity (Tonk *et al.* 2013), coral genera that had a minimum of five representatives *per* region (Atauro or Timor) were investigated separately. This involved the five most common coral genera: *Acropora, Porites, Pocillopora, Platygyra* and *Montipora* (Figure 2.7).
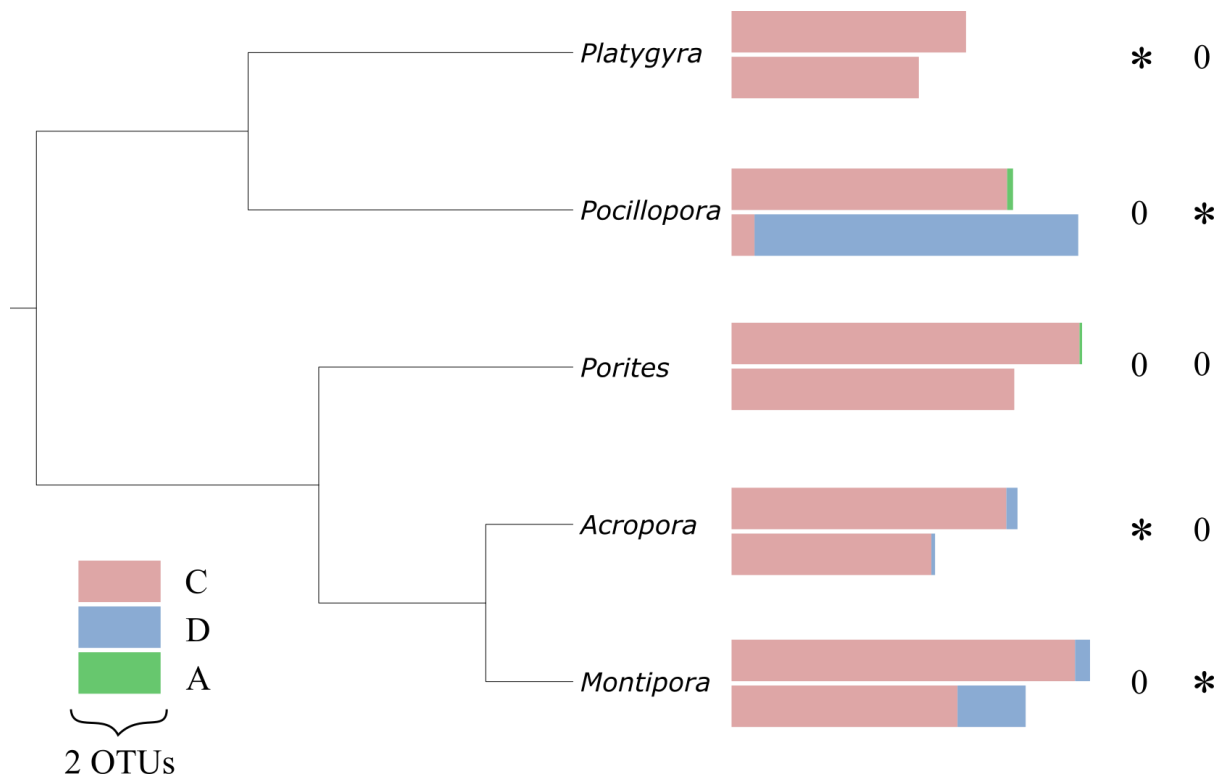
Figure 2.7: Mean OTU richness and composition for the five most common coral genera, for both Atauro Island sites (top bar) and Timor sites (bottom bar). The first column of symbols on the right indicates significance in terms of OTU richness (* = significant difference between Atauro and Timor, 0 = non-significant), while the second column indicates significance in terms of OTU composition (* = significant difference between Atauro and Timor, 0 = non-significant).

Differences in terms of OTU richness between Atauro Island and Timor sites were assessed with one-way ANOVA, while differences in cladal composition were assessed with χ^2 tests. *Porites* showed no differences between the sites in terms of either richness or composition, hosting only clade C, plus one clade A at background levels at Atauro. However, there was a fairly strong trend of increased richness at Atauro Island ($p = 0.051$). In contrast, there was a significant difference in terms of richness for both *Platygyra* ($p = 0.007$) and *Acropora* ($p < 0.0001$), with Atauro Island having 1.4 times and 1.3 times more OTUs, respectively. However, the composition of OTUs did not differ ($p > 0.05$ for both genera). In contrast, *Pocillopora* and *Montipora* showed no differences in OTU richness ($p > 0.05$), but significant differences in composition ($p < 0.0001$ and $p = 0.0002$, respectively). This difference was particularly pronounced in *Pocillopora*, with 93% of OTUs belonging to clade D in Timor, matching the dominant sequence analysis which found only clade Ds at Timor sites. In contrast, there were no clade D OTUs found at Atauro sites.

2.4. Discussion

2.4.1. Commentary on methods

This study utilised both direct Sanger sequencing and Next Generation Sequencing (NGS) to assess the diversity of symbiont communities at Atauro Island and adjacent sites at Timor. While the use of NGS to study *Symbiodinium* is not novel, running ~380 samples on a single MiSeq is significantly more than is generally attempted, with 96 being the typical maximum (*e.g.* Ziegler *et al.* 2017a). This approach was shown to be highly successful. Only two samples could be considered unsuccessful (0.5%), and high-quality data were produced which supported and added to the direct sequencing results, which had failure rates of 14% (*cob*) and 42% (psbA^{ncr}). Running a large number of samples on a single lane was highly cost-effective – assuming the researcher has access to indexing primers (a NZ$1500 outlay being sufficient for 80,000 samples), the sequencing cost per sample was <NZ$7, comparable to a standard PCR/direct sequencing cost while achieving a much higher sequencing depth. Other approaches used to quantify background symbiont populations are less appropriate. DGGE is prohibitively time-consuming over large numbers and insensitive (Quigley *et al.* 2014), while plasmid cloning is very expensive and artificially inflates rare variants (LaJeunesse and Thornhill 2011). While increasing the number of samples on a single lane does decrease sequencing depth *per* sample, if meta-community dynamics are not being studied then depth is less important. In addition, typically sequences that only appear at frequencies >1% in at least one sample are used anyway (Thomas *et al.* 2014, Ziegler *et al.* 2017a), so the decrease in sensitivity is likely to be negligible. Therefore, this method is strongly recommended for future studies of *Symbiodinium* diversity. The psbA^{ncr} could also technically be sequenced *via* NGS, but the long amplicon length means that this could only be done in one direction. For future *Symbiodinium* systematics, the development of additional short, highly variable gene regions amenable to NGS methods is recommended.

Further, the utility of using multiple genes was demonstrated. This is clear for the five samples that had co-dominant symbiont populations (see section 2.2.6). A single gene would not have discovered this, and failed to appreciate the full diversity of relationships present. It was also able to highlight potential inconsistencies in nomenclature; for example *cob* type C3 being a different symbiont type to ITS2 type C3 (Table 2.2). If this is not appreciated it could lead to confusion and incorrect comparisons between studies. In addition, it also challenged the assertion that the *cob* gene always contains less resolution than other regions, as it was able to detect further diversity within ITS2 type C15 (Table 2.2). This supports the use of a hierarchical approach to quantify symbiont populations, using genes with different sensitivities (LaJeunesse and Thornhill 2011).

Overall, these methods combined to find significant variation in symbiont populations at Atauro Island and also Timor. The following sections explore this diversity for the direct sequencing results (2.4.2) and OTU results (2.4.3). In each case, the contrasting patterns at Atauro and Timor will first be considered, before extending the discussion to comparisons with similar studies at other locations. As *per* the Results, symbiont types are referred to by their ITS2 designation unless otherwise specified.

2.4.2. Diversity in the dominant sequence analyses

There were significant differences in both symbiont richness and composition between sites (Figures 2.2 – 2.5). In some cases, this can be explained by differential host sampling between the sites. For example, 43% of *cob* type C1 was found at Beloi Saddle Patch. However, Beloi Saddle Patch also hosted 38% of the *Astreopora* sampled, the predominant host of that symbiont type. This appears to be through chance, as Turak and DeVantier (2013) do not report any notable increase of *Astreopora* at that site (though it could have changed since their survey in late 2012). A similar phenomenon could also explain some of the abundance differences of clade D between Atauro and Timor. The genus *Pocillopora* hosted type D1 at Timor sites, and 48% of all *Pocillopora* in the study was sampled at Hera West. However, this does not explain the presence of clade D exclusively in *Pocillopora* at Hera West and Lamsana Inlet West, when it appears in a variety of other genera at Atauro Island. Other differences also remain unexplained, such as the exclusive presence of type C3z at Beloi Harbour and Beloi Saddle Patch. At those two sites, C3z occurred in *Acropora, Montipora* and *Favites*, which were common across all sites. In addition, symbiont types C1 and C1x showed strong partitioning between Atauro and Timor sites, despite them having similar host genera (they share six host genera, largely from the families Fungiidae and Faviidae). Differential host sampling does not explain the richness differences between Atauro Island and Timor, both in terms of overall sequence diversity (Figure 2.2) and the number of unique sequences for the *cob* and ITS2 regions (see Table A1.4). On this evidence, there does appear to be some differentiation in symbiont richness and community structure between Atauro Island and Timor, disproving the first hypothesis of this study (that Atauro and Timor sites have similar symbiont diversity and composition). A contributing factor could be the underlying oceanographic conditions. The two regions are separated by the Ombai Strait, a channel over 3000 m deep (Atmadipoera *et al.* 2009). This features a strong north-south flow (Sprintall *et al.* 2004), which may prevent gene flow between Atauro and Timor, and encourage the production of distinctive *Symbiodinium* communities. However, this does not necessarily explain the greater diversity displayed at Atauro Island, especially given the reduced diversity in terrestrial fauna such as birds (Trainor and Soares 2004) and reptiles (Kaiser *et al.* 2013).

Despite the differences, there were also many conserved relationships between the regions, which serve to distinguish these reefs from other systems. Clearly, at Atauro Island and Timor sites, *Symbiodinium*

C40 is a common generalist symbiont, fulfilling the role that type C3 typically inhabits elsewhere in the Indo-Pacific. For example, *Diploastrea* and *Favites* are largely C40 here, when they are dominated by C3 elsewhere (LaJeunesse *et al.* 2004a). In addition, *Acropora* is mainly (and often exclusively) inhabited by C3 as the dominant symbiont (LaJeunesse *et al.* 2003; Davies *et al.* 2016), while here it hosts C40 as well as a range of other symbionts in the C3 radiation. However, C40 has also been shown to be the dominant symbiont at sites in both Western Australia (Silverstein *et al.* 2011) and Micronesia (Davies *et al.* 2016). While such results may be attributed to stochasticity or general flexibility in these relationships, it is very interesting to note that from the OTU analysis, C3 was actually the most common symbiont in this study, in terms of the number of samples it appeared in (52% of all samples, Table 2.4). The fact that it was only the dominant symbiont in 1% of samples would indicate that C40 holds some sort of selective or competitive advantage over C3, or that there is some functional benefit to corals keeping C3 at background levels. Additionally, it could be explained by intragenomic variation, though given that C3 appears at background levels in 191 samples and C40 only appears in 152 samples in total, it would also require multiple other types to have residual C3 repeats in their genomes. This would be an interesting direction of further study.

There are several other noticeable differences between this study and other studies. While type C3z has previously been reported, it appears in the literature largely in the western Indian Ocean (African coast and Madagascar) (LaJeunesse *et al.* 2010b; Chauka 2012; Yang *et al.* 2012). To the author's knowledge, this is the first time that C3z has been recorded inside the Coral Triangle. Finally, several of the non-scleractinians also display divergent symbiont relationships. *Millepora*, which exclusively hosts C66 as its dominant symbiont, has largely been reported to host type A7 elsewhere (LaJeunesse *et al.* 2003; Tonk *et al.* 2013). When it does host clade C, it appears to be only C57 (LaJeunesse *et al.* 2004a). *Heliopora* hosts largely unique previously unidentified types, which are not shared with other host taxa or indeed conspecifics. This is a particularly unusual result given they appear to obtain their symbionts horizontally: studies have shown their planulae lack zooxanthellae (Harii *et al* 2002). There is little literature on the symbionts of *Heliopora*, but this wide diversity between individual colonies suggests it may be worthy of greater consideration.

There were also relationships found here that are widely conserved across multiple studies. *Stylophora* has been shown to host only symbionts from the C1 radiation (Fitt *et al.* 2009), as it did here. In this case it was a previously unreported subtype C1z, likely a local variant. In addition, *Galaxea* has been reported to host either D1 or C1 as its dominant symbiont type (LaJeunesse *et al.* 2004b; Zhou and Huang 2011), exactly as it did here. *Porites* almost exclusively appears to host *Symbiodinium* C15, in associations that are highly stable through time (LaJeunesse *et al.* 2003; Fitt *et al.* 2009; Zhou and Huang 2011; Putnam *et al.* 2012). *Montipora* has also been reported to host C15, though the authors that reported this expressed mild surprise that it hosts the same type that has been well-established to

occur in *Porites* (LaJeunesse *et al.* 2003). My results bring further clarity to this, by showing that C15 encompasses two different types that clearly partition into the two genera. All four of these genera (*Stylophora, Galaxea, Porites, Montipora*) are largely vertical transmitters of symbionts to their offspring (Baird *et al.* 2009); this study hence affirms the general hypothesis that vertically transmitting corals are more predisposed to hosting specific (as opposed to generalist) symbionts (Stat *et al.* 2008; Fabina *et al.* 2012).

The results for *Pocillopora* also match previous studies, where it has been shown to host either *S. glynii* (D1) or symbionts of the C1 radiation (*e.g.* C1b-c) as their homologous (*i.e.* native) type depending on location (Rowan 2004; Magalon *et al.* 2007; LaJeunesse *et al.* 2008; McGinley *et al.* 2012). These relationships have been shown to be highly stable, and are hence attributed to long-term ecological processes in addition to the host cell environment (LaJeunesse *et al.* 2008). It has been shown that such assemblages are independent of the species of *Pocillopora* sampled (Cunning *et al.* 2013), so it is unlikely that these results are explainable by sampling different species of *Pocillopora* at the different sites. The major contributor to *Pocillopora* hosting these different homologous types in distinct but close geographic areas appears to be stress, particularly past bleaching events (Glynn *et al.* 2001; Rowan 2004). In this case, however, it appears implausible that bleaching has caused the different symbionts between Atauro Island and Timor sites, given that there is no evidence of coral bleaching at Timor reefscapes dating to before the first global bleaching event in 1998 (Turak and DeVantier 2013). However, it is possible that other stressors such as terrestrial runoff or nutrient enrichment can also lead to the observed changes in *Pocillopora* symbiont types (LaJeunesse *et al.* 2008). Turak and DeVantier (2013) found lower visibility due to terrestrial runoff as well as higher litter levels at the Timor sites, particularly Lamsana Inlet, something they attribute to the higher human population adjacent to these sites. Given that Timor is significantly agricultural, it is likely that the terrestrial runoff is nutrient-enriched to some extent. This combination of factors appears the most plausible reason for the differential symbiont populations hosted by *Pocillopora* across the sites surveyed here, and other studies have reported clade D symbionts in *Pocillopora* from more turbid areas (LaJeunesse *et al.* 2010b, 2014; Wham *et al.* 2017). Further, Stat *et al.* (2013) found a correlation between sedimentation and clade D *Symbiodinium* in *Montipora* in Kaneohe Bay, Hawaii. The current study also found a greater proportion of clade D in *Montipora* at Timor (Figure 2.7), which further supports this explanation. While further research is required to definitely prove this, human-mediated impacts on the reefs appears a possible mechanism for the general diversity differences displayed between Timor and Atauro Island.

In terms of ITS2 diversity, Atauro Island appears very symbiont-rich compared with other studies and Timor (Table 2.3). Only the Hawaii study reports a high number of symbiont types *per* host genus, a result which was considered to represent an extremely high level of diversity (LaJeunesse *et al.* 2004b). In addition, all the other studies in Table 2.3 conducted sampling at both deep and shallow sites, and

many of the symbiont types found were exclusively at deep sites. The current study only sampled in shallow regions, and therefore diversity estimates would likely be further elevated if symbiont assessments at other depths were carried out. Further, in the Hawaii study the diversity was largely in three vertically (*i.e.* maternally) transmitting genera, a symbiont acquisition mode which has been shown to promote symbiont specialisation (Douglas 1998) and is thought to be the major cause of the high diversity observed (LaJeunesse *et al.* 2004b). In contrast, the different types found here were spread reasonably evenly over 17 host genera, encompassing a wide range of horizontally- and vertically-transmitting corals. Finally, many of the genera were probably under-sampled (54% had ≤5 samples taken), and increased sampling may drastically increase the observed diversity. This provides partial support for the second hypothesis of this study, in that Atauro appears to have a higher level of symbiont diversity than other regions around the world, while Timor does not. However, the psbA^{ncr} values in Table 2.3 would suggest the opposite. This result could reflect a higher rate of evolution at Timor sites caused by environmental stress (Moran and Alexander 2014), or, more likely, that sampling had not reached a saturation point yet for this gene (Figure A1.4).

To conclude this section, the sites studied here show very high symbiont diversity, though with a significant discordance between Atauro Island and Timor. Host-symbiont assemblages are reasonably consistent with other studies, though also show novel relationships (*e.g.* C3z) which further set this region apart from other reefs examined in the literature.

2.4.3. Total diversity in Atauro and Timor corals: the OTU-based approach

Overall, the OTU results from the NGS support the direct sequencing results. There were clear differences in OTU cladal composition between Atauro and Timor sites, shown both overall (Figure 2.6a) and for the common genera *Pocillopora* and *Montipora* (Figure 2.7). In addition, there were many more OTUs unique to Atauro Island than there were to Timor (Figure 2.6b). This included the presence of clades G and A at background levels. While these are reasonable associations to uncover, it should also be noted that this may have been caused by contamination on the surface of the coral, especially as these clades were never recovered as the dominant sequence. There was also an observable non-significant drop in OTU richness from Atauro to Timor, a pattern that was particularly apparent in both *Acropora* and *Platygyra*. In this case, differences in the number of host genera sampled can be categorically eliminated as an explanation, as richness was assessed on a *per*-sample basis. Therefore, a difference in symbiont communities between Atauro and Timor is further supported by the NGS data.

However, other results also support broad-level homogeneity between the two regions, especially Table 2.4. The fact that the third most common OTU found here is a previously undefined symbiont type, C40p, which appears commonly in both regions, would indicate that there is potential for gene flow. In

addition, the high prevalence of types C3 and C115 at background levels, despite being nearly or totally absent as the dominant type, suggests a shared evolutionary history between the two regions. This pattern could be the result of intragenomic variation, or cryptic functional benefit. For instance, there is the potential for symbiont combinations to convey an interactive benefit to a host that cannot be achieved by symbiont populations in isolation (Ziegler *et al.* 2017b).

Comparison with other studies is more difficult, as slightly different processing methods could have a profound influence on the number and nature of OTUs recorded. In addition, the OTU approach itself is not without limitations. Different 'species' (in the sense of a reproductively isolated population) may display different levels of genetic variation (Correa and Baker 2009). Using a single cut-off point (97% similarity) for collapsing sequences into discrete units may mean that some OTUs could potentially host multiple species, while others may not encompass the full range of a single species. While a 97% cut-off was claimed to resolve species (Arif *et al.* 2014) and has since been adopted for NGS of *Symbiodinium* (*e.g.* Thomas *et al.* 2014, Ziegler *et al.* 2017a), a paucity of definitive information on what constitutes a *Symbiodinium* 'species' means further work is recommended to have complete confidence in an OTU-based approach. Despite this, there are some striking comparisons to be made. In the present study, 8.5% (32/377) of the samples hosted more than one clade of *Symbiodinium*. Other studies also report this statistic, and return values of 8.5% (Thomas *et al.* 2014), 9% (Putnam *et al.* 2012) and 6% (Tonk *et al.* 2013), a number which appears remarkably consistent across divergent reefscapes. In saying that, higher levels of up to 78% have been reported (Mieog *et al.* 2007), though this is attributable to recent bleaching events (Mieog *et al.* 2007; Rouzé *et al.* 2017). Hence, the reasonably low percentage reported here further supports the assertion that these reefs have not experienced significant bleaching. While there may be an element of coincidence to this congruence, it may also reflect some aspect of evolutionary history that maintains background clades at this frequency. A possible mechanism is beyond the remit of this thesis, but it provides a potentially fascinating area of future study.

The richness of OTUs is also higher than found in other studies. This is consistent for both total OTUs and OTUs *per* host genus. For example, across *Acropora, Montipora, Pocillopora* and *Porites*, Ziegler and colleagues (2017a) found on average 2.9 OTUs *per* sample, in contrast to the 7.4 here. Thomas *et al.* (2014) found 4 OTUs *per Acropora* colony, in contrast to the 6.2 here. In addition, the diversity appears to be much more even in the current study. The most common OTU by number of sequences accounts for 34% of all sequences here; in Western Australia, by comparison, this figure is 98% (Thomas *et al.* 2014). While these comparisons must be treated with caution, these results support the direct sequencing results that Atauro sites harbour abnormally high *Symbiodinium* diversity that is reasonably evenly distributed.

2.4.4. Conclusions

The hypothesis that symbiont diversity and composition would be similar between Atauro and Timor sites has been strongly disproven, with Atauro Island displaying high diversity relative to Timor and significantly different symbiont communities, particularly a high proportion of clade D symbionts at Timor sites. Thanks to the OTU analysis, variable host sampling and bleaching can largely be eliminated as causes. In addition, Turak and DeVantier (2013) documented no clear differences in coral communities between the regions. Therefore, a combination of oceanographic conditions and environmental (*e.g.* turbidity) stress appears to be the most likely explanation for this. However, there are several aspects of the symbiont community shared between regions which distinguish them from other regions and support the conservation value of these sites. The hypothesis that these regions would show elevated diversity relative to other reefs is partially supported; while Timor sites showed similar levels, Atauro diversity was higher than most other studies, even with only shallow sites and small numbers of some genera being represented. Given the highly diverse macrofauna documented at this site (Conservation International 2016), there may be a link between symbiont diversity and reef health. Turak and DeVantier (2013) observed environmental degradation consistent with human influences at Timor sites, and while this does not appear to have significantly affected the coral community, this study points to a potentially significant reduction in symbiont diversity. Therefore, it is possible that by the time corals start to show negative effects, the adaptive potential of high symbiont diversity may have been lost without anyone realising. This should encourage further thorough analyses of symbiont diversity, in order to accurately predict and perhaps ameliorate the deleterious effects of human-induced environmental change.

Chapter 3: Multi-gene evidence of symbiont hybridisation at Atauro Island, Timor-Leste

3.1. Introduction

In order for coral reefs to survive the stresses of a changing climate, genetic adaptation over rapid evolutionary timescales has to occur. Adaptation in the coral itself may go some way to provisioning for the environmentally challenging conditions predicted to come (Rodriguez *et al.* 2009). However, given that the response of corals to environmental conditions is inextricably linked to the diversity and performance of their symbionts, increasing attention is being focused on the evolutionary potential within the genus *Symbiodinium*.

Coral symbionts were thought to be nearly exclusively asexual (Trench 1997; LaJeunesse 2005), thanks to their isolated position sequestered inside host cells, and the hypothesis that endosymbiotic sex would encourage exploitation of the host (Law and Lewis 1983). However, previous work in other taxa has shown that intracellular symbionts can sexually reproduce (Chesnick and Cox 1987). In general, it is thought that many such organisms may have cryptic sexual cycles that have previously been unappreciated, in addition to the production of clonal populations *via* asexual reproduction (Heitman 2010). Now, there is significant evidence that *Symbiodinium* also displays a mixed reproductive strategy, with periods of asexuality interspersed with occasional to frequent sex. While it has never been explicitly observed, there are distinct and observable traces of sex in their genomes (*e.g.* Baillie *et al.* 2000; LaJeunesse 2001; Santos and Coffroth 2003; Santos *et al.* 2004; Pettay *et al.* 2011; Baums *et al.* 2014; Chi *et al.* 2014; LaJeunesse *et al.* 2014; Thornhill *et al.* 2014; Levin *et al.* 2016; see also review by Thornhill *et al.* 2017).

Sexual reproduction can facilitate adaptation by generating novel genetic variation, which selection can then act on. However, symbiont sex may not be enough to ensure the survival of coral reefs into the future. The maximum recorded increased tolerance conveyed by symbionts is 1.5ºC (Berkelmans and van Oppen 2006), while increases up to 3ºC are predicted by the end of the century (Kirtman *et al.* 2013). Therefore, even if sex does produce increased thermal tolerance in *Symbiodinium*, it may not be sufficient to facilitate coral survival. Further, bursts of diversification in *Symbiodinium* have traditionally occurred in periods of cooler temperatures (Pochon and Pawlowski 2006); warmer temperatures may lead to significantly reduced diversity as many symbiont types may die out. This removal of genetic diversity eliminates the raw material for adaptation *via* sex. Put plainly, corals are unlikely to be able to keep pace with the warming climate, even with the adaptive potential of symbiont sex (Baskett *et al.* 2009; van Oppen *et al.* 2015).

However, the presence of sex raises the possibility of hybridisation, sexual reproduction between divergent genetic lineages. This can allow for rapid adaptation, by facilitating macro-evolutionary jumps (Willis *et al.* 2006; Dittrich-Reed and Fitzpatrick 2013). An example is introgressive hybridisation, where the F1 hybrids subsequently mate with one or both parent populations; this can transfer a large quantity of genetic material between the two parent lineages in the space of a few generations. In addition, hybridisation can also produce offspring with elevated fitness ('hybrid vigour'), which can even outcompete the parent species (Ellstrand and Hoffman 1990; Rhymer and Simberloff 1996). Importantly, instances of hybridisation have also been shown to increase in taxing conditions (Rhymer and Simberloff 1996; Moran and Alexander 2014). Therefore, the possibility of hybridisation in coral symbionts raises tantalising opportunities for adaptation at the required pace and scale for survival. Other biologically similar taxa show evidence of adaptive hybridisation, which encourages the hypothesis that *Symbiodinium* could hybridise.

In particular, there is evidence that a range of dinoflagellates hybridise. For example, Brosnahan and colleagues (2010) recorded reproduction between two genetically diverse groups (I and III) of *Alexandrium tamarense*, though the zygote became inviable after three cell divisions and died. A more successful example is between *Dinophysis acuminata* and *D. norvegica*. Successful hybrids between these two species have been recorded from both Norway (Edvardsen *et al.* 2003) and Scotland (Hart *et al.* 2007). In addition, there is evidence from genetic data that hybridisation also occurs in the genera *Protoperidinium, Preperidinium* and *Diplopsalis* (Gribble and Anderson 2007). Of course, these examples all come from exclusively free-living taxa, with no significant barriers to encountering other species. However, there is also evidence from plant-fungi relationships that endosymbionts can successfully hybridise. The endophytes *Epichloë* spp. are pathogenic or mutualistic fungi that inhabit a wide range of grasses. Hybridisation appears to be a major mechanism for diversification in this genus, and has been reported to occur inside the grasses *Lolium perenne* (Schardl *et al.* 1994), *Festuca arundinacea* (Tsai *et al.* 1994), *Bromus laevipes* (Charlton *et al.* 2014) and *Poa alsodes* (Shymanovich *et al.* 2017). In several cases, multiple cases of hybridisation have been recorded, and evidence put forward that those hybrids are fitter than non-hybrids (Schardl *et al.* 1994; Moon *et al.* 2004). It should be noted, however, that these endophytes grow in intercellular space in their hosts, rather than being sequestered in host cells as *Symbiodinium* are. However, in other respects they are remarkably similar. Both are generally asexual with evidence of cryptic sex, infect a broad variety of hosts, and are often transferred *via* host embryos, where hybridisation could occur.

Hybridisation remains understudied as a mechanism for evolutionary leaps (Dittrich-Reed and Fitzpatrick 2013), both in general and especially for *Symbiodinium*. However, the above evidence that both dinoflagellates and endosymbionts can hybridise increases the plausibility of it as an evolutionary mechanism. The extensive presence of background symbiont populations inside hosts, the observation

that corals themselves hybridise (Willis *et al.* 2006, Combosch and Vollmer 2015), and the existence of a free-living state (Coffroth *et al.* 2006; Nitschke *et al.* 2016) mean it is highly possible (even likely) that at some point diverse symbiont communities may interact, with the possibility for sexual reproduction. This is especially possible for those vertically transmitting corals that host diverse symbiont consortia in their eggs and where genotypically distinct symbionts are hence not physically segregated (Padilla-Gamiño *et al.* 2012). Because researchers have not been focused on hybridisation as an adaptive mechanism, evidence for its existence could easily have been glossed over. Some possible examples for *Symbiodinium* are given in the following paragraph.

LaJeunesse and colleagues (2003) reported an ITS2 sequence variant they called C1c and treated as an intragenomic variant, as it was only observed in DGGE profiles associated with type C1. However, it was then discovered to be an independent type and called C45 (LaJeunesse 2005). Therefore, the additive DGGE pattern shown in LaJeunesse *et al.* (2003) could have in fact resulted from the hybridisation of C1 and C45 (Wilkinson 2015). In the same paper, LaJeunesse (2005) defined type C3m using the ITS2 region, which has co-dominant characteristics of both C1 and C3. This was attributed to either sexual recombination or homoplasy. A similar scenario was also recorded in symbiont type C3h, an apparent intermediary between C3 and C21 (LaJeunesse *et al.* 2004a). This time, the pattern was hypothesised to be due to incomplete lineage sorting or sexual recombination between the two different types. Given the unambiguous existence of 'pure' C3 and C21 in the samples, sexual recombination is possibly a more credible explanation than having a symbiont 'caught between' the two established genotypes. Finally, one of the only studies that actually targeted hybridisation found evidence for its existence. Wilkinson and colleagues (2015) reported two symbiont types but three distinct symbiont populations inside a single *Pocillopora* colony: C100 symbionts, C109 symbionts, and symbionts having co-dominant C100 and C109 repeats in the same cell. Again, the extensive presence of the two 'pure' populations means incomplete lineage sorting is a less parsimonious explanation than hybridisation. However, it cannot be completely eliminated as a possibility. In addition, this study took place at Lord Howe Island, the world's southern-most coral reef, and therefore may not be widely applicable across less marginal, low-latitude sites.

Hence, there is a body of indirect evidence for sexual recombination between diverse symbiont types (hybridisation *sensu lato*), and this warrants in-depth study. The current study aimed to gather further defendable evidence as to whether hybridisation occurs in coral symbionts. Because it is very difficult to observe hybridisation directly, it is generally inferred through genetic signals. One of the most common of these is incongruence between gene regions, which can be caused by hybridisation in a number of ways. Because nuclear genes are inherited biparentally, while organelle genes are inherited uniparentally, sexual reproduction between different species will result in organelle genes resembling one parent only, while the nuclear genome will have clear traces of both parents (Rieseberg *et al.* 1996).

In extreme cases, repeated backcrosses with a parent type can result in organelle capture, where novel, discordant nuclear-organellar combinations are observed (Folk *et al.* 2017; Figure 3.1).
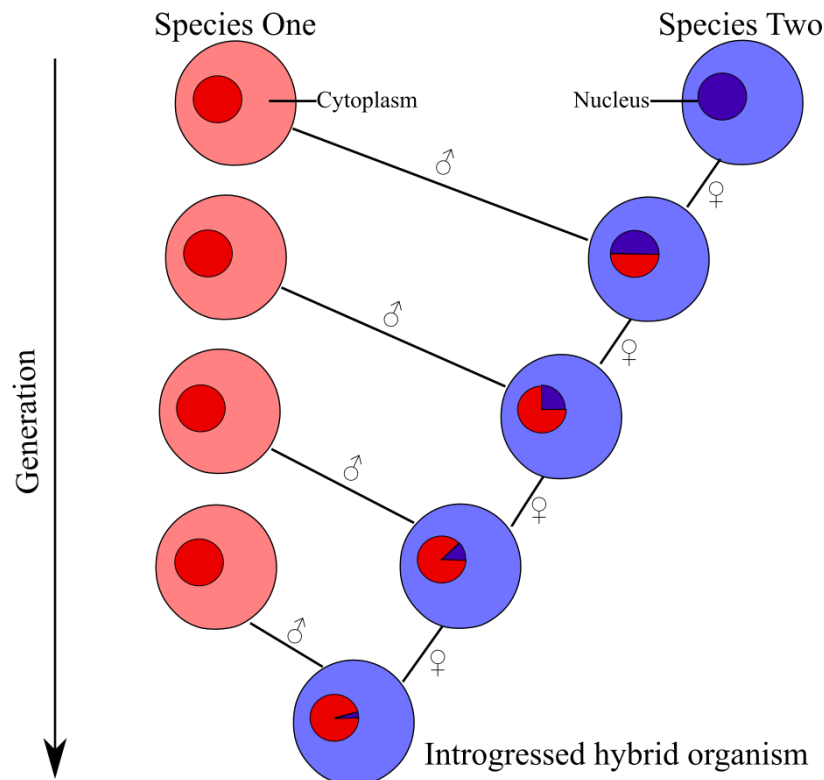
Figure 3.1: Theory of introgressive hybridisation causing incongruence, where one or a few representatives of Species Two contact a population of Species One. If Species Two acts as the maternal contributor in the initial hybridisation event, and the hybrids continue to do so, after several generations, then the hybrids will largely have the nuclear genome of Species One (red) but the organellar genome of Species Two (blue). It has been found that minority taxa will nearly always be the 'female' parent of hybrids due to gamete fusion typically being egg-limited rather than sperm-limited (Rieseberg *et al.* 1996). Figure adapted for *Symbiodinium* from Rieseberg *et al.* (1996).

Following a hybridisation event, selection can also act to produce incongruence between gene regions: there may be elevated (or reduced) fitness of certain nuclear-cytoplasmic combinations, or selection pressure may be different for nuclear and cytoplasmic genomes (*e.g.* a greater selection pressure acting on nuclear genes) (Rieseberg *et al.* 1996). Therefore, identifying incongruence between gene regions is a common method for assessing potential hybridisation (Planet *et al.* 2006, Govindarajulu *et al.* 2015), and was utilised in the current study.

The chosen location for this study of Atauro Island and the north coast of Timor, is in the Coral Triangle and therefore widely applicable to other important reef systems. The hypothesis tested was that hybridisation between distinct *Symbiodinium* genotypes has occurred at these sites, as evidenced by various gene regions (*cob*, ITS2, psbA^{ncr}) having experienced different evolutionary histories. If proven to be correct, then this may provide a plausible mechanism for coral reef survival in the future.

3.2. Methods

3.2.1. Incongruence tests utilised

A brief introduction to the incongruence tests utilised in this chapter is provided below, as a basic mathematical understanding of their function is necessary when their results are interpreted and discussed. Ideally, a statistical test would be able to test the null hypothesis 'Dataset X and Dataset Y are congruent', against an alternate hypothesis 'Dataset X and Dataset Y are incongruent.' Regretfully, no such test exists for phylogenetic data. As such, other tests with slightly different hypotheses have been frequently employed as an approximation. Two of these tests will be utilised here.

The Incongruence Length Difference (ILD) test was originally developed by Farris and colleagues (1994), using a criterion of maximum parsimony. At its simplest, the test is used to compare two data partitions (nucleotide alignments) X and Y, of arbitrary length. The shortest number of steps (evolutionary events) required to build tree X and tree Y from the partitions independently is calculated and summed. The test then generates random partitions P and Q, the same sequence length as X and Y, by randomly drawing characters (*i.e.* columns in the alignment, a single nucleotide position) from X and Y, and placing them in the partitions P and Q. This is repeated for a defined number of replicates, and the test statistic is the number of replicates for which X+Y produces a more parsimonious tree than P+Q. The null hypothesis is that the defined partitions (X, Y) are no more parsimonious than random partitions (P, Q) while the alternate hypothesis is that the defined partitions are significantly more parsimonious than random partitions. Functionally, this can be used to test if two datasets have undergone separate evolutionary histories (Planet 2006). The implication is that if X and Y are indeed more parsimonious, they encode contrary evolutionary information that is lost when randomised, and hence leads to less parsimonious trees in P and Q. In this way, incongruence can be assessed. While the example here is two partitions, the test can technically be scaled up to any number of partitions. This test uses parsimony as its criterion, and though there is no reason why it should not be as successful using maximum likelihood or distance methods, currently there are no programs available using these criteria (Planet 2006; Leigh *et al.* 2011).

An explicit tree-based test was developed using the criterion of maximum likelihood (ML), the Shimodaira-Hasegawa (SH) test (Shimodaira and Hasegawa 1999). This is designed to test a multiplicity of trees, allows for the fact that the optimal tree may be included, and establishes a 'confidence set' of trees that explain the data equally well. The null hypothesis is that all tested trees are equally good explanations of the data, while the alternate is that some or all tested trees are not equally good explanations of the data. In practice, this test identifies the best tree for a given dataset (*i.e.* a multiple sequence alignment), and then presents output as to whether other candidate trees are

statistically distinct from that best tree. The output hence appears as pairwise comparisons between two trees. This procedure can be used to test for incongruence in datasets X and Y, using trees T_X and T_Y made from those datasets. If T_X and T_Y are equally likely for all or most characters in X and in Y (tested in two separate tests), the test will find a p-value >0.05, and it can be concluded that X and Y are congruent, as their trees do an equally good job of explaining each other's data. If they are incongruent, it is expected that T_X will be significantly better than T_Y when considering dataset X, and *vice versa* for T_Y and Y. Alterations have since been made to this test, as it can be very conservative. The weighted SH test (WSH) was also defined by Shimodaira and Hasegawa (1999), which divides the test statistic by the variance for each set of trees tested. The Approximately Unbiased (AU) test was developed by Shimodaira (2002) as a derivation of the SH test, and generally finds more accurate results when there are many candidate trees, or some trees are particularly unlikely (Shimodaira 2002; Strimmer and Rambaut 2002). The SH test (both normal and weighted) and the AU test were utilised for testing procedures.

3.2.2. Data Assembly

Raw sequence data were obtained *via* the methods outlined in Chapter 2. As the tests require a single sequence *per* sample, the most dominant ITS2 sequence from the Next Generation Sequencing in each sample was extracted. Before analyses, data were filtered extensively. First, only samples that had successful sequences for all three gene regions were chosen, as the tests require exactly the same taxa lists for each tree or partition. This eliminated a large number of samples, as generally unsuccessful sequences were not consistent within samples (*i.e.* an unsuccessful psbA^{ncr} sequence would render the sample invalid for use in the analyses, even if the *cob* and ITS2 sequences were successful). Following this, only samples that could be placed in an unambiguous alignment were used, which eliminated several samples with highly divergent psbA^{ncr} sequences. This also meant that clade D samples were not included, as their psbA^{ncr} sequences were not alignable with those of clade C. They were not included as a separate analysis due to the small amount of successful psbA^{ncr} sequences and the fact that the clade D *cob* gene was completely invariant. This limited the data to between 18 and 28 samples *per* site ($\bar{x} = 22.7$), with a total of 159 samples used. However, this number is typical of other studies assessing incongruence, and provides the resolution required (*e.g.* Dolphin *et al.* 2000; Govindarajulu *et al.* 2015).

For the Shimodaira-Hasegawa and Approximately Unbiased tests, alignments were created and manually edited in Geneious v8.0.5, using the in-built Geneious alignment algorithm with all default settings (gap open penalty = 12, extension = 3). Each site (BBR, BHB, BLS, BSP, HEW, LIE, LIW; see Table 2.1 for abbreviations) had a separate alignment for each gene region (*cob*, psbA^{ncr}, ITS2), leading to 21 separate alignments. Additional holistic datasets for each gene region were created for

Atauro Island (93 samples) and Timor (66 samples), to facilitate broad-scale island comparisons. All alignments had 787, 369 and 531 columns for the *cob*, ITS2 and psbA^{ncr} regions respectively. In total, 27 separate datasets were assembled. The samples with co-dominant clade C and D populations (see Chapter 2) were not included in the analyses, while sample BHB151 (co-dominant clade B and C) was used as an outgroup for these analyses (see Appendix A2.2).

For the Incongruence Length Difference test, the three gene regions were concatenated for each site, with each region then treated as a separate partition (*cob*: 1-787; ITS2: 788-1156; psbA^{ncr} 1157-1687; total of 1687 columns, Figure 3.2b). This was carried out for each site, plus for Atauro Island samples and Timor samples as above (total of nine different concatenations).

For the parsimony analyses, gaps were coded as a fifth character state (except in the *cob* datasets, where no gaps existed). All analyses described below used the program PAUP* 4.0a159 (Swofford 2002) unless otherwise specified. Note that in PAUP*, the incongruence length difference test is called the partition homogeneity test.

3.2.3. Incongruence Length Difference (ILD) test

The individual site analyses were originally carried out with 100 replications, using a MaxTrees value (number of trees stored at any one time) of 1000. For results that had p-values <0.2, a more thorough confirmatory analysis was run with 1000 replicates and a MaxTrees value of 10000. In all cases, the p-values between the two sets of tests differed by <0.015, and therefore the tests with original p-values >0.2 would be extremely unlikely to change the result if the more extensive tests had been run on them.

The Atauro Island and Timor analyses were also carried out with 100 replications, with MaxTrees set to 100 to compensate for the additional computational burden of the larger dataset. Exploratory analyses of subsampled datasets (not presented) showed that the difference in MaxTrees setting had minimal impact on the performance of the test. All other settings used for the tests were the PAUP* defaults. Conclusions were drawn at $\alpha = 0.05$.

3.2.4. Shimodaira-Hasegawa (SH) and Approximately Unbiased (AU) tests

Maximum likelihood trees were generated for all gene regions by individual site (all possible combinations of {*cob*, psbA^{ncr}, ITS2} and {BBR, BHB, BLS, BSP, HEW, LIE, LIW} *i.e.* 21 different trees). Trees were also made for each gene region for Atauro Island and Timor datasets (*i.e.* six trees). The appropriate evolutionary model was determined for each of the 27 datasets individually by first making a neighbour-joining tree using a Jukes-Cantor distance measure and then running the

`automodel` command. The appropriate evolutionary model for each dataset was then employed when making the maximum likelihood trees (see Table A2.1, Appendix 2). A basic heuristic search was run to generate a base tree or trees, which was then bootstrapped. All bootstrapping procedures used a heuristic search with random sequence addition. For the individual site datasets, the following procedure was used. Bootstrapping of *cob* datasets involved 1000 replicates, using a heuristic search algorithm with all other settings as the PAUP* defaults and unlimited MaxTrees. For the ITS2 and psbA^{ncr} datasets, bootstrapping had 100 replicates, with MaxTrees set at 1000 for the initial heuristic search and 100 for the bootstrapping. In addition, for the psbA^{ncr} datasets, the number of addition sequence replicates was set to 2 (*versus* the default of 10). For the Atauro Island and Timor datasets, bootstrapping involved 1000 replicates using the `faststep` search option for all three gene regions. Nodes with <50% bootstrap support were collapsed into polytomies. This procedure yielded 27 maximum likelihood trees, one for each gene region for each of the nine datasets.

A set of 100 random trees was also generated for each dataset, using the `generate random` command employing an equiprobable model. These trees are random in the sense that they represent a subset of possible topologies for the data in question. For the SH and AU tests these additional trees are necessary to gain an accurate p-value. In theory, every single possible tree topology of the data should be present, to ensure that the 'true' maximum likelihood tree is available to be chosen by the test, and thus provide a baseline (Goldman *et al.* 2000). Indeed, this *a priori* specification of trees is necessary for the appropriate calculation of the null distribution for the test statistic (Planet 2006). However, given that the number of possible topologies increases exponentially with the addition of taxa, this criterion is functionally impossible to meet for most modern studies. As such, a random subset of all possible topologies is chosen instead (*e.g.* see Robinson *et al.* 2005).

Because the SH test assesses whether competing trees are equally likely hypotheses of the data, the choice of dataset will affect the conclusions of the test (*e.g.* it may be expected that for dataset X, tree T_X made from that dataset may be statistically better than another tree T_Y, even if they do not inherently disagree. This would not be evidence for incongruence, just the test behaving in its originally intended manner). Because of this, for each site, reciprocal SH tests were run (Figure 3.2c). For example, for site BBR, the *cob* BBR alignment was used as the base, and all three trees (from the *cob* BBR, ITS2 BBR and psbA^{ncr} BBR alignments) were compared with the SH (both weighted and unweighted) and AU tests. This was then repeated using the ITS2 BBR and psbA^{ncr} BBR alignments as bases to compare the same three trees (Figure 3.2c). 10000 RELL bootstrap replicates were used for calculation of p-values. Because there were six pairwise comparisons carried out for each site (best tree *vs.* other two trees for *cob*, psbA^{ncr} and ITS2 regions), a within-site Bonferroni correction was applied by dividing α by six, meaning p-values were considered significant if less than 0.0085.

3.2.5. *Post hoc* analyses

Based on the original analyses, several datasets displayed consistent evidence of incongruence or contrary patterns (see Results). To verify these results, further ILD tests were executed, using only two gene regions at a time (*e.g.* for a single site, the following concatenations were assembled and tested: *cob vs*. ITS2; *cob vs*. psbA^{ncr}; ITS2 *vs*. psbA^{ncr}; Figure 3.2b). This allowed the location of incongruence to be established (in terms of between-gene regions), as the original ILD tests could not say which partitions were incongruent, only that incongruence existed. An extra site which had consistently shown no evidence of incongruence (LIW) was used as a control.

Following that, the datasets which continued to show incongruence had their ML trees input into Dendroscope 3.0 (Huson and Scornavacca 2012), and pairwise tanglegrams were constructed to identify the source of incongruence (Figure 3.2d). In addition, tree hybridisation networks were created using the Autumn algorithm (Huson and Linz 2016), implemented in Dendroscope 3.0. This algorithm attempts to make a consensus tree from two input trees, and highlights the taxa that cannot be reconciled. This was used to verify the visual inspection of the tanglegrams.
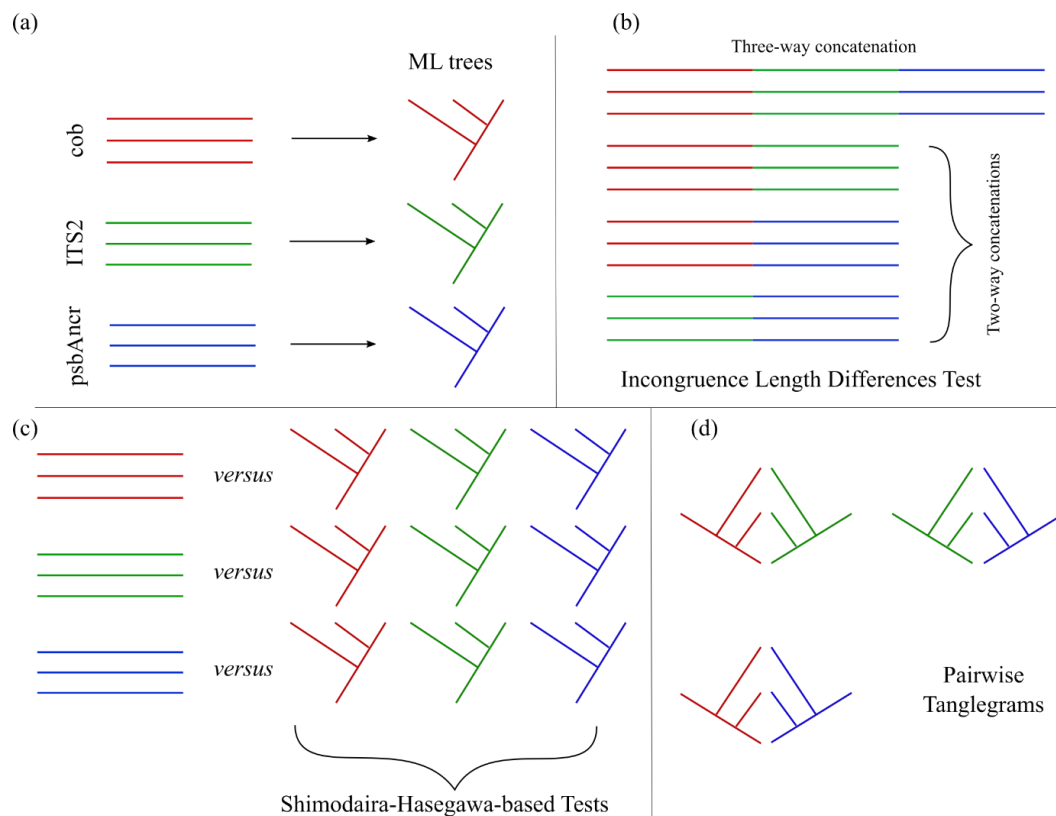
Figure 3.2: Graphical representation of methods, using a single site as an example. (a) Trees were made from the gene regions individually using maximum likelihood (ML) criteria. (b) Data were merged into an overall concatenation for the initial Incongruence Length Difference test, as well as in pairwise fashion for the *post hoc* ILD tests. (c) All three gene trees were tested against each gene region, using Shimodaira-Hasegawa-based tests. (d) Trees from each gene region were compared in pairwise fashion to identify sources of incongruence.

3.3. Results

3.3.1. Incongruence Length Difference test

The Timor sites showed no evidence of incongruence among the *cob*, psbA^{ncr} and ITS2 gene regions, either when considered all together or as separate sites (ILD test, p = 1). This p-value is not concerning; it simply indicates that among the replicates, the partitions were never more parsimonious than random partitions. The Atauro dataset as a whole did not show statistically substantiated evidence of incongruence, though it approached significance (p = 0.07). Looking at each Atauro site individually, BBR and BLS were strongly congruent (p = 0.99), while BHB was close to significance (p = 0.066) and BSP was strongly incongruent between partitions (p = 0.001). This could explain the equivocal result demonstrated by the overall Atauro dataset.

3.3.2. Shimodaira-Hasegawa and Approximately Unbiased tests

In 25 of the 27 tests conducted, the best tree chosen was the one that was made from that gene region originally (*i.e.* for a test with the ITS2 region as its base, the ITS2 tree was chosen as the best tree). The exceptions were the Atauro and Timor datasets, where the psbA^{ncr} and ITS2 trees were chosen as the best explanation of the *cob* dataset, respectively. Generally, the SH and WSH tests failed to find any incongruence. The occasional exception was when the psbA^{ncr} region was compared to the other two. Similarly, the AU test always found incongruence between this and the other two gene regions. This set of results is more likely due to issues with the tests (see Discussion), and therefore the results of the AU test for the *cob* and ITS2 gene regions are the major focus of these results.

There was a very clear island-wide partitioning of results when it came to the AU test. All Timor sites showed no incongruence for either the *cob* or ITS2 gene regions; all three trees (*cob*, ITS2, psbA^{ncr}) did an equally good job of explaining these two regions. This was also displayed for the overall Timor analysis. In contrast, the Atauro dataset showed high levels of incongruence in the AU test. Overall, the ITS2 tree (but not the psbA^{ncr} tree) made from all Atauro samples was incongruent with the *cob* dataset, and both the *cob* and psbA^{ncr} trees were incongruent with the ITS2 dataset. Looking at individual sites, for BBR, BHB and BSP datasets, the ITS2 tree (but not the psbA^{ncr} tree) was incongruent with the *cob* dataset, and for BHB, BLS and BSP, both the *cob* and psbA^{ncr} trees were incongruent with the ITS2 dataset. The three most consistently incongruent datasets in these tests (Atauro, BHB, BSP) correspond to the three lowest p-values returned by the ILD tests.

Taken in broad summary, the AU reveals incongruence between the organellar (*cob* and psbA^{ncr}) and nuclear (ITS2) gene regions. In all cases, the AU test was unable to reject congruence between the *cob*

and psbA[ncr] regions. However, it did reject congruence between the ITS2 and psbA[ncr] regions (using the ITS2 region as a base), and showed reciprocal incongruence between the *cob* and ITS2 region (using both the ITS2 and *cob* regions as a base). See Table A2.3 for all results.

In addition, these results now mean that three datasets (BHB, BSP, and Atauro) displayed incongruence or near-incongruence in both sets of tests. Therefore, these datasets were carried forward to *post hoc* testing.

3.3.3. *Post hoc* analyses

First, additional ILD tests were carried out using two partitions at a time. Two other sites were included: site BLS, which showed incongruence between the ITS2 region and the other two in the AU test, and site LIW, which was included to ensure that the tests still successfully supported congruence where appropriate.

Table 3.1: Results of pairwise Incongruence Length Difference tests. Conclusions were drawn at $\alpha = 0.05$. Bolded values indicate significance.

Dataset	Partitions tested	p-value
BHB	*cob* vs. ITS2	**0.006**
	cob vs. psbA[ncr]	0.847
	ITS2 vs. psbA[ncr]	**0.021**
BSP	*cob* vs. ITS2	**0.011**
	cob vs. psbA[ncr]	0.223
	ITS2 vs. psbA[ncr]	**0.001**
Atauro	*cob* vs. ITS2	**0.01**
	cob vs. psbA[ncr]	1
	ITS2 vs. psbA[ncr]	**0.01**
BLS	*cob* vs. ITS2	1
	cob vs. psbA[ncr]	1
	ITS2 vs. psbA[ncr]	0.778
LIW	*cob* vs. ITS2	1
	cob vs. psbA[ncr]	1
	ITS2 vs. psbA[ncr]	1

These results strongly support the main tests, particularly the AU test. There is clear incongruence between the nuclear ITS2 region and the other two organellar gene regions, which are congruent when considered together. Site LIW is strongly congruent at all regions. This shows that these two-way tests are functioning as expected. BLS is also strongly congruent; while the AU test indicated incongruence, the other tests do not and so it was not carried forward as a candidate for hybridisation.

Tanglegrams were made for BHB, BSP, Atauro, and LIW (as a control), with potentially incongruent branches verified by attempting to hybridise the two trees to create a consensus. The tanglegrams for BHB and BSP are presented below.
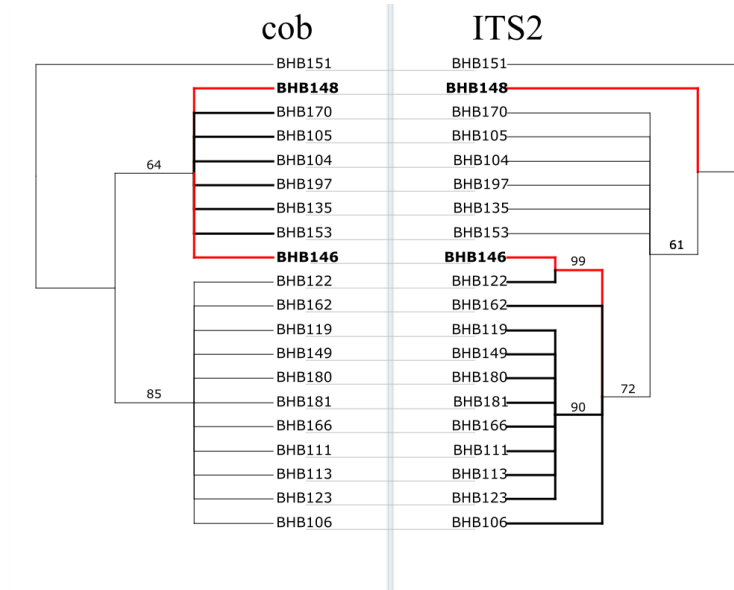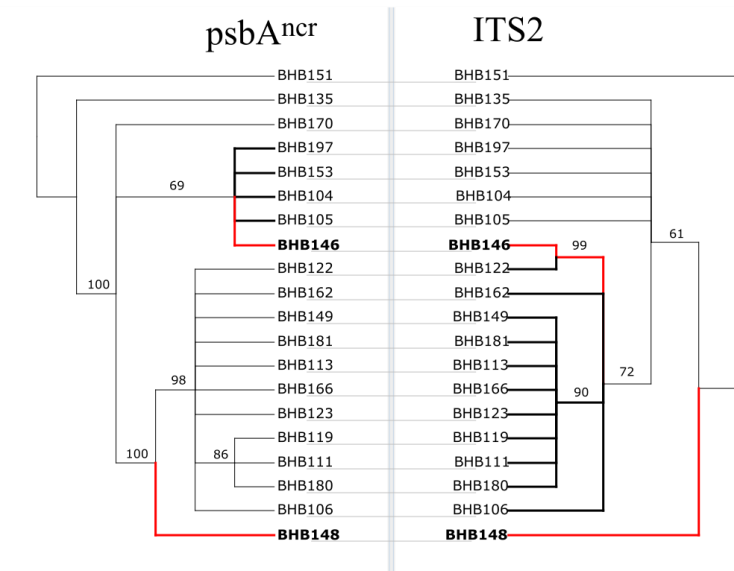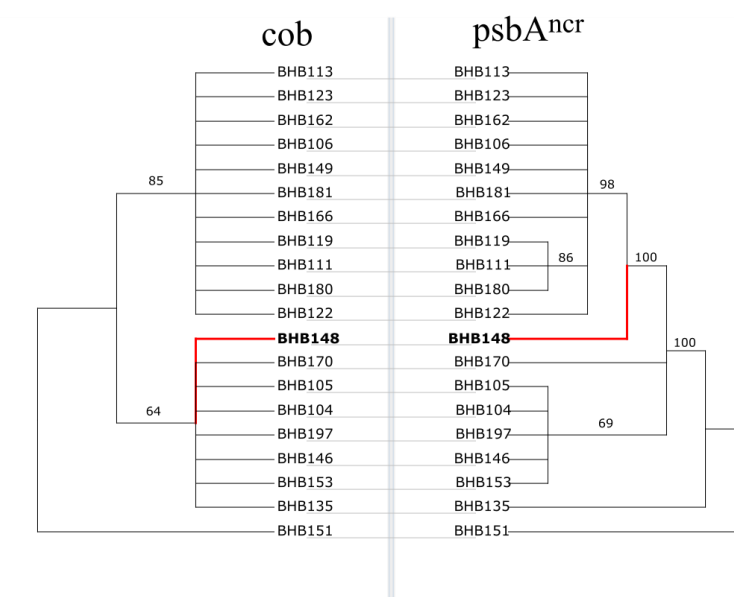
Figure 3.3 (previous page): Pairwise tanglegrams for site BHB. Red branches with bolded taxa labels indicate incongruent samples, while dark black branches designate monophyletic groups to emphasise the differential location of incongruent samples. Branch labels are ML bootstrap values (100 replicates). (a) *cob vs.* ITS2: found incongruent by ILD and AU tests. (b) psbA^{ncr} *vs.* ITS2: found incongruent by ILD and AU tests. (c) *cob vs.* psbA^{ncr}: found congruent by ILD and AU tests.

A visual inspection of the tanglegram for site BHB (Figure 3.3) supports the results of the statistical tests. Comparing the two organellar genes with the ITS2 region (Figure 3.3a, b) reveals two incongruent samples, BHB146 and BHB148, something affirmed by the tree hybridisation analyses, which showed that these two samples prevented a congruent consensus tree. For the organellar gene regions, BHB146 belongs to the C1 subclade (symbiont types C42a and C1v respectively, bolded in the *cob* and psbA^{ncr} trees). For the ITS2 region, it is identified as type C1#, which groups more closely with the C3 group (bolded in the ITS2 trees; see Table A1.3 for why it is called C1# but groups with C3 sequences). BHB148 is identified by the alphanumeric system as a member of the C1 radiation for all three gene regions: symbiont type C42a for the *cob*, and type C1d for the psbA^{ncr} and ITS2 regions. However, C1d appears highly divergent in this case; for the psbA^{ncr} it is a sister to one of the common groups, while it groups as a sister to all other samples in the ITS2 tree.

Interestingly, sample BHB148 is also incongruent between the two organellar genes (Figure 3.3c), despite this incongruence not being detected in the tests. This can be considered positive; it shows that the tests are not functioning in an excessively sensitive manner. This is supported by the LIW tanglegrams (not presented), a site that showed no evidence of incongruence in any of the tests. The tanglegrams and tree hybridisation analysis showed no incongruence at all between the *cob vs.* ITS2 or *cob vs.* psbA^{ncr} regions, but one potentially incongruent sample between the psbA^{ncr} and ITS2 regions. Therefore, there is high consistency between the statistical tests and visual inspection, with some allowance for conservancy on the part of the tests.

Much like the BHB analyses, the BSP tanglegrams (Figure 3.4) support the statistical analyses. The tree hybridisation analysis found no specifically incongruent relationships between the cob and ITS2 regions (Figure 3.4a), however the four bolded samples contribute to generating a reasonably different tree structure. This would support the results of the incongruence tests. In particular is sample BSP387, which is sister to the major C3 group in the *cob* tree, being part of the ambiguous *Symbiodinium* C1/3 group (see Chapter 2) but part of the monophyletic type C40 group in the ITS2 tree, though it is a distinct type called C31 (see Table A1.3). The same pattern is observed in Figure 3.4c, which finds BSP387 to be definitively incongruent between *cob* and psbA^{ncr} sequences, even though the tests did not find these gene regions to be incongruent. However, the incongruence between the psbA^{ncr} and ITS2 regions is very closely supported by the tanglegram (Figure 3.4b), with five samples (bolded) identified as potentially incongruent.
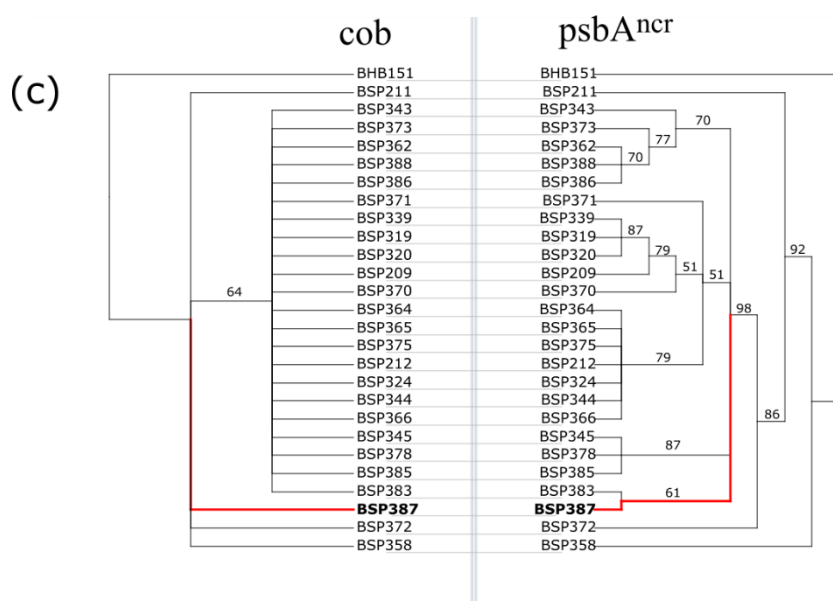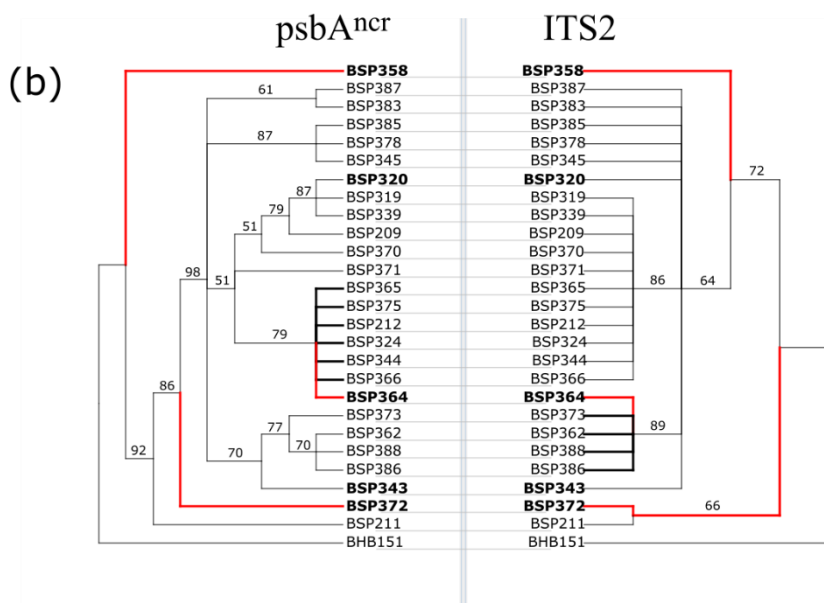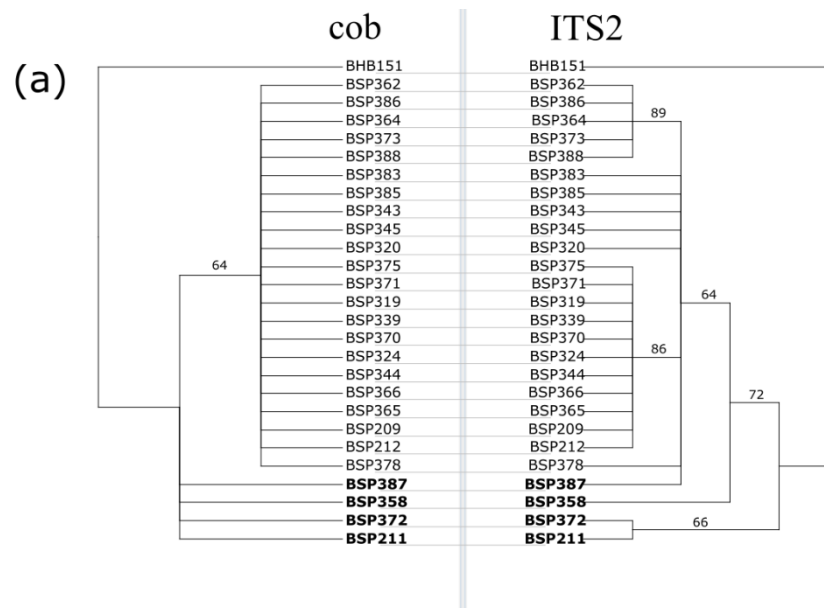
Figure 3.4 (previous page): Pairwise tanglegrams for site BSP. Red branches with bolded taxa labels indicate incongruent samples, while dark black branches designate monophyletic groups to emphasise the differential location of incongruent samples. Bolded taxa without red branches indicate potentially incongruent samples. Branch labels are ML bootstrap values (100 replicates). (a) *cob vs.* ITS2: found incongruent by ILD and AU tests. (b) psbA^{ncr} *vs.* ITS2: found incongruent by ILD and AU tests. (c) *cob vs.* psbA^{ncr}: found congruent by ILD and AU tests.

Of particular significance are the following. BSP364 belongs to two different previously defined subclades: type C40 for psbA^{ncr} (specifically C40i), and type C3z for ITS2. BSP372 follows a very similar trend to BHB148, being designated as symbiont type C1d alphanumerically for both ITS2 and psbA^{ncr}, but grouping with the broad C3 group in the psbA^{ncr} tree and as sister to most other samples in the ITS2 tree (compare Figures 3.3b and 3.4b). Finally, BSP343 shows clear incongruence between the organellar and nuclear genes regions. The psbA^{ncr} is C40l, just a more specific designation of the *cob* C3 (see Table 2.2), which groups it most closely with the C3z clade (Figure 3.4b), while the ITS2 region features type C3u, which places it as distinct from both the C3z and C40 groups.

Given the larger number of taxa, tanglegrams and tree hybridisation analyses for Atauro Island are presented in Appendix A2.3 (Figures A2.1 – A2.3). Their overall results are remarkably consistent with the individual site analyses, and are presented in Table 3.2.

Table 3.2: Summary of incongruent samples inferred from tanglegrams and tree hybridisation analyses. Samples that are bolded indicate they are shared across more than one dataset.

Dataset	Comparison	Incongruent Samples
BHB	*cob vs.* ITS2	**BHB146,** BHB148
	psbA^{ncr} *vs.* ITS2	**BHB146,** BHB148
	cob vs. psbA^{ncr}	BHB148
BSP	*cob vs.* ITS2	**BSP387**
	psbA^{ncr} *vs.* ITS2	BSP320, **BSP343**, BSP358, **BSP364**, BSP372
	cob vs. psbA^{ncr}	**BSP387**
Atauro	*cob vs.* ITS2	**BHB146, BSP387**
	psbA^{ncr} *vs.* ITS2	**BHB146, BSP343, BSP364**
	cob vs. psbA^{ncr}	**BSP387**

Despite the Atauro dataset hosting a diverse range of samples from four sites, the tree analyses showed that incongruence was caused by exactly the same samples as found by the individual site analyses, affirming BHB and BSP as sites with incongruent samples. Further, no other sites contributed any incongruent samples. In addition, Table 3.2 shows that there were always the fewest or equal-fewest incongruences between the *cob* and psbA^{ncr} regions, while the ITS2 comparisons displayed the most incongruence. This is strongly supportive of the AU test results as well as Table 3.1, which all indicate that incongruence largely occurs between the organellar and nuclear genomes of *Symbiodinium*. Of the four clearly incongruent samples (BHB146, BSP343, BSP364, BSP387) there is no general clear pattern in coral host (host genera: *Montipora*, *Pavona, Acropora* and *Symphyllia*).

Finally, these patterns were confirmed by inspecting the raw sequence alignments. An example using sample BSP364 is given in Figure 3.5 below. This confirms that it possesses different genetic signals in the psbA^{ncr} and ITS2 regions.

```
                         255    260     265
                          |      |       |
(a) ITS2    BSP344    TGCTTGCAACTGCTGG
            BSP362    TGCTTGCGACCGCTGG
            BSP364    TGCTTGCAACTGCTGG
            BSP366    TGCTTGCAACTGCTGG
            BSP373    TGCTTGCGACCGCTGG


                        345    350    355    360
                         |      |      |      |
(b) psbA^ncr BSP344   ACGAAAAAGAAAAATA
            BSP362    ATG-AAAAGAAAAAGA
            BSP364    ATG-AAAAGAAAAAGA
            BSP366    ACGAAAAAGAAAAATA
            BSP373    ATG-AAAAGAAAAAGA
```

Figure 3.5: 15 base-pair alignments showing sample BSP364 (nucleotides at polymorphic sites bolded) and a few related sequences. (a) When the ITS2 region is considered, BSP364 is identical to samples BSP344 and BSP366 (type C3z). (b) When the psbA^{ncr} region is considered, BSP364 is identical to BSP362 and BSP373 (type C40i).

3.4. Discussion

3.4.1. True incongruence in *Symbiodinium*?

For a discussion of symbiont hybridisation to be compelling, the first burden-of-proof is to establish genuine patterns of incongruence (*i.e.* not artefacts of the tests). There are many factors, such as character sampling and bias due to differential gene length, which can give false signals of incongruence (Som 2014). Once the limitations of the tests are understood, it is clear that this burden-of-proof has been met in this study.

The SH and AU tests display an issue, with most trees being incongruent for the psbA^{ncr} region. The psbA^{ncr} region is highly variable, and hence a more complex tree is required to explain it. The *cob* and ITS2 trees with multiple polytomies could not do this as effectively, and hence a result of incongruence is returned. Therefore, the *cob* and ITS2 results are likely more reliable. The SH and weighted SH tests also suffer from excessive conservancy (Shimodaira 2002; Planet 2006), observed by their failure to reject congruence in the large majority of cases (the exception was for the Atauro dataset). This can be caused by including unlikely trees in the analysis. While high variance between trees will allow a correct identification of incongruence, extraneous variance (which naturally increases as the number of candidate trees or unlikely trees is increased) will retard the test's ability to detect incongruence

(Kishino and Hasegawa 1989, Strimmer and Rambaut 2002). In contrast, the AU test has been shown to be less affected by these factors (Shimodaira 2002). In the datasets which have broadly been established to be congruent, there was a 61% average reduction in p-value going from the SH to AU test. For the incongruent datasets, this inflates to 94%, which would indicate that this may be the cause of the differences between the test types. As the least biased test, the AU p-values are the most reliable. This test found incongruence between the ITS2 region and the other two regions for the BHB, BSP, and Atauro datasets (Table A2.3). These results are also compelling because they are differential: they show consistently different patterns between datasets and are therefore likely responding to genuine phylogenetic signals.

The ILD analyses provide some of the strongest evidence for incongruence, especially when comparing the two partitions. This test has been well-criticised for being overly sensitive, especially when comparing partitions of different resolutions (Barker and Lutzoni 2002). The refutation of this is simple: in all cases, it found congruence between the $psbA^{ncr}$ and cob regions, the two most different in terms of resolution (Table 3.1), so this is clearly not contributing to the positive results between the organellar and nuclear partitions. This test clearly displays incongruence between the ITS2 regions and other two regions for the BHB, BSP, and Atauro datasets (Table A2.2, Table 3.1).

One discrepancy was observed in the tanglegram and tree hybridisation results: samples BHB148 and BSP372 (both *Symbiodinium* C1d for all regions) were observed as incongruent in the site-wise analyses, but not in the wider Atauro dataset. Given that the different gene regions have the same nomenclature, it appears that in these two cases the incongruence can be attributed to highly divergent evolutionary rates in the $psbA^{ncr}$ and ITS2 regions, which cause resolution differences and group the samples differently in the single site tanglegrams (Figures 3.3, 3.4). However, in the larger Atauro dataset which incorporates more variation, the samples can be better contextualised and are shown to be congruent (Figures A2.1 – A2.3). However, in the other cases the visual tree results affirm the statistical tests.

This leaves the irrefutable conclusion of incongruence between gene regions in *Symbiodinium*. All tests support incongruence between the nuclear and organellar genomes of datasets BHB, BSP, and Atauro. The tanglegrams concur with these results, and consistently recover the same samples implicated in this incongruence: BHB146, BSP343, BSP364, and BSP387 (Table 3.2). Incongruence is also occasionally recovered *via* visual inspection between organellar gene regions, but this is not statistically supported. Generally, the organellar genes are congruent, or at least significantly more so than between the organelles and the nucleus. Given that only a quarter of the samples causing incongruence overall belong to vertically transmitting hosts (one *Montipora* species) it does not appear that there is any particular link between coral transmission mode and potential hybridisation, at least in this study.

3.4.2. Hybridisation in *Symbiodinium*?

While incongruence has been comprehensively established, there are many factors other than hybridisation which can cause it. One hypothesised to be quite common but insidious in its undetectable nature is heterotachy, shifts in site-specific evolutionary rates through time (Som 2014). For example, Lopez *et al.* (2002) showed that, for the *cob* gene, 95% of variable sites have inconstant substitution rates through time in vertebrate lineages. While there is no particular way to identify heterotachy or exclude it as a cause, except with a very large number of sequences, maximum likelihood methods in particular have been shown to be robust to even intermediate levels of heterotachy (Som 2014).

A more plausible explanation is incomplete lineage sorting (ILS), often considered the most common cause of incongruence (Degnan and Rosenberg 2009). This is due to polymorphisms not segregating fully during speciation events, leading to phylogenetic signals in gene trees that conflict with the overall species tree. This has been shown to be quite common in the ITS2 region, thanks to its multiple-copy nature (Thornhill *et al.* 2007, see also Figure 1.4). Through this mechanism, ancestral polymorphisms may persist at low levels in the genome. The results observed in Chapter 2, with C3 sequences occurring at background levels in a large number of samples, are suggestive that some of the sequences recovered may be intragenomic variants in a single cell, rather than an independent population of symbionts. Therefore, it is possible that the divergent sequences recovered actually represent a single symbiont population, which has multiple ancestral polymorphisms present *via* ILS. Through stochastic DNA processes such as unequal crossing over, slipped-strand mispairing and transposition, these intragenomic variants may be eliminated or promoted in the multiple-copy array (Nei and Rooney 2005). Hence, in the samples from a single population, one ancestral polymorphism may be dominant in the ITS2 region of some, while a different ancestral polymorphism may be dominant in others. This would cause the patterns observed in this study, with the ITS2 region being occasionally incongruent with the organellar regions.

Ideally, a statistical test would be carried out to differentiate between hybridisation and ILS, and such tests do exist. However, they require inputs of information which are not currently available for *Symbiodinium*, such as: (a) An understanding of the effective population size N_e (Pelser *et al.* 2010); (b) a large number of genes, at least some of which must be adjacent (Pollard *et al.* 2006; Meng and Kubatko 2009); or (c) strictly bifurcating trees and clearly defined species (Sang and Zhong 2000; Joly *et al.* 2009). Therefore, ILS as a cause of the observed incongruence cannot be statistically refuted. However, there is good evidence that the patterns observed here are more likely to be caused by symbiont hybridisation.

First, the pattern of incongruence observed, with organellar cytoplasmic genes being different to nuclear genes, accords with a large body of prior theory on hybridisation. Nuclear genes are largely inherited biparentally, and the ITS2 region is no exception (Baldwin *et al.* 1995; Rybalka *et al.* 2013). However, the cytoplasm tends to be inherited maternally (Rieseberg *et al.* 1996). This difference is largely due to gametogenesis and fertilisation, where the male gamete typically only contains nuclear information, while the female gamete (egg) contains the cytoplasm that will be passed on to the zygote. Therefore, if an organism encounters a population of another species and produces viable hybrids, theory predicts that over time, repeated backcrossing with the more common species (introgression) will produce hybrids with divergent organellar and nuclear signals (Figure 3.1).

While the nature of the sexual life cycle has yet to be fully elucidated in *Symbiodinium*, previous evidence has shown that other unicellular dinoflagellates produce gametes (Brawley and Johnson 1992). In addition, the presence of 'plus' and 'minus' mating types, analogous to gender, has been shown in the dinoflagellate *Alexandrium tamarense* (Brosnahan 2011). Therefore, it is reasonable to assume that *Symbiodinium* also produce distinct gametes (as opposed to conducting sex *via* fusion, for example), making this mechanism eminently plausible. The documentation of functional meiotic genes in *Symbiodinium* (Chi *et al.* 2014, Levin *et al.* 2016) supports this assertion. Such a pattern of discordance between cytoplasmic and nuclear genes caused by hybridisation has been recorded for taxa as diverse as plants (Rieseberg *et al.* 1996; Pelser *et al.* 2010; Sun *et al.* 2015), beetles (Sota and Vogler 2001) and indeed corals (van Oppen *et al.* 2001b). In general, hybridisation is predicted to cause incongruence between nuclear and cytoplasmic markers in both multicellular and unicellular taxa (Bull *et al.* 1993). Other factors due to hybridisation, such as semigamy or differential fitness of nuclear-cytoplasmic combinations, can also cause incongruence between gene nuclear and cytoplasmic gene trees (Rieseberg *et al.* 1996). Therefore, the fact that this was the pattern observed in this study is strong circumstantial evidence that hybridisation is the explanation.

In addition, hybridisation is made more likely in comparison to ILS by the fact that all of the incongruent ITS2 sequences were previously defined types (*i.e.* not unique sequences), that were also present in non-incongruent relationships in the analyses. One example from the results will serve to demonstrate this. BSP364 had a generic *Symbiodinium* C3 sequence for the *cob* gene, was a C40 type for the psbA[ncr] region, and C3z for the ITS2 region. Significantly, there were also samples recovered which were type C40 for both the psbA[ncr] and ITS2 regions (bolded branch group for psbA[ncr] in Figure 3.4b, and matching samples in the ITS2 tanglegram), and samples which were type C3z for both regions (bolded branch group for ITS2 in Figure 3.4b and corresponding samples in the psbA[ncr] tanglegram). This confirms that they are clearly separate types, supported by the fact that they differ by four base pairs in the ITS2 sequence and 64 base pairs in the psbA[ncr] region (including a 49 base pair deletion in the C40 sequences), indicating that this is not just a non-diagnostic polymorphism (Wilkinson *et al.* 2015). The

implications for this being caused by ILS are given in Figure 3.6. Only the psbA^{ncr} and ITS2 genes are presented, as the *cob* gene was invariant in this case.
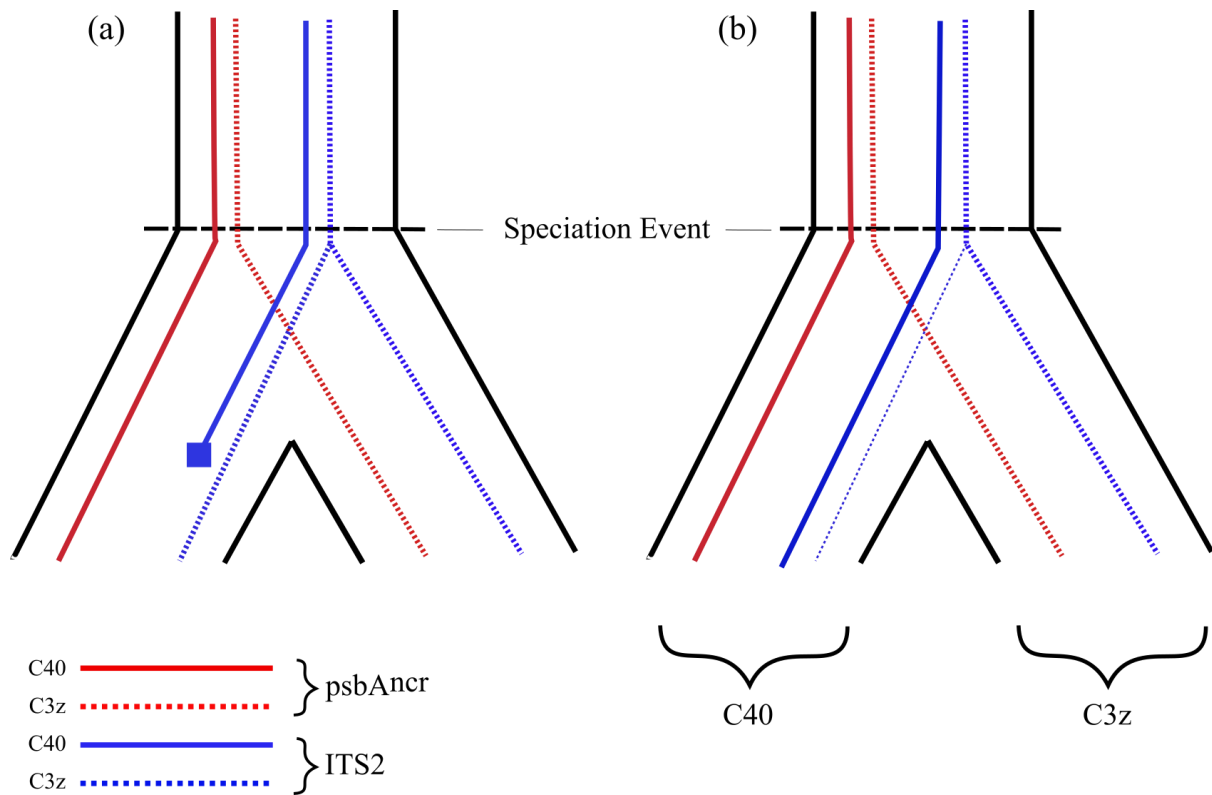
Figure 3.6: Predictions under Incomplete Lineage Sorting. (a) General pattern expected for ILS. A single ancestral population with polymorphism in both the psbA^{ncr} and ITS2 regions is present before a speciation event. After speciation, the ITS2 polymorphism fails to segregate, while through stochastic processes the C40 polymorphism is eliminated and leads to incongruence between nuclear and chloroplast genes. (b) The process of ILS that would be required for this example. The ITS2 region fails to segregate after speciation; despite the extensive presence of C40 alleles, a small subpopulation of symbionts with dominant C3z alleles is maintained (weak dashed blue line) in the C40 population and both are recovered in present-day sampling, at the same site, as pure C3z populations.

Figure 3.6b graphically represents the process that would be required for the observed patterns to be due to ILS. Given that symbiont sex is now strongly supported (Thornhill *et al.* 2017), it seems unlikely that a divergent ancestral polymorphism could be maintained as the *dominant* sequence in some samples within type C40, as it would be expected that repeated recombination would eventually remove C3z traces from the C40 genome, or *vice versa* (Figure 3.6a). It would appear a more reasonable conclusion that a hybridisation event has occurred between symbiont types C40 and C3z, *via* a mechanism such as that presented in Figure 3.1, leading to incongruence between organellar and nuclear genes. In addition, the NGS data from Chapter 2 further challenge ILS as the cause of incongruence. If it was to be caused by ILS, then *both* variants would be expected to occur in the ITS2 region of the samples (with one at low frequency), but this was not consistently observed. For example, sample BSP343 was identified as *Symbiodinium* type C40 for the organelle regions, and type C3u for ITS2. However, the NGS data

revealed no trace of type C40 in that sample for the ITS2 region, contrasting with the prediction of ILS. In fact, in only two of the four cases were both the divergent types observed in the ITS2 region of the sample, suggesting that ILS is a less plausible explanation than hybridisation. In addition, the divergences observed (*i.e.* C40/C3u, C40/C3z *etc.*) all coalesce at the 'ancestral' types C1 or C3, rather than one representing an intermediate evolutionary step to the other (see Figure 2.3). Therefore, ILS would also predict these ancestral sequences to be in the ITS2 genome in low frequencies. However, once again, in only two of the four cases was this observed (and the two samples in which it was observed are not the same two samples as above).

Finally, even if ILS was operating on these samples, in one case there would still be incongruence present. Sample BSP387 has ITS2 type C31 as the dominant sequence, which is incongruent with the *cob* gene identity C1/3. The most common ITS2 background populations in this sample are C3d/C21 and C15, which would also be incongruent with the *cob* gene. Therefore, even if the C31 repeat has risen to ascension through random processes associated with ILS, the sample would still display incongruence. This is strong evidence that the patterns observed are more likely to be due to genuine hybridisation. While it is acknowledged that hybridisation and ILS are not mutually exclusive and the incongruences observed could be caused by a combination of both, the weight of evidence suggests that these results are more likely a result of interspecific hybridisation between distinct symbiont types.

Finally, an appeal to basic common sense can be made. ILS, and indeed most other analytical factors, are random or would be expected to affect all sites. The results obtained, however, are anything but random, with two sites consistently being recovered as incongruent in contrast to all others, despite those incongruences coming from a range of host species. The fact that they both came from Atauro Island, with its high diversity when compared with the Timor sites (see Chapter 2) speaks to the fact that this is a genuine and significant pattern. While ILS cannot be categorically dismissed, all available evidence suggests that hybridisation is occurring at Beloi Harbour and Beloi Saddle Patch (Right).

3.4.3. Synthesis with previous research

No previous study on *Symbiodinium* seriously considers symbiont hybridisation, except that of Wilkinson and colleagues (2015), which also finds evidence for its existence. However, aside from the potential examples of hybridisation mentioned in the Introduction of this chapter (LaJeunesse *et al.* 2003, 2004a; LaJeunesse 2005), three other studies bear mention.

Sampayo *et al.* (2009) also focused on the basis that hybridisation can cause incongruence between genes from different organelles, and built trees from mitochondrial, chloroplast and rDNA nuclear gene regions to test this. Based on these trees, they concluded that different symbiont lineages (types) within

clade C are reproductively isolated. However, this may be an overly general conclusion considering that only three species of pocilloporid corals were sampled (*Pocillopora damicornis*, *Stylophora pistillata* and *Seriatopora hystrix*). Further, this was done largely on the visual inspection of trees, and they used the cp23S chloroplast gene region, which is less sensitive than the psbA^{ncr}. Interestingly, they did also use the ILD test to formally test incongruence, which returned a p-value of 0.01. The authors recognised that this signifies incongruence, but then go on to say that this supported the concatenated analysis. This is a misunderstanding, as incongruence implies conflicting genetic signals in the data which invalidate a holistic grouping. Therefore, this study did not definitively disprove hybridisation, and may have actually provided some evidence towards it.

Pochon and colleagues (2014) assessed six genes from three different organelles (mitochondrion, nucleus and chloroplast). In all cases, they found evidence of incongruence between pairwise comparisons of genes, using the AU test. While they go on to discuss the implications for concatenation in some detail, the cause of these incongruences was never attributed or discussed in detail again. Because only one test was attempted, and given the acknowledged limitations of incongruence tests, it is difficult to draw confident conclusions from that result alone, something which also demonstrates the utility of the approach taken in this chapter. However, that study (Pochon *et al* 2014) adds to the body of evidence that the genus *Symbiodinium* has not evolved in a simple linear fashion.

Another study from Pochon *et al.* (2006) found the surprising result of incongruence between whole clades rendered from nr28S and cp23S data, using the SH test. However, when they removed all but two members of each clade, the test then showed congruence between datasets. This indicated incongruence was being caused by the accumulation of within-clade mismatches between the nucleus and chloroplasts, something which is also broadly agreeable with a hypothesis of hybridisation.

These studies can hardly be considered to provide convincing evidence of hybridisation. However, it is reasonably striking that four studies conduct an explicit statistical test of incongruence within *Symbiodinium* (Pochon *et al.* 2006, 2014; Sampayo *et al.* 2009; this study), and all four find evidence for its existence. At the very least, this justifies a more careful consideration of patterns of incongruence within this genus, and its potential causes.

3.4.4. Conclusions

This study cannot be considered unequivocal evidence for hybridisation within *Symbiodinium*. However, it takes a significant step towards justifying an in-depth exploration of genetic incongruence between organelles in coral symbionts. Given the multiple-copy nature of the ITS2 region, corroboration with an additional nuclear marker would lend further weight to these results, and is

recommended for future studies. In addition, the clear differences in instances of putative hybridisation between Atauro and Timor sites suggest some external, environmental influence on its occurrence. This will be explored in Chapter 4.

All incongruence tests have flaws, largely because none of them were originally designed to test incongruence, just whether certain phylogenetic hypotheses fit data better. However, the use of multiple tests and a thorough understanding of their flaws allows for a certain level of confidence in their results. When they agree with both each other and a visual representation of the data, as they do here, incongruence can be assuredly stated. Despite the clear advantages of this approach, few studies use multiple tests or optimality criteria, or consider the statistical basis of the tests. This leaves many studies drawing conclusions with confidence despite unnecessarily equivocal results.

In this study, there is clearly evidence for incongruence between the ITS2 region of the symbiont nucleus and the two organellar gene regions, psbA^{ncr} and *cob*. This pattern conforms to the hypothesis of hybridisation between divergent taxa. While incomplete lineage sorting remains a possibility, it is a less intuitive explanation, especially in the light of incongruent samples having clearly distinct, predefined types which were recovered in non-incongruent samples, and the failures of background populations to consistently align to its predictions. Therefore, hybridisation appears to be a credible mechanism for adaptive change in *Symbiodinium*. Ascertaining the frequency and extent of this may be vital to predicting the fate of coral reefs in an environmentally unpredictable future.

Chapter 4: General Discussion

The preceding chapters in this thesis have explored the diversity of symbiont populations at Atauro Island and nearby Timor sites. Chapter 2 demonstrated that both geographic regions host highly diverse symbiont communities, in patterns that generally conform to those of the Indo-Pacific region. For example, a few host-generalist symbionts such as C40, C1 and C1x are present, as well as several conserved, specific relationships such as *Porites* and *Montipora* associating with C15p and C15q, respectively (Section 2.3.5). However, there were also several novel relationships and symbiont types documented, as well as a large community of cryptic background symbionts within each coral colony (*e.g.* Table 2.3). These differences were also observed over more local geographic scales, with Atauro Island having demonstrably higher diversity than Timor (Figures 2.2 and 2.6), something which cannot be attributed to differential host sampling or other analytical factors. This extends to an analysis of potential hybridisation. Chapter 3 provides reasonably strong evidence for non-linear evolution in *Symbiodinium*, but only at Atauro Island. The identity of putative hybrid samples, as revealed by their incongruence in different gene trees, was found to be diverse. While the cob and psbA[ncr] identities were generally common symbiont types (C40 or C1), the ITS2 identities belonged to the less common types C1#, C3z, C3u and C31, though it is significant that all of these are previously defined types. While other explanations such as incomplete lineage sorting cannot be eliminated, the observed patterns align very closely with theoretical predictions for the effect of hybridisation, and this is therefore invoked as the most plausible reason for the incongruences found.

This discussion will begin with a brief consideration of the methods used in this study. Subsequently, the potential reasons for the patterns of prospective hybridisation are explored, informed by the symbiont diversity of Atauro Island and Timor. Following this, the potential implications of hybridisation for evolutionary advancement are discussed, before concluding with a section on considerations arising from this thesis.

4.1. Development of methods and future recommendations

This study displays a novel application of Next Generation Sequencing (NGS) with the Illumina platform, multiplexing over 380 samples on a single lane. This made costs comparable to standard PCR while significantly increasing sequencing depth. Given the high prevalence of background symbiont populations (Silverstein *et al.* 2012; Kennedy *et al.* 2015), this approach clearly has high efficacy for *Symbiodinium* diversity studies and is strongly encouraged for future studies. In addition, this study also utilised markers from multiple organelles. While this is becoming increasingly common (*e.g.* Sampayo *et al.* 2009; Pochon *et al.* 2014), in general it is only to 'back up' the results of other genes, and

differences are treated as a problem (*e.g.* procedural or analytical mistakes) that should be eliminated (Leigh *et al.* 2011; Som 2014). However, this study proves that, in many cases, differences between the gene regions are just as informative as the information encoded in the regions themselves. For example, these differences have facilitated the identification of co-dominant symbiont populations and nomenclature issues (Chapter 2), as well as provided strong evidence of non-linear evolution within *Symbiodinium* (Chapter 3). Therefore, it is strongly recommended that future studies treat the small incongruences between different gene regions and organelles not as a problem, but as an opportunity to further understand evolution and diversity within coral symbionts.

However, there are still issues with the interpretation of genetic data in this study, particularly concerning the multiple-copy nature of the ITS2 region. While the OTU-based approach is designed to limit the impact of intragenomic variation, it can still reduce confidence in the results. Ideally, other gene regions which are less afflicted by intragenomic variation should also undergo NGS to increase sureness in attributing background populations. However, this is limited by the fact that the Illumina MiSeq platform still only supports sequence reads of 300 bp or less (illumina.com/systems/sequencing-platforms/miseq/specifications). The limited number of currently-developed markers for *Symbiodinium* (see Chapter 1) means that there are very few options for an NGS-based verification of ITS2 patterns. Therefore, a priority should be the development of additional markers <300 bp for *Symbiodinium*, particularly in the mitochondrial and chloroplast genomes. If these markers are then utilised in the manner described in Chapters 2 and 3, thorough, defendable and wide-ranging studies of symbiont diversity and evolution can be carried out at costs scarcely greater than those currently incurred.

4.2. The observed patterns of putative hybridisation

The following two subsections (4.2.1 and 4.2.2) expand upon the discussion in Chapter 3, and provide further evidence and explanations for hybridisation by incorporating the analyses in Chapter 2.

4.2.1. Does hybridisation make sense in terms of the observed diversity?

Chapter 2 demonstrated the possibility that different gene regions may identify different symbiont populations, due to co-dominance inside a coral host (see Section 2.2.6). It is plausible that such a mechanism could cause the incongruence here. For example, sample BHB146 is a putative hybrid between *Symbiodinium* types C1# and C1, and while the dominant ITS2 type was C1#, the OTU data revealed that C1 was present at background frequencies in the sample. Potentially, there were actually two clear symbiont populations, C1# and C1, which were co-dominant and reproductively isolated. However, overall the OTU data suggest that this explanation can likely be dismissed. On average, the most common ITS2 sequence found in the putative hybridising samples was 2.25 times more common

than the next sequence in the sample (range 1.91 – 2.55). Of the 377 samples included in the OTU analysis, 122 had a smaller difference than this between their first and second most common sequences. Even if the smallest difference in the potential hybridising samples is considered (most common sequence 1.91 times greater than second most common), there are 101 samples in the dataset with a difference less than that. If co-dominant symbiont populations were the explanation for this, it would be expected that a much greater proportion of the samples across all sites would have displayed these incongruent patterns, rather than the four observed at two sites only. In addition, only in two of the four cases (samples BHB146 and BSP364) was the other possible parent type found among the background ITS2 types. Therefore, the within-host diversity does suggest divergent parent *Symbiodinium* taxa for a single symbiont population.

The assertion that hybrids are present would be made more compelling if there was evidence of both 'pure' parent populations also present at the sites. For most examples, this does indeed appear to be the case. All of the putative hybrids have a reasonably common *Symbiodinium* type as one of their parents (either C40 or a C1 variant), which were frequently found in congruent relationships. The other possible parents were all comparatively rarer (ITS2 types C3z, C1#, C3u and C31). However, they are also well-established as independent evolutionary lineages. They all appeared as the dominant sequence, an important criterion for establishing evolutionary divergence (Smith *et al.* 2009). In addition, they are all types that have been previously reported and established as independent (van Oppen *et al.* 2011b; Zhou and Huang 2011; Chauka 2012; Franklin *et al.* 2012; Baker *et al.* 2017), and therefore their candidacy as hybrid progenitors is well supported. Symbiont types C3z and C1# also appear in congruent relationships multiple times, which is strongly supportive of their presence as distinct populations (as opposed to intragenomic variants). Of course, this assumes reasonably recent hybridisation as the cause of incongruence – it is possible that the hybridisation events in the more distant evolutionary past could have caused the incongruence. Therefore, the clear attribution of distinct parent types is not necessary for proof of hybridisation, though the fact that there appear to be in several cases here encourages the hypothesis of recent hybridisation at these sites.

4.2.2. What explains the differential patterns of putative hybridisation at different sites?

4.2.2.1. Can differences in observed cases of putative hybridisation be explained by symbiont diversity?

The symbiont diversity at each of the sites provides a good explanation as to why putative hybridisation was only observed at two Atauro sites. As previously established, the hypothesised *Symbiodinium* parent populations for the hybrid samples feature some uncommon symbiont types (ITS2 types C3z, C3u, C31, and C1#). They only appear in samples 36 times (including dominant and background occurrences). Of those 36 samples, 31 of them were collected at Atauro Island; the only time they

appeared at Timor sites was type C1# at background levels at Lamsana Inlet East. Therefore, this appears to be a reasonable explanation for the differential pattern of possible hybridisation displayed, at least between Atauro and Timor sites – the lowered diversity means that there are simply fewer symbiont populations that could hybridise.

This conclusion operates under the assumption that hybridisation is a purely chance event, and the more closely related types that coexist, the greater the probability of an interspecific mating event occurring. This model has support in the literature, with other studies finding evidence that hybridisation is attributable purely to chance events associated with co-existence of closely related species (*e.g.* Nolte *et al.* 2005; Yaakub *et al.* 2007). This is also supported in a slightly different way by comparisons with the two studies of Pochon and colleagues (2006, 2014), which also found statistical evidence of incongruence. In both cases they were meta-analyses, which drew symbiont samples from a wide array of locations. Therefore, the wider sampling encompassing high levels of diversity allowed for the detection of potential hybridisation, which supports it being a chance event driven by the presence of increased genetic variability. In contrast, studies on plants have suggested that hybridisation is not due to chance, but degrading environmental conditions, which causes species to shift ranges or decrease in population size (Rhymer and Simberloff 1996; Moran and Alexander 2014). This serves to increase the likelihood of sexual contact between two species, because with lowered population size the chance of encountering a different species (*versus* a conspecific) increases. However, this appears to be a less plausible explanation here, as the geographic region that is arguably more environmentally stressful (Timor, see Section 4.2.2.3) provided no evidence of hybridisation. It appears more likely that high diversity facilitates the potential for symbiont hybridisation. As only seven sites were assessed in total in this study, it is difficult to state this with certainty, but it represents a testable hypothesis for *Symbiodinium* that should be explored further.

Interestingly, the cryptic types C3u, C31, C3z and C1# are not just rare at Timor, but throughout the Indo-Pacific. In all four cases, this is the first time that they have been reported inside the Coral Triangle (Figure 4.1). Obviously, it is unlikely that they only appear inside the Coral Triangle at Atauro Island; it is more reasonable to assume that they exist at other sites, but remain undetected due to incomplete sampling. However, this still suggests that they are uncommon types, and therefore potential hybridisation may be limited by their absence elsewhere. Therefore, establishing the possibility of hybridisation may be studied further by accounting for why these rare types are absent elsewhere.

Figure 4.1: All previous reported locations of the four rare putative hybridising types in this thesis. Note their complete absence from the Coral Triangle (blue). Data from Smith *et al.* (2009); LaJeunesse *et al.* (2010b); Zhou and Huang (2011); van Oppen *et al.* (2011b); Chauka (2012); Franklin *et al.* (2012); Yang *et al.* (2012).

4.2.2.2. Can differences in observed cases of putative hybridisation be explained by host diversity or under-sampling?

One of the most plausible reasons for different symbiont communities in different geographic regions is the coral species found there (*e.g.* LaJeunesse *et al.* 2004a, Kemp *et al.* 2015). However, this does not appear to be the explanation here. The potentially hybridising symbiont types were found in a broad range of host genera at both dominant and background levels (Table A1.5). For example, *Symbiodinium* type C3z was found in *Acropora, Favia, Favites* and *Montipora*, among other genera, which were common at all sites in both geographic regions. The only potential exception is *Symbiodinium* C3u, which was found in the genera *Sandalolitha, Ctenactis* and *Symphyllia*, which were rare at all sites. However, in general there was no clear pattern evident, which suggests that the absence of rare symbiont types is not due to different coral species at Timor sites.

Given that putative hybridisation appears to involve rare types in this study, it is possible that there was simply not enough sampling to detect hybrids at all sites. This explanation could perhaps explain the lack of evidence for hybridisation at Beloi Barrier Reef and Beloi Lagoon South at Atauro Island – most of the rare types were found at these sites, and it is possible that if more sampling was done, there would have been incongruent samples there too. However, this explanation is less likely for Timor. Calculating the proportion of potential hybridising types *per* genus *per* sample at Atauro Island and extrapolating this to Timor, the four rare types would still be expected to occur ~23 times, a number far more than the five observed. Therefore, the lowered presence of putative hybridising types at Timor is not adequately explained by low potential for discovery.

4.2.2.3. Can differences in observed cases of putative hybridisation be explained by environmental conditions?

Chapter 2 hypothesised that the differences in diversity between the two geographic regions were explained by environmental conditions. In light of the evidence presented in Chapter 3 as well, and in the absence of other explainable factors, this appears to be the most likely mechanism for the observed differences. This is possibly temperature-related: Turak and DeVantier (2013) reported an average of 26.3ºC at Atauro sites (range 25ºC-27ºC) and 27ºC at Timor sites (range 26ºC-28ºC). While this difference is reasonably small, often temperature increases of as little as 1ºC can stimulate bleaching (Hill *et al.* 2014). Therefore, temperature could potentially have a role in the patterns observed. In addition, there were clear differences in turbidity (as assessed by water visibility) between the sites, with Timor being much more turbid (Turak and DeVantier 2013). As phototrophs, light is one of the most important resources for *Symbiodinium*, and plays a large role in structuring symbiont communities (Ulstrup and van Oppen 2003; Bongaerts *et al.* 2015a). As such, turbidity has a major influence on symbionts, having the potential to depress their photosynthetic rates (Fabricius 2005) and place their coral hosts under significant physiological pressure (Pollock *et al.* 2014). Some symbionts have the ability to photo-acclimate, or to maintain photosynthetic capacity under lowered light conditions (Ulstrup *et al.* 2008; Browne *et al.* 2014); it is therefore reasonable to assume that, given the opportunity, corals would preferentially associate with these types in suboptimal light conditions. In the case of both increased turbidity and increased temperature, clade D symbionts have generally been shown to be more resilient, and are the homologous type in corals under such conditions (*e.g.* LaJeunesse *et al.* 2010a, 2010b, 2014; Ladner *et al.* 2012; Stat *et al.* 2013; Wham *et al.* 2017). Significantly, Chapter 2 found a greater proportion of clade D symbionts at Timor sites; it may be that these symbionts bring greater benefit to coral hosts under the slightly warmer, more turbid conditions at this location. In addition, two of the most common clade C symbiont types observed at Timor sites, C40 and C15, are ranked as the 7th and 17th most thermally tolerant symbiont types in a recent meta-analysis (Swain *et al.* 2017). In contrast, type C1 (which replaces D1 in *Pocillopora* at Atauro Island) is ranked 42nd out of the 64 types assessed. Therefore, the symbiont composition observed at Timor sites is strongly supportive of this location generally being more 'stressful'.

In contrast, at least three of the rare potential hybridising types in this study (C3u, C3z, and C31) have been shown to be intolerant of environmental stress (Chauka 2016; Swain *et al.* 2017). This may explain their absence at Timor sites (and indeed elsewhere in the world): there may be negative selection for suboptimal or rare types, as they do not contribute to a maximal photosynthetic rate under lowered light or increased temperature (Suggett *et al.* 2008; Ainsworth *et al.* 2016). This selection pressure does not appear to be operating at Atauro Island to the same extent. The results in Chapter 2 are strongly supportive of this, with a greater proportion of rare clade C types at background levels at Atauro sites

(Figure 2.6b). This leads to the conclusion that hybridisation may only be possible (or at least far more likely) at high-quality reef sites, with high levels of symbiont diversity. Other studies have also recorded higher levels of symbiont diversity at higher-quality sites (*e.g.* Howells *et al.* 2009). This accords with the result of Pochon and Pawlowski (2006), who found that rapid diversification in *Symbiodinium* occurred in relatively cooler periods in geological history. Potentially, these periods with reduced heat stress released corals from the pressure of hosting only optimal symbionts, and promoted diverse symbiont assemblages that potentially contributed to the bursts of diversification *via* hybridisation. In addition, this suggests that rare symbiont types may have an important role to play in symbiont communities, if they can lead to highly stress-tolerant hybrids (Figure 4.2). The efficacy of this mechanism is discussed below.
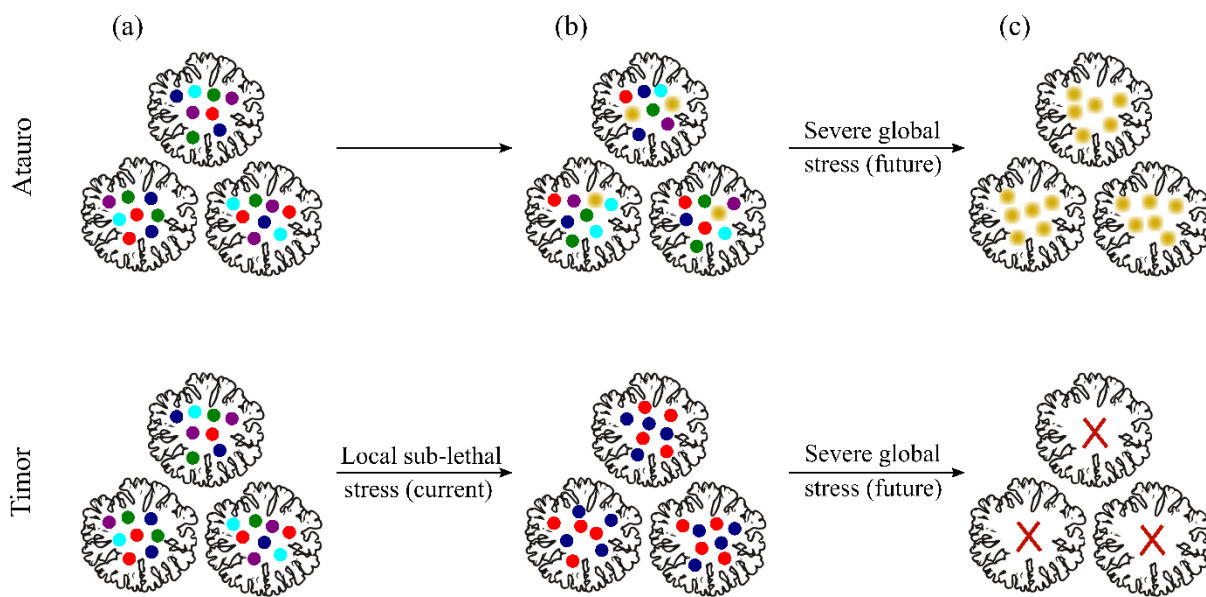
Figure 4.2: Conceptual diagram of the hypothesis advanced in this chapter. (a) Original high levels of symbiont diversity (coloured circles) in coral populations in both geographic regions (Timor and Atauro Island). (b) Local environmental stress serves to reduce diversity and therefore hybridisation potential at Timor sites, while maintenance of diversity at Atauro allows occasional production of (golden) hybrid types. (c) Hybrids may have higher realised fitness under future stress scenarios, allowing coral persistence at Atauro but death at Timor. Coral outline from Wilkinson *et al.* (2015).

4.3. Evolutionary benefits of hybridisation

Evidence suggests that hybridisation may be increasingly important as an evolutionary mechanism for coral reefs. Acclimation, where occasional heat-stress events allow a tolerance to warming oceans to be generated, has been largely discredited as a method by which corals could survive into the future (Ainsworth *et al.* 2016). While most bleaching events have been characterised by the occasional well-adapted host-symbiont combination surviving and proliferating (*e.g.* Jones *et al.* 2008), conditions are changing too rapidly for this to sustain coral reefs (Baskett *et al.* 2009). It has even been hypothesised

that the adaptive limit of coral reefs has been reached (Császár *et al.* 2010), at least through the standard mechanisms of somatic mutation and intraspecific sexual recombination.

However, hybridisation is a highly credible mechanism for the appearance of new forms (Dittrich-Reed and Fitzpatrick 2013). It facilitates rapid evolutionary change, by producing high levels of variation at thousands of genes *per* generation (Rieseberg *et al.* 2003). In many cases, hybridisation can lead to inviable or low-functioning combinations, and many authors express doubt as to the evolutionary validity of this mechanism (*e.g.* Brosnahan *et al.* 2010, see also review by Arnold and Martin 2010). However, it is generally recognised that hybrids exist on a continuum of significantly depressed to highly elevated fitness, even for different hybrids in the same genus (Schardl and Craven 2003; Hamilton *et al.* 2010; Dittrich-Reed and Fitzpatrick 2013). Most importantly, hybridisation has the ability to produce extreme phenotypes, through transgressive segregation. This is where favourable alleles from both parents combine to produce significantly elevated hybrid fitness, relative to either parent (Rieseberg *et al.* 2003). For example, the hybrid sunflower *Helianthus paradoxus* is adapted to extreme conditions such as highly saline soils, in contrast to its two progenitors *H. annuus* and *H. petiolaris* (Rieseberg *et al.* 2003). This is significant as it can promote species that will survive in highly stressful conditions, even if the parent species do not live in those conditions.

This is eminently possible for *Symbiodinium*, especially as there are three separate reservoirs of genetic material to act on (mitochondrial, chloroplast and nuclear genomes). It has been shown that cytoplasmic genomes have a significant role to play in adaptive potential (Leinonen *et al.* 2011; Bock *et al.* 2014), and different cytoplasmic-nuclear combinations can display novel functions (Derks *et al.* 1992), which brings significant evolutionary potential to the different combinations observed in Chapter 3. For example, the cotton plant *Gossypium bickii* has different ancestors for cytoplasmic and nuclear genomes following a hybridisation event but there is little evidence for subsequent nuclear recombination; this could be taken as evidence that the current nuclear-cytoplasm combination is under significant positive selection (Wendel *et al.* 1991). In extreme cases, this can manifest itself as cytoplasmic male sterility (CMS): cytoplasmic genes from one parent are turned off, until they come into contact with certain nuclear genes from the other parent, when they are then turned on again (Frank 1989; McCauley 2013). This suggests that there can be significant evolutionary benefit to certain nuclear-cytoplasm combinations. Given the evidence presented for nuclear-organellar recombination in *Symbiodinium*, this could be a rich potential source of adaptive traits. Importantly, other studies of hybridisation (*e.g.* Rieseberg *et al.* 2003) suggest that these adaptive hybrids can be generated under normal conditions, before expanding to open marginal niches if given the opportunity. While fidelity for particular symbiont taxa is generally displayed by coral hosts (Chapter 2), they have also demonstrated the ability to take up new symbionts in new environments (Baird *et al.* 2007), and so they could associate with adaptive hybrid types in the future.

4.4. Considerations arising from this thesis

This study suggests that hybridisation between divergent *Symbiodinium* lineages can occur; however, it may require high levels of symbiont diversity to do so. It has been suggested that "disaster taxa" such as type D1a (*S. trenchii*) may be able to see corals through stressful periods, and that the future of coral reefs lies with one or a few resilient generalists (*e.g.* Correa and Baker 2011; Silverstein *et al.* 2015). However, this study suggests the opposite: high symbiont diversity is required to maintain adaptive potential.

A good example of this is in the Persian/Arabian Gulf, where *S. thermophilum* was described by Hume and colleagues (2015). This thermally resistant symbiont has allowed corals to survive in the world's hottest sea, a body of water which became hot over reasonably short evolutionary timescales (Hume *et al.* 2016). While this symbiont is now common in this body of water, modelling suggests that it arose from a rare variant that suddenly diversified and spread when conditions changed (Hume *et al.* 2016). This supports the notion that a rare beneficial hybrid could spread and provide resistance over reasonably short timescales. The authors go on to explicitly state that protection of symbiont diversity is required, from which stress-tolerant genotypes can become more common under severe natural selection. This is supported by the hypothesis of Ziegler and colleagues (2017b), who observe that the dominant members of symbiont communities in corals are insufficient to explain the fine-scale environmental responses displayed by reef communities. They instead suggest that background symbiont populations have an important role to play, though they do not explicitly state how. I suggest that hybridisation may be a contributor to this role; the evidence in this study, as well as other possible cases documented in the introduction to Chapter 3, indicate that rare or background types can produce novel *Symbiodinium* types through sexual recombination with more common types.

This naturally leads to the question as to where symbiont hybridisation may occur. When in symbiosis, *Symbiodinium* cells are sequestered in vacuoles inside host cells (Davy *et al.* 2012), with little chance for contact with other symbionts. The exception is in gametes; coral oocytes of vertically-transmitting hosts have been demonstrated to contain diverse symbiont assemblages (Padilla-Gamiño *et al.* 2012). Therefore, it may be hypothesised that vertically-transmitting corals would show greater evidence of hybridisation, something which has been observed for the fungal symbiont *Neotyphodium* and its grass hosts (Hamilton *et al.* 2010). This is supported by the genetic study of Pettay *et al.* (2011), which finds evidence for sexual recombination within *S. glynni*, a symbiont vertically transmitted by *Pocillopora*. However, only one of the four hosts of putative hybrids in the current study are vertical transmitters. The others take up their symbionts from the environment, which has been shown to host a wide range of *Symbiodinium* taxa, both in the water column (Coffroth *et al.* 2006) and in benthic sediments (Nitschke *et al.* 2016). Therefore, it is possible that diverse symbionts could sexually reproduce *ex*

hospite and be taken up by other corals. Additional results from my study suggest that there is broad sharing of symbiont types among host genera. For example, *Symbiodinium* C40n (psbA^{ncr}) was only found at sites LIW and LIE (on ten occasions), but in eight different horizontally transmitting genera of corals. Therefore, there is good evidence that symbionts could hybridise and spread in both vertically and horizontally transmitting coral hosts. However, determining the location of symbiont sex would make studies of hybridisation more compelling.

The performance of putative hybrids is another crucial piece of information for determining the validity of hybridisation as a mechanism of long-term coral survival. This cannot be assessed here, as the ITS2 and psbA^{ncr} regions are non-coding, the *cob* gene is highly conserved and the samples collected were too small to provide any convincing parameters of performance while in symbiosis. Ideally, effort will be put into quantifying the physiological and ecological performance of putative hybrids in combination with genetic data, as this is the most convincing way to distinguish functionally-distinct lineages (Arnold and Martin 2010; Charlton *et al.* 2014). Some work is already taking place in this area, both for *Symbiodinium* species generally (LaJeunesse *et al.* 2012, 2014) and for potential hybrids (Wilkinson *et al.* 2016). However, much more is required, and a combination of phylogenetic and ecological methods should be a feature of future *Symbiodinium* studies.

Finally, this study demonstrates the need to understand *Symbiodinium* diversity more thoroughly. The weight of evidence for sexual recombination is strong, yet precise knowledge of where and how *Symbiodinium* has sex is still lacking. Until this is resolved, hypotheses like the one advocated in this thesis remain equivocal. This study has also shown new host-symbiont relationships and added to the already prodigious list of *Symbiodinium* types. Given that the ecological success of corals is owed largely to their symbiosis with their dinoflagellate symbionts, a complete understanding of the scale and nature of these relationships must be a priority.

4.5. Concluding remarks

The reefs of northern Timor and associated islands match the general patterns of the Indo-Pacific, while displaying high diversity and several novel relationships. These results are interpreted with confidence thanks to the multi-gene approach taken here, as well as the use of Next Generation Sequencing. The use of these methods is strongly advocated for future studies, particularly NGS which has been demonstrated to be highly cost-effective while still producing high quality data when used in the manner described here. Atauro Island appears to have a greater diversity of *Symbiodinium* than Timor, a result which extends to strong evidence of hybridisation at select Atauro Island sites only. This suggests that, while hybridisation is a potential adaptive mechanism, it may be limited to high-quality sites which are able to maintain high levels of symbiont diversity. Importantly, the corals of Timor are not different to

those at Atauro Island (Turak and DeVantier 2013), which suggests that even regions with high coral diversity may be losing symbiont diversity. As such, symbiont diversity should be studied and managed as much as coral diversity to maintain maximum adaptive potential, something which is, regretfully, increasingly necessary.

References

Aihara Y, Takahashi S, Minagawa J (2016) Heat induction of cyclic electron flow around photosystem I in the symbiotic dinoflagellate Symbiodinium. *Plant Physiology*, **171**, 522-529.

Ainsworth TD, Heron SF, Ortiz JC, Mumby PJ, Grech A, Ogawa D, ..., Leggat W (2016) Climate change disables coral bleaching protection on the Great Barrier Reef. *Science*, **352**(6283), 338-342.

Allen GR (2008) Conservation hotspots of biodiversity and endemism for Indo-Pacific coral reef fishes. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **18**, 541-556.

Anthony KR, Kline DI, Diaz-Pulido G, Dove S, Hoegh-Guldberg O (2008) Ocean acidification causes bleaching and productivity loss in coral reef builders. *Proceedings of the National Academy of Sciences*, **105**, 17442-17446.

Arif C, Daniels C, Bayer T, Banguera-Hinestroza E, Barbrook A, Howe CJ, LaJeunesse TC, Voolstra CR (2014) Assessing Symbiodinium diversity in scleractinian corals via next-generation sequencing-based genotyping of the ITS2 rDNA region. *Molecular Ecology*, **23**, 4418-4433.

Arnold ML, Martin NH (2010) Hybrid fitness across time and habitats. *Trends in Ecology and Evolution*, **25**, 530-536.

Atmadipoera A, Molcard R, Madec G, Wijffels S, Sprintall J, Koch-Larrouy, A, ..., Supangat A (2009) Characteristics and variability of the Indonesian throughflow water at the outflow straits. *Deep Sea Research Part I: Oceanographic Research Papers*, **56**, 1942-1954.

Ayling AM, Ayling AL, Edyvane KS, Penny S, de Carvalho N, Fernandes A, Amaral AL (2009) Preliminary biological resource survey of fringing reefs in the proposed Nino Konis Santana Marine Park, Timor Leste. Report to the Northern Territory Department of Natural Resources, Environment & the Arts.

Baillie BK, Monje V, Silvestre V, Sison M, Belda-Baillie CA (1998) Allozyme electrophoresis as a tool for distinguishing different zooxanthellae symbiotic with giant clams. *Proceedings of the Royal Society of London B: Biological Sciences*, **265**, 1949-1956.

Baillie BK, Belda-Baillie CA, Maruyama, T (2000a) Conspecificity and Indo-Pacific distribution of Symbiodinium genotypes (Dinophyceae) from giant clams. *Journal of Phycology*, **36**, 1153-1161.

Baillie BK, Belda-Baillie CA, Silvestre V, Sison M, Gomez AV, Gomez ED, Monje V (2000b) Genetic variation in Symbiodinium isolates from giant clams based on random-amplified-polymorphic DNA (RAPD) patterns. *Marine Biology*, **136**, 829-836.

Baird AH, Cumbo VR, Leggat W, Rodriguez-Lanetty M (2007) Fidelity and flexibility in coral-algal symbioses. *Marine Ecology Progress Series*, **347**, 307-309.

Baird AH, Guest JR, Willis BL (2009) Systematic and biogeographical patterns in the reproductive biology of scleractinian corals. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 551-571.

Baker AC (2001) Reef corals bleach to survive change. *Nature*, **411**, 765-766.

Baker AC (2002) Ecology (Communication arising): Is coral bleaching really adaptive?. *Nature*, **415**, 602.

Baker AC (2003) Flexibility and specificity in coral-algal symbiosis: diversity, ecology, and biogeography of Symbiodinium. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 661–689.

Baker AC, Starger CJ, McClanahan TR, Glynn PW (2004) Coral reefs: corals' adaptive response to climate change. *Nature*, **430**, 741.

Baker AC, Correa AM, Cunning R (2017) Diversity, distribution and stability of Symbiodinium in reef corals of the eastern tropical pacific. In: *Coral Reefs of the Eastern Tropical Pacific* (eds Glynn PW, Manzello DP, Enochs IC), pp. 405-420. Springer, Netherlands.

Baker DM, Andras JP, Jordán-Garza AG, Fogel ML (2013) Nitrate competition in a coral symbiosis varies with temperature among Symbiodinium clades. *The ISME journal*, **7**, 1248-1251.

Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ (1995) The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden*, 247-277.

Banaszak AT, LaJeunesse TC, Trench RK (2000) The synthesis of mycosporine-like amino acids (MAAs) by cultured, symbiotic dinoflagellates. *Journal of Experimental Marine Biology and Ecology*, **249**, 219-233.

Barbrook AC, Visram S, Douglas AE, Howe CJ (2006) Molecular diversity of dinoflagellate symbionts of Cnidaria: the psbA minicircle of Symbiodinium. *Protist*, **157**, 159-171.

Barbrook AC, Voolstra CR, Howe CJ (2014) The chloroplast genome of a Symbiodinium sp. clade C3 isolate. *Protist*, **165**, 1-13.

Barker FK, Lutzoni FM (2002) The utility of the incongruence length difference test. *Systematic Biology*, **51**, 625-637.

Barshis DJ, Ladner JT, Oliver TA, Palumbi SR (2014) Lineage-specific transcriptional profiles of Symbiodinium spp. unaltered by heat stress in a coral host. *Molecular Biology and Evolution*, **31**, 1343-1352.

Baskett ML, Gaines SD, Nisbet RM (2009) Symbiont diversity may help coral reefs survive moderate climate change. *Ecological Applications*, **19**, 3-17.

Baums IB, Devlin-Durante MK, LaJeunesse TC (2014) New insights into the dynamics between reef corals and their associated dinoflagellate endosymbionts from population genetic studies. *Molecular Ecology*, **23**, 4203-4215.

Bay LK, Doyle J, Logan M, Berkelmans R (2016) Recovery from bleaching is mediated by threshold densities of background thermo-tolerant symbiont types in a reef-building coral. *Royal Society Open Science*, **3**, 160322.

Benayas JMR, Newton AC, Diaz A, Bullock JM (2009) Enhancement of biodiversity and ecosystem services by ecological restoration: a meta-analysis. *Science*, **325**, 1121-1124.

Berkelmans R, van Oppen MJH (2006) Flexible partners in coral symbiosis: a 'nugget of hope' for coral reefs in an era of climate change. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 2305-2312.

Bieri T, Onishi M, Xiang T, Grossman AR, Pringle JR (2016) Relative contributions of various cellular mechanisms to loss of algae during cnidarian bleaching. *PloS One*, **11**, e0152693.

Biquand E, Okubo N, Aihara Y, Rolland V, Hayward DC, Hatta M, ..., Takahashi S (2017) Acceptable symbiont cell size differs among cnidarian species and may limit symbiont diversity. *The ISME Journal*, **11**, 1702-1712.

Blank RJ, Trench RK (1985) Speciation and symbiotic dinoflagellates. *Science*, **229**, 656-659.

Bock DG, Andrew RL, Rieseberg LH (2014) On the adaptive value of cytoplasmic genomes in plants. *Molecular Ecology*, **23**, 4899-4911.

Bongaerts P, Carmichael M, Hay KB, Tonk L, Frade PR, Hoegh-Guldberg O (2015a) Prevalent endosymbiont zonation shapes the depth distributions of scleractinian coral species. *Royal Society Open Science*, **2**, 140297.

Bongaerts P, Frade PR, Hay KB, Englebert N, Latijnhouwers KR, Bak RP, ..., Hoegh-Guldberg O (2015b) Deep down on a Caribbean reef: lower mesophotic depths harbor a specialized coral-endosymbiont community. *Scientific reports*, DOI: 10.1038/srep07652.

Bouchard JN, Yamasaki H (2009) Implication of nitric oxide in the heat-stress-induced cell death of the symbiotic alga Symbiodinium microadriaticum. *Marine Biology*, **156**, 2209-2220.

Boulotte NM, Dalton SJ, Carroll AG, Harrison PL, Putnam HM, Peplow LM, van Oppen MJ (2016) Exploring the Symbiodinium rare biosphere provides evidence for symbiont switching in reef-building corals. *The ISME journal*, **10**, 2693-2701.

Brandt K (1881) Über das Zusammenleben von Thieren undAlgen. *Verhandlungen der Physiologischen Gesellschaft zu Berlin,* 1881/1882, 22-26.

Brawley SH, Johnson LE (1992) Gametogenesis, gametes and zygotes: an ecological perspective on sexual reproduction in the algae. *British Phycological Journal*, **27**, 233-252.

Brosnahan, ML (2011) Life cycle studies of the red tide dinoflagellate species complex Alexandrium tamarense. Doctoral dissertation, Massachusetts Institute of Technology.

Brosnahan ML, Kulis DM, Solow AR, Erdner DL, Percy L, Lewis J, Anderson DM (2010) Outbreeding lethality between toxic Group I and nontoxic Group III Alexandrium tamarense spp. isolates: predominance of heterotypic encystment and implications for mating interactions and biogeography. *Deep Sea Research Part II: Topical Studies in Oceanography*, **57**, 175-189.

Browne NK, Precht E, Last KS, Todd PA (2014) Photo-physiological costs associated with acute sediment stress events in three near-shore turbid water corals. *Marine Ecology Progress Series*, **502**, 129-143.

Bruno JF, Selig ER, Casey KS, Page CA, Willis BL, Harvell CD, ..., Melendy AM (2007) Thermal stress and coral cover as drivers of coral disease outbreaks. *PLoS biology*, **5**, e124.

Buddemeier RW, Fautin DG (1993) Coral bleaching as an adaptive mechanism. *Bioscience*, **43**, 320-326.

Bull JJ, Huelsenbeck JP, Cunningham CW, Swofford DL, Waddell PJ (1993) Partitioning and combining data in phylogenetic analysis. *Systematic Biology*, **42**, 384-397.

Burriesci MS, Raab TK, Pringle JR (2012) Evidence that glucose is the major transferred metabolite in dinoflagellate–cnidarian symbiosis. *Journal of Experimental Biology*, **215**, 3467-3477.

Byler KA, Carmi-Veal M, Fine M, Goulet TL (2013) Multiple symbiont acquisition strategies as an adaptive mechanism in the coral Stylophora pistillata. *PloS One*, **8**, e59596.

Cantin NE, van Oppen MJH, Willis BL, Mieog JC, Negri AP (2009) Juvenile corals can acquire more carbon from high-performance algal symbionts. *Coral Reefs*, **28**, 405-414.

Carlos AA, Baillie BK, Kawachi M, Maruyama T (1999) Phylogenetic position of Symbiodinium (Dinophyceae) isolates from tridacnids (Bivalvia), cardiids (Bivalvia), a sponge (Porifera), a soft coral (Anthozoa), and a free-living strain. *Journal of Phycology*, **35**, 1054-1062.

Carpenter KE, Abrar M, Aeby G, Aronson RB, Banks S, Bruckner A, ..., Edgar GJ (2008) One-third of reef-building corals face elevated extinction risk from climate change and local impacts. *Science*, **321**, 560-563.

Carilli J, Donner SD, Hartmann AC (2012) Historical temperature variability affects coral response to heat stress. *PLoS One*, **7**, e34418.

Charlton ND, Craven KD, Afkhami ME, Hall BA, Ghimire SR, Young CA (2014) Interspecific hybridization and bioactive alkaloid variation increases diversity in endophytic Epichloë species of Bromus laevipes. *FEMS Microbiology Ecology*, **90**, 276-289.

Chauka LJ (2012) Diversity of symbiotic algae in the genus Symbiodinium in Scleractinian corals of Tanzania. *Western Indian Ocean Journal of Marine Science*, **11**, 67-76.

Chauka LJ (2016) Tanzanian Reef Building Corals May Succumb to Bleaching Events: Evidences from Coral-Symbiodinium Symbioses. In: *Estuaries: A Lifeline of Ecosystem Services in the Western Indian Ocean* (eds Salif D, Scheren P, Ferdinand Machiwa J), pp. 161-168. Springer International, Switzerland.

Charlton ND, Craven KD, Afkhami ME, Hall BA, Ghimire SR, Young CA (2014) Interspecific hybridization and bioactive alkaloid variation increases diversity in endophytic Epichloë species of Bromus laevipes. *FEMS Microbiology Ecology*, **90**, 276-289.

Cheung WW, Lam VW, Sarmiento JL, Kearney K, Watson REG, Zeller D, Pauly D (2010) Large-scale redistribution of maximum fisheries catch potential in the global ocean under climate change. *Global Change Biology*, **16**, 24-35.

Chesnick JM, Cox ER (1987) Synchronized sexuality of an algal symbiont and its dinoflagellate host, Peridinium balticum (Levander) Lemmermann. *Biosystem*s, **21**, 69-78.

Chi J, Parrow MW, Dunthorn M (2014) Cryptic sex in Symbiodinium (Alveolata, Dinoflagellata) is supported by an inventory of meiotic genes. *Journal of Eukaryotic Microbiology*, **61**, 322-327.

Chou Q, Russell M, Birch DE, Raymond J, Bloch W (1992) Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. *Nucleic Acids Research*, **20**, 1717-1723.

Cinner JE (2014) Coral reef livelihoods. *Current Opinion in Environmental Sustainability*, **7**, 65-71.

Cinner JE, McClanahan TR, Graham NAJ, Daw TM, Maina J, Stead SM, ..., Bodin Ö (2012) Vulnerability of coastal communities to key impacts of climate change on coral reef fisheries. *Global Environmental Change*, **22**, 12-20.

Coffroth MA, Santos SR, Goulet TL (2001) Early ontogenetic expression of specificity in a cnidarian-algal symbiosis. *Marine Ecology Progress Series*, **222**, 85–96.

Coffroth MA, Santos SR (2005) Genetic diversity of symbiotic dinoflagellates in the genus Symbiodinium. *Protist*, **156**, 19-34.

Coffroth MA, Lewis CF, Santos SR, Weaver JL (2006) Environmental populations of symbiotic dinoflagellates in the genus Symbiodinium can initiate symbioses with reef cnidarians. *Current Biology*, **16**, R985-R987.

Coffroth MA, Poland DM, Petrou EL, Brazeau DA, Holmberg JC (2010) Environmental symbiont acquisition may not be the solution to warming seas for reef-building corals. *PLoS One*, **5**, e13258.

Coker DJ, Wilson SK, Pratchett MS (2014) Importance of live coral habitat for reef fishes. *Reviews in Fish Biology and Fisheries*, **24**, 89-126.

Combosch DJ, Vollmer SV (2015) Trans-Pacific RAD-Seq population genomics confirms introgressive hybridization in Eastern Pacific Pocillopora corals. *Molecular Phylogenetics and Evolution*, **88**, 154-162.

Conservation International (2016) conservation.org/NewsRoom/pressreleases/Pages/Biodiversity-survey-reveals-reefs-in-Timor-Leste-s-Atauro-Island-hold-the-worlds-highest-reef-fish-species-average-.aspx (18/8/16)

Correa AMS, Baker AC (2009) Understanding diversity in coral-algal symbiosis: a cluster-based approach to interpreting fine-scale genetic variation in the genus Symbiodinium. *Coral Reefs*, **28**, 81-93.

Correa AMS, McDonald MD, Baker AC (2009) Development of clade-specific Symbiodinium primers for quantitative PCR (qPCR) and their application to detecting clade D symbionts in Caribbean corals. *Marine Biology*, **156**, 2403-2411.

Correa AMS, Baker AC (2011) Disaster taxa in microbially mediated metazoans: How endosymbionts and environmental catastrophes influence the adaptive capacity of reef corals. *Global Change Biology*, **17**, 68-75.

Costanza R, d'Arge R, De Groot R, Farber S, Grasso M, Hannon B, ..., Raskin RG (1998) The value of the world's ecosystem services and natural capital. *Ecological Economics*, **25**, 3-16.

Császár NBM, Ralph PJ, Frankham R, Berkelmans R, van Oppen MJH (2010) Estimating the potential for adaptation of corals to climate warming. *PLoS One*, **5**, e9751.

Cumbo VR, Baird AH, van Oppen MJH (2013) The promiscuous larvae: flexibility in the establishment of symbiosis in corals. *Coral Reefs*, **32**, 111-120.

Cunning R, Glynn PW, Baker AC (2013) Flexible associations between Pocillopora corals and Symbiodinium limit utility of symbiosis ecology in defining species. *Coral Reefs*, **32**, 795-801.

Cunning R, Gillette P, Capo T, Galvez K, Baker AC (2015a) Growth tradeoffs associated with thermotolerant symbionts in the coral Pocillopora damicornis are lost in warmer oceans. *Coral Reefs*, **34**, 155-160.

Cunning R, Yost DM, Guarinello ML, Putnam HM, Gates RD (2015b) Variability of Symbiodinium communities in waters, sediments, and corals of thermally distinct reef pools in American Samoa. *PloS One*, **10**, e0145099.

D'Angelo C, Hume BC, Burt J, Smith EG, Achterberg EP, Wiedenmann J (2015) Local adaptation constrains the distribution potential of heat-tolerant Symbiodinium from the Persian/Arabian Gulf. *The ISME journal*, **9**, 2551-2560.

Davies SW, Wham DC, Kanke MR, Matz MV (2016) Ecological factors rather than physical barriers shape genetic structure of algal symbionts in Micronesian corals. *bioRxiv*, 037994.

Davy SK, Allemand D, Weis VM (2012) Cell biology of cnidarian-dinoflagellate symbiosis. *Microbiology and Molecular Biology Reviews*, **76**, 229-261.

De Bary A (1879) *Die Erscheinung der Symbiose* (ed Trubner KJ). Strasbourg, Germany.

Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference, and the multispecies coalescent. *Trends in Ecology and Evolution*, **24**, 332-340.

Derks FHM, Hakkert JC, Verbeek WHJ, Colijn-Hooymans CM (1992) Genome composition of asymmetric hybrids in relation to the phylogenetic distance between the parents. Nucleus-chloroplast interaction. *Theoretical and Applied Genetics*, **84**, 930-940.

Díaz S, Demissew S, Joly C, Lonsdale WM, Larigauderie A (2015) A Rosetta Stone for nature's benefits to people. *PLoS Biology*, **13**, e1002040.

Dittrich-Reed DR, Fitzpatrick BM (2013) Transgressive hybrids as hopeful monsters. *Evolutionary Biology*, **40**, 310-315.

Dolphin K, Belshaw R, Orme CDL, Quicke DL (2000) Noise and incongruence: interpreting results of the incongruence length difference test. *Molecular phylogenetics and evolution*, **17**, 401-406.

Douglas AE (1998) Host benefit and the evolution of specialization in symbiosis. *Heredity*, **81**, 599-603.

Douglas AE (2008) Conflict, cheats and the persistence of symbioses. *New Phytologist*, **177**, 849-858.

Dubinsky Z, Stambler N (1996) Marine pollution and coral reefs. *Global Change Biology*, **2**, 511-526.

Edinger EN, Jompa J, Limmon GV, Widjatmoko W, Risk MJ (1998) Reef degradation and coral biodiversity in Indonesia: effects of land-based pollution, destructive fishing practices and changes over time. *Marine Pollution Bulletin*, **36**, 617-630.

Edmunds PJ, Comeau S, Lantz C, Andersson A, Briggs C, Cohen A, ..., Muller EB (2016) Integrating the effects of ocean acidification across functional scales on tropical coral reefs. *BioScience*, **66**, 350-362.

Edvardsen B, Shalchian-Tabrizi K, Jakobsen KS, Medlin LK, Dahl E, Brubak S, Paasche E (2003) Genetic variability and molecular phylogeny of Dinophysis species (Dinophyceae) from Norwegian waters inferred from single cell analyses of rDNA. *Journal of Phycology*, **39**, 395-408.

Ellstrand NC, Hoffman CA (1990) Hybridization as an avenue of escape for engineered genes. *Bioscience*, **40**, 438-442.

Fabina NS, Putnam HM, Franklin EC, Stat M, Gates RD (2012) Transmission mode predicts specificity and interaction patterns in coral-Symbiodinium networks. *PLoS One*, **7**, e44970.

Fabricius KE (2005) Effects of terrestrial runoff on the ecology of corals and coral reefs: review and synthesis. *Marine Pollution Bulletin*, **50**, 125-146.

Fay SA, Weber MX, Lipps JH (2009) The distribution of Symbiodinium diversity within individual host foraminifera. *Coral Reefs*, **28**, 717-726.

Faircloth BC, Glenn TC (2014) Protocol: Preparation of an AMPure XP substitute (AKA Serapure). DOI: 10.6079/J9MW2F26

Falkowski PG, Dubinsky Z, Muscatine L, Porter JW (1984) Light and the bioenergetics of a symbiotic coral. *Bioscience*, **34**, 705-709.

Farris JS, Källersjö M, Kluge AG, Bult C (1994) Testing significance of incongruence. *Cladistics*, **10**, 315-319.

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution*, **17**, 368-376.

Finney JC, Pettay DT, Sampayo EM, Warner ME, Oxenford HA, LaJeunesse TC (2010) The relative significance of host-habitat, depth, and geography on the ecology, endemism, and speciation of coral endosymbionts in the genus Symbiodinium. *Microbial Ecology*, **60**, 250–263.

Fischer J, Lindenmayer DB, Manning AD (2006) Biodiversity, ecosystem function, and resilience: ten guiding principles for commodity production landscapes. *Frontiers in Ecology and the Environment*, **4**, 80-86.

Fitt WK, Chang SS, Trench RK (1981) Motility patterns of different strains of the symbiotic dinoflagellate Symbiodinium (= Gymnodinium) microadriaticum (Freudenthal) in culture. *Bulletin of Marine Science*, **31**, 436-443.

Fitt WK, Gates RD, Hoegh-Guldberg O, Bythell JC, Jatkar A, Grottoli AG, ..., Iglesias-Prieto R (2009) Response of two species of Indo-Pacific corals, Porites cylindrica and Stylophora pistillata, to short-term thermal stress: the host does matter in determining the tolerance of corals to bleaching. *Journal of Experimental Marine Biology and Ecology*, **373**, 102-110.

Folk RA, Mandel JR, Freudenstein JV (2017) Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms. *Systematic Biology*, **66**, 320-337.

Forsman ZH, Hunter CL, Fox GE, Wellington GM (2006) Is the ITS region the solution to the "species problem" in corals? Intragenomic variation and alignment permutation in Porites, Siderastrea and outgroup taxa. In *Proceedings of the 10th Int. Coral Reef Symposium*, **1**, 14-23.

Frank SA (1989) The evolutionary dynamics of cytoplasmic male sterility. *The American Naturalist*, **133**, 345-376.

Franklin EC, Stat M, Pochon X, Putnam HM, Gates RD (2012) GeoSymbio: a hybrid, cloud-based web application of global geospatial bioinformatics and ecoinformatics for Symbiodinium–host symbioses. *Molecular Ecology Resources*, **12**, 369-373.

Fredricks DN, Smith C, Meier A (2005) Comparison of six DNA extraction methods for recovery of fungal DNA as assessed by quantitative PCR. *Journal of Clinical Microbiology*, **43**, 5122-5128.

Freudenthal HD (1962) Symbiodinium Gen Nov and Symbiodinium microadriaticum Sp Nov, a zooxanthella—taxonomy, life cycle, and morphology. *Journal of Protozoology*, **9**, 45-52.

Geneious version 8.0.5 created by Biomatters. Available from http://www.geneious.com

Gibbin EM, Davy SK (2013) Intracellular pH of symbiotic dinoflagellates. *Coral Reefs*, **32**, 859-863.

Glynn PW (1983) Extensive 'bleaching' and death of reef corals on the Pacific coast of Panamá. *Environmental Conservation*, **10**, 149-154.

Glynn PW (1996) Coral reef bleaching: facts, hypotheses and implications. *Global Change Biology*, **2**, 495-509.

Glynn PW, Maté JL, Baker AC, Calderón MO (2001) Coral bleaching and mortality in Panama and Ecuador during the 1997–1998 El Niño–Southern Oscillation event: spatial/temporal patterns and comparisons with the 1982–1983 event. *Bulletin of Marine Science*, **69**, 79-109.

Goldman N, Anderson JP, Rodrigo AG (2000) Likelihood-based tests of topologies in phylogenetics. *Systematic Biology*, **49**, 652-670.

Gómez F (2012) A checklist and classification of living dinoflagellates (Dinoflagellata, Alveolata). *Cicimar Oceanides*, **27**, 65-140.

Gordon AL, Susanto RD (2001) Banda Sea surface-layer divergence. *Ocean Dynamics*, **52**, 2-10.

Govindarajulu R, Parks M, Tennessen JA, Liston A, Ashman TL (2015) Comparison of nuclear, plastid, and mitochondrial phylogenies and the origin of wild octoploid strawberry species. *American Journal of Botany*, **102**, 544-554.

Graham NAJ, Nash KL (2013) The importance of structural complexity in coral reef ecosystems. *Coral Reefs*, **32**, 315-326.

Grégoire V, Schmacka F, Coffroth MA, Karsten U (2017) Photophysiological and thermal tolerance of various genotypes of the coral endosymbiont Symbiodinium sp.(Dinophyceae). *Journal of Applied Phycology*, DOI: 10.1007/s10811-017-1127-1.

Gribble KE, Anderson DM (2007) High intraindividual, intraspecific, and interspecific variability in large-subunit ribosomal DNA in the heterotrophic dinoflagellates Protoperidinium, Diplopsalis, and Preperidinium (Dinophyceae). *Phycologia*, **46**, 315-324.

Grupstra CG, Coma R, Ribes M, Leydet KP, Parkinson JE, McDonald K, ..., Coffroth MA (2017) Evidence for coral range expansion accompanied by reduced diversity of Symbiodinium genotypes. *Coral Reefs*, **36**, 981-985.

Guest JR, Baird AH, Maynard JA, Muttaqin E, Edwards AJ, Campbell SJ, ..., Chou LM (2012) Contrasting patterns of coral bleaching susceptibility in 2010 suggest an adaptive response to thermal stress. *PLoS One*, **7**, e33353.

Hamilton CE, Dowling TE, Faeth SH (2010) Hybridization in endophyte symbionts alters host response to moisture and nutrient treatments. *Microbial Ecology*, **59**, 768-775.

Harii S, Kayanne H, Takigawa H, Hayashibara T, Yamamoto M (2002) Larval survivorship, competency periods and settlement of two brooding corals, Heliopora coerulea and Pocillopora damicornis. Marine *Biology*, **141**, 39-46.

Hansen G, Daugbjerg N (2009) Symbiodinium natans sp. nov.: A "free-living" dinoflagellate from Tenerife (Northeast-Atlantic Ocean). *Journal of Phycology*, **45**, 251-263.

Hart MC, Green DH, Bresnan E, Bolch CJ (2007) Large subunit ribosomal RNA gene variation and sequence heterogeneity of Dinophysis (Dinophyceae) species from Scottish coastal waters. *Harmful Algae*, **6**, 271-287.

Hawkins TD, Bradley BJ, Davy SK (2013) Nitric oxide mediates coral bleaching through an apoptotic-like cell death pathway: evidence from a model sea anemone-dinoflagellate symbiosis. *The FASEB Journal*, **27**, 4790-4798.

Hawkins TD, Krueger T, Wilkinson SP, Fisher PL, Davy SK (2015) Antioxidant responses to heat and light stress differ with habitat in a common reef coral. *Coral Reefs*, **34**, 1229-1241.

Heitman J (2010) Evolution of eukaryotic microbial pathogens via covert sexual reproduction. *Cell Host & Microbe*, **8**, 86-99.

Hennige SJ, Suggett DJ, Warner ME, McDougall KE, Smith DJ (2009) Photobiology of Symbiodinium revisited: bio-physical and bio-optical signatures. *Coral Reefs*, **28**, 179-195.

Hentschel U, Steinert M, Hacker J (2000) Common molecular mechanisms of symbiosis and pathogenesis. *Trends in Microbiology*, **8**, 226-231.

Hicks C, McClanahan T, Cinner J, Hills J (2009) Trade-offs in values assigned to ecological goods and services associated with different coral reef management strategies. *Ecology and Society*, **14**.

Higuchi T, Yuyama I, Nakamura T (2015) The combined effects of nitrate with high temperature and high light intensity on coral bleaching and antioxidant enzyme activities. *Regional Studies in Marine Science*, **2**, 27-31.

Hill R, Fernance C, Wilkinson SP, Davy SK, Scott A (2014) Symbiont shuffling during thermal bleaching and recovery in the sea anemone Entacmaea quadricolor. *Marine Biology*, **161**, 2931-2937.

Hillyer KE, Tumanov S, Villas-Bôas S, Davy SK (2016) Metabolite profiling of symbiont and host during thermal stress and bleaching in a model cnidarian–dinoflagellate symbiosis. *Journal of Experimental Biology*, **219**, 516-527.

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

Hoegh-Guldberg O, Smith GJ (1989) The effect of sudden changes in temperature, light and salinity on the population density and export of zooxanthellae from the reef corals Stylophora pistillata Esper and Seriatopora hystrix Dana. *Journal of Experimental Marine Biology and Ecology*, **129**, 279-303.

Hoegh-Guldberg O, Salvat B (1995) Periodic mass-bleaching and elevated sea temperatures: bleaching of outer reef slope communities in Moorea, French Polynesia. *Marine Ecology Progress Series*, **121**, 181-190.

Hoegh-Guldberg O, Jones RJ, Ward S, Loh WK (2002) Ecology (Communication arising): Is coral bleaching really adaptive?. *Nature*, **415**, 601.

Hoegh-Guldberg O, Mumby PJ, Hooten AJ, Steneck RS, Greenfield P, Gomez E, ..., Knowlton N (2007) Coral reefs under rapid climate change and ocean acidification. *Science*, **318**, 1737-1742.

Hovasse R (1923) Endodinium chattoni (nov. gen, nov. sp.) parasite des velelles. Un type exceptionnel de variation du nombre des chromosomes. *Bulletin Biologique de la France et de la Belgique*, **57**, 107-130.

Howells EJ, van Oppen MJH, Willis BL (2009) High genetic differentiation and cross-shelf patterns of genetic diversity among Great Barrier Reef populations of Symbiodinium. *Coral Reefs*, **28**, 215-225.

Howells EJ, Beltran VH, Larsen NW, Bay LK, Willis BL, van Oppen MJH (2012) Coral thermal tolerance shaped by local adaptation of photosymbionts. *Nature Climate Change*, **2**, 116-120.

Howells EJ, Willis BL, Bay LK, Oppen MJ (2013) Spatial and temporal genetic structure of Symbiodinium populations within a common reef-building coral on the Great Barrier Reef. *Molecular Ecology*, **22**, 3693-3708.

Howells EJ, Willis BL, Bay LK, van Oppen MJH (2016a) Microsatellite allele sizes alone are insufficient to delineate species boundaries in Symbiodinium. *Molecular Ecology*, **25**, 2719-2723.

Howells EJ, Abrego D, Meyer E, Kirk N, Burt J (2016b) Host adaptation and unexpected symbiont partners enable reef-building corals to tolerate extreme temperatures. *Global Change Biology*, **22**, 2702-2714.

Hughes TP, Baird AH, Bellwood DR, Card M, Connolly SR, Folke C, …, Lough JM (2003) Climate change, human impacts, and the resilience of coral reefs. *Science*, **301**, 929-933.

Hughes TP, Kerry JT, Álvarez-Noriega M, Álvarez-Romero JG, Anderson KD, Baird AH, ..., Bridge TC (2017) Global warming and recurrent mass bleaching of corals. *Nature*, **543**, 373.

Hume BC, D'Angelo C, Smith EG, Stevens JR, Burt J, Wiedenmann J (2015) Symbiodinium thermophilum sp. nov., a thermotolerant symbiotic alga prevalent in corals of the world's hottest sea, the Persian/Arabian Gulf. *Scientific reports*, **5**, 8562

Hume BC, Voolstra CR, Arif C, D'Angelo C, Burt JA, Eyal G, ..., Wiedenmann J (2016) Ancestral genetic diversity associated with the rapid spread of stress-tolerant coral symbionts in response to Holocene climate change. *Proceedings of the National Academy of Sciences*, **113**, 4416-4421.

Huson D, Linz S (2016) Autumn Algorithm–Computation of Hybridization Networks for Realistic Phylogenetic Trees. *IEEE/ACM transactions on computational biology and bioinformatics.*

Iglesias-Prieto R, Matta JL, Robins WA, Trench RK (1992) Photosynthetic response to elevated temperature in the symbiotic dinoflagellate Symbiodinium microadriaticum in culture. *Proceedings of the National Academy of Sciences*, **89**, 10302-10305.

Iglesias-Prieto R, Beltran VH, LaJeunesse TC, Reyes-Bonilla H, Thome PE (2004) Different algal symbionts explain the vertical distribution of dominant reef corals in the eastern Pacific. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **271**, 1757-1763.

illumina.com/systems/sequencing-platforms/miseq/specifications.html (accessed 20/2/18)

Jeong HJ, Lee SY, Kang NS, Yoo YD, Lim AS, Lee MJ, Kim HS, Yih WH, LaJeunesse TC (2014) Genetics and morphology characterize the dinoflagellate Symbiodinium voratum, n. sp., (Dinophyceae) as the sole representative of Symbiodinium clade E. *Journal of Eukaryotic Microbiology*, **61**, 75-94.

Joly S, McLenachan PA, Lockhart PJ (2009) A statistical approach for distinguishing hybridization and incomplete lineage sorting. *The American Naturalist*, **174**, E54-E70.

Jones RJ, Hoegh-Guldberg O, Larkum AWD, Schreiber U (1998) Temperature-induced bleaching of corals begins with impairment of the CO2 fixation mechanism in zooxanthellae. *Plant, Cell & Environment*, **21**, 1219-1230.

Jones AM, Berkelmans R, van Oppen MJH, Mieog JC, Sinclair W (2008) A community change in the algal endosymbionts of a scleractinian coral following a natural bleaching event: field evidence of acclimatization. *Proceedings of the Royal Society of London B Biological Sciences*, **275**, 1359-1365.

Kaiser H, Sanchez C, Heacox S, Kathriner A, Ribeiro AV, Soares ZA, …, O'Shea M (2013) First Report on the Herpetofauna of Ataúro Island, Timor-Leste. *CheckList*, **9**.

Kawaguti S (1944) On the physiology of reef corals VII. Zooxanthellae of the reef corals is Gymnodium sp., dinoflagellate; its culture in vitro. *Palau Tropical Biological Station Studies*, **2**, 675-680.

Kawaguti S (1964) Zooxanthellae in the corals are intercellular symbionts. *Proceedings of the Japan Academy*, **40**, 545-548.

Kemp DW, Thornhill DJ, Rotjan RD, Iglesias-Prieto R, Fitt WK, Schmidt GW (2015) Spatially distinct and regionally endemic Symbiodinium assemblages in the threatened Caribbean reef-building coral Orbicella faveolata. *Coral Reefs*, **34**, 535-547.

Kennedy EV, Foster NL, Mumby PJ, Stevens JR (2015) Widespread prevalence of cryptic Symbiodinium D in the key Caribbean reef builder, Orbicella annularis. *Coral Reefs*, **34**, 519-531.

Kirtman B, Power SB, Adedoyin AJ, Boer GJ, Bojariu R, Camilloni I, ..., Prather M (2013) Near-term climate change: projections and predictability. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the*

Intergovernmental Panel on Climate Change (eds. Stocker TF, Qin D, …, Plattner GK). Cambridge University Press, Cambridge.

Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *Journal of Molecular Evolution*, **29**, 170-179.

Kopp C, Domart-Coulon I, Escrig S, Humbel BM, Hignette M, Meibom A (2015) Subcellular Investigation of Photosynthesis-Driven Carbon Assimilation in the Symbiotic Reef Coral Pocillopora damicornis. *mBio* **6**.

Kopp C, Domart-Coulon I, Barthelemy D, Meibom A (2016) Nutritional input from dinoflagellate symbionts in reef-building corals is minimal during planula larval life stage. *Science Advances*, **2**, e1500681.

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and Environmental Microbiology*, **79**, 5112-5120.

Krueger T, Hawkins TD, Becker S, Pontasch S, Dove S, Hoegh-Guldberg O, ..., Davy SK (2015) Differential coral bleaching—Contrasting the activity and response of enzymatic antioxidants in symbiotic partners under thermal stress. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, **190**, 15-25.

Ladner JT, Barshis DJ, Palumbi SR (2012) Protein evolution in two co-occurring types of Symbiodinium: an exploration into the genetic basis of thermal tolerance in Symbiodinium clade D. *BMC Evolutionary Biology*, **12**, 217.

LaJeunesse TC (2001) Investigating the biodiversity, ecology, and phylogeny of endosymbiotic dinoflagellates in the genus Symbiodinium using the internal transcribed spacer region: in search of a ″species″ level marker. *Journal of Phycology*, **37**, 866-880.

LaJeunesse TC (2002) Diversity and community structure of symbiotic dinoflagellates from Caribbean coral reefs. *Marine Biology*, **141**, 387-400.

LaJeunesse TC (2005) "Species" radiations of symbiotic dinoflagellates in the Atlantic and Indo-Pacific since the Miocene-Pliocene transition. *Molecular Biology and Evolution*, **22**, 570-581.

LaJeunesse TC, Trench RK (2000) Biogeography of two species of Symbiodinium (Freudenthal) inhabiting the intertidal sea anemone Anthopleura elegantissima (Brandt). *The Biological Bulletin*, **199**, 126-134.

LaJeunesse TC, Loh WK, van Woesik R, Hoegh-Guldberg O, Schmidt GW, Fitt WK (2003) Low symbiont diversity in southern Great Barrier Reef corals, relative to those of the Caribbean. *Limnology and Oceanography*, **48**, 2046-2054.

LaJeunesse TC, Bhagooli R, Hidaka M, DeVantier L, Done T, Schmidt GW, ..., Hoegh-Guldberg O (2004a) Closely related Symbiodinium spp. differ in relative dominance in coral reef host communities across environmental, latitudinal and biogeographic gradients. *Marine Ecology Progress Series*, **284**, 147-161.

LaJeunesse TC, Thornhill DJ, Cox EF, Stanton FG, Fitt WK, Schmidt GW (2004b) High diversity and host specificity observed among symbiotic dinoflagellates in reef coral communities from Hawaii. *Coral Reefs*, **23**, 596-603.

LaJeunesse TC, Lee S, Bush S, Bruno JF (2005) Persistence of non-Caribbean algal symbionts in Indo-Pacific mushroom corals released to Jamaica 35 years ago. *Coral Reefs*, **24**, 157-159.

LaJeunesse TC, Pinzon JH (2007) Screening intragenomic rDNA for dominant variants can provide a consistent retrieval of evolutionarily persistent ITS (rDNA) sequences. *Molecular Phylogenetics and Evolution*, **45**, 417-422.

LaJeunesse TC, Bonilla HR, Warner ME, Wills M, Schmidt GW, Fitt WK (2008) Specificity and stability in high latitude eastern Pacific coral-algal symbioses. *Limnology and Oceanography*, **53**, 719-727.

LaJeunesse TC, Smith R, Walther M, Pinzón J, Pettay DT, McGinley M, ..., Reyes-Bonilla H (2010a) Host–symbiont recombination versus natural selection in the response of coral–dinoflagellate symbioses to environmental disturbance. *Proceedings of the Royal Society of London B: Biological Sciences*, **277**, 2925-2934.

LaJeunesse TC, Pettay DT, Sampayo EM, Phongsuwan N, Brown B, Obura DO, ..., Fitt WK (2010b) Long-standing environmental conditions, geographic isolation and host–symbiont specificity influence the relative ecological dominance and genetic diversification of coral endosymbionts in the genus Symbiodinium. *Journal of Biogeography*, **37**, 785-800.

LaJeunesse TC, Thornhill DJ (2011) Improved resolution of reef-coral endosymbiont (Symbiodinium) species diversity, ecology, and evolution through psbA non-coding region genotyping. *PLoS One*, **6**, e29013.

LaJeunesse TC, Parkinson JE, Reimer JD (2012) A genetics-based description of Symbiodinium minutum sp. nov. and S. psygmophilum sp. nov. (Dinophyceae), two dinoflagellates symbiotic with cnidaria. *Journal of Phycology*, **48**, 1380-1391.

LaJeunesse TC, Wham DC, Pettay DT, Parkinson JE, Keshavmurthy S, Chen CA (2014) Ecologically differentiated stress-tolerant endosymbionts in the dinoflagellate genus Symbiodinium (Dinophyceae) Clade D are different species. *Phycologia*, **53**, 305-319.

LaJeunesse TC, Lee SY, Gil-Agudelo DL, Knowlton N, Jeong HJ (2015) Symbiodinium necroappetens sp. nov. (Dinophyceae): an opportunist 'zooxanthella'found in bleached and diseased tissues of Caribbean reef corals. *European Journal of Phycology*, **50**, 223-238.

Law R, Lewis DH (1983) Biotic environments and the maintenance of sex–some evidence from mutualistic symbioses. *Biological Journal of the Linnean Society*, **20**, 249-276.

Lee MJ, Jeong HJ, Jang SH, Lee SY, Kang NS, Lee KH, ..., LaJeunesse TC (2016) Most Low-Abundance "Background" Symbiodinium spp. Are Transitory and Have Minimal Functional Significance for Symbiotic Corals. *Microbial Ecology*, 1-13.

Leigh JW, Lapointe FJ, Lopez P, Bapteste E (2011) Evaluating phylogenetic congruence in the post-genomic era. *Genome Biology and Evolution*, **3**, 571-587.

Leigh JW, Bryant D (2015) PopART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, **6**, 1110-1116.

Leinonen PH, Remington DL, Savolainen O (2011) Local adaptation, phenotypic differentiation, and hybrid fitness in diverged natural populations of Arabidopsis lyrata. *Evolution*, **65**, 90-107.

Lesser MP (2006) Oxidative stress in marine environments: biochemistry and physiological ecology. *Annual Review Physiology*, **68**, 253-278.

Lesser MP (2011) Coral bleaching: causes and mechanisms. In: *Coral reefs: an ecosystem in transition* (eds Dubinsky Z, Stambler N), pp. 405-419. Springer, Dordrecht.

Levin RA, Beltran VH, Hill R, Kjelleberg S, McDougald D, Steinberg PD, van Oppen MJ (2016) Sex, scavengers, and chaperones: transcriptome secrets of divergent Symbiodinium thermal tolerances. *Molecular Biology and Evolution*, **33**, 2201-2215.

Lewis CL, Coffroth MA (2004) The acquisition of exogenous algal symbionts by an octocoral after bleaching. *Science*, **304**, 1490-1492.

Lopez P, Casane D, Philippe H (2002) Heterotachy, an important process of protein evolution. *Molecular Biology and Evolution*, **19**, 1-7.

Madduppa H, Schupp PJ, Faisal MR, Sastria MY, Thoms C (2017) Persistent outbreaks of the "black disease" sponge Terpios hoshinota in Indonesian coral reefs. *Marine Biodiversity*, **47**, 149-151.

Magalon H, Flot JF, Baudry E (2007) Molecular identification of symbiotic dinoflagellates in Pacific corals in the genus Pocillopora. *Coral Reefs*, **26**, 551-558.

Matthews JL, Crowder CM, Oakley CA, Lutz A, Roessner U, Meyer E. ..., Davy SK (2017) Optimal nutrient exchange and immune responses operate in partner specificity in the cnidarian-dinoflagellate symbiosis. *Proceedings of the National Academy of Sciences*, **114**, 13194-13199.

McAllister DE (1991) What is the status of the world's coral reef fishes. *Sea Wind*, **5**, 14-18.

McCauley DE (2013) Paternal leakage, heteroplasmy, and the evolution of plant mitochondrial genomes. *New Phytologist*, **200**, 966-977.

McGinley MP, Aschaffenburg MD, Pettay DT, Smith RT, LaJeunesse TC, Warner ME (2012) Symbiodinium spp. in colonies of eastern Pacific Pocillopora spp. are highly stable despite the prevalence of low-abundance background populations. *Marine Ecology Progress Series*, **462**, 1-7.

McGinty ES, Pieczonka J, Mydlarz LD (2012) Variations in reactive oxygen release and antioxidant activity in multiple Symbiodinium types in response to elevated temperature. *Microbial Ecology*, **64**, 1000-1007.

Meng C, Kubatko LS (2009) Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: a model. *Theoretical Population Biology*, **75**, 35-45.

Mieog JC, van Oppen MJ, Cantin NE, Stam WT, Olsen JL (2007) Real-time PCR reveals a high incidence of Symbiodinium clade D at low levels in four scleractinian corals across the Great Barrier Reef: implications for symbiont shuffling. *Coral Reefs*, **26**, 449-457.

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

Moon CD, Craven KD, Leuchtmann A, Clement SL, Schardl CL (2004) Prevalence of interspecific hybrids amongst asexual fungal endophytes of grasses. *Molecular Ecology*, **13**, 1455-1467.

Moore RB, Ferguson KM, Loh WK, Hoegh-Guldberg O, Carter DA (2003) Highly organized structure in the non-coding region of the psbA minicircle from clade C Symbiodinium. *International Journal of Systematic and Evolutionary Microbiology*, **53**, 1725-1734.

Moran EV, Alexander JM (2014) Evolutionary responses to global change: lessons from invasive species. *Ecology Letters*, **17**, 637-649.

Muyzer G, Smalla K (1998) Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie van Leeuwenhoek*, **73**, 127-141.

Muscatine L, Hand C (1958) Direct evidence for the transfer of materials from symbiotic algae to the tissues of a coelenterate. *Proceedings of the National Academy of Sciences*, **44**, 1259-1263.

Muscatine L, Cook CB, Pardy RL, Pool RR (1974) Uptake, recognition and maintenance of symbiotic Chlorella by Hydra viridis. *In Symposia of the Society for Experimental Biology*, **29**, 175-203.

Muscatine L, Porter JW (1977) Reef corals: mutualistic symbioses adapted to nutrient-poor environments. *Bioscience*, **27**, 454-460.

Muscatine L, Falkowski PG, Porter JW, Dubinsky Z (1984) Fate of photosynthetic fixed carbon in light-and shade-adapted colonies of the symbiotic coral Stylophora pistillata. *Proceedings of the Royal Society of London B: Biological Sciences*, **222**, 181-202.

Muscatine L, Goiran C, Land L, Jaubert J, Cuif JP, Allemand D (2005) Stable isotopes (δ^{13}C and δ^{15}N) of organic matrix from coral skeleton. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 1525-1530.

Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. *Annual Review of Genetics,* **39**, 121-152.

Nolte AW, Freyhof J, Stemshorn KC, Tautz D (2005) An invasive lineage of sculpins, Cottus sp. (Pisces, Teleostei) in the Rhine with new habitat adaptations has originated from hybridization between old phylogeographic groups. *Proceedings of the Royal Society of London B: Biological Sciences*, **272**, 2379-2387.

Odum HT, Odum EP (1955) Trophic structure and productivity of a windward coral reef community on Eniwetok Atoll. *Ecological Monographs*, **25**, 291-320.

Oksanen J, Blanchet FG, Kindt R, Legendre P, O'hara RB, Simpson GL, ..., Wagner H (2017) vegan: Community Ecology Package. R package version 2.4-5

Oliver TA, Palumbi SR (2011) Many corals host thermally resistant symbionts in high-temperature habitat. *Coral Reefs*, **30**, 241-250.

Ortiz JC, González-Rivero M, Mumby PJ (2013) Can a thermally tolerant symbiont improve the future of Caribbean coral reefs?. *Global Change Biology*, **19**, 273-281.

Padilla-Gamiño JL, Pochon X, Bird C, Concepcion GT, Gates RD (2012) From parent to gamete: vertical transmission of Symbiodinium (Dinophyceae) ITS2 sequence assemblages in the reef building coral Montipora capitata. *PLoS One*, **7**, e38440.

Palumbi SR, Barshis DJ, Traylor-Knowles N, Bay RA (2014) Mechanisms of reef coral resistance to future climate change. *Science*, **344**, 895-898.

Pandolfi JM, Connolly SR, Marshall DJ, Cohen AL (2011) Projecting coral reef futures under global warming and ocean acidification. *Science*, **333**, 418-422.

Parkinson JE, Baums IB (2014) The extended phenotypes of marine symbioses: ecological and evolutionary consequences of intraspecific genetic diversity in coral–algal associations. *Frontiers in Microbiology*, **5**, 445.

Parkinson JE, Banaszak AT, Altman NS, LaJeunesse TC, Baums IB (2015a) Intraspecific diversity among partners drives functional variation in coral symbioses. *Scientific Reports*, **5**, 15667.

Parkinson JE, Coffroth MA, LaJeunesse TC (2015b) New species of Clade B Symbiodinium (Dinophyceae) from the greater Caribbean belong to different functional guilds: S. aenigmaticum sp.nov., S. antillogorgium sp.nov., S. endomadracis sp.nov., and S. pseudominutum sp.nov. *Journal of Phycology*, **51**, 850-858.

Parkinson JE, Baumgarten S, Michell CT, Baums IB, LaJeunesse TC, Voolstra CR (2016) Gene expression variation resolves species and individual strains among coral-associated dinoflagellates within the genus Symbiodinium. *Genome Biology and Evolution*, **8**, 665-680.

Pelser PB, Kennedy AH, Tepe EJ, Shidler JB, Nordenstam B, Kadereit JW, Watson LE (2010) Patterns and causes of incongruence between plastid and nuclear Senecioneae (Asteraceae) phylogenies. *American Journal of Botany*, **97**, 856-873.

Pettay DT, LaJeunesse TC (2007) Microsatellites from clade B Symbiodinium spp. specialized for Caribbean corals in the genus Madracis. *Molecular Ecology Notes*, **7**, 1271-1274.

Pettay DT, Wham DC, Pinzon JH, LaJeunesse TC (2011) Genotypic diversity and spatial–temporal distribution of Symbiodinium clones in an abundant reef coral. *Molecular Ecology*, **20**), 5197-5212.

Pinzon JH, LaJeunesse TC (2011) Species delimitation of common reef corals in the genus Pocillopora using nucleotide sequence phylogenies, population genetics and symbiosis ecology. *Molecular Ecology*, **20**, 311-325.

Planet PJ (2006) Tree disagreement: measuring and testing incongruence in phylogenies. *Journal of Biomedical Informatics*, **39**, 86-102.

Pochon X, Pawlowski J, Zaninetti L, Rowan R (2001) High genetic diversity and relative specificity among Symbiodinium-like endosymbiotic dinoflagellates in soritid foraminiferans. *Marine Biology*, **139**, 1069-1078.

Pochon X, Montoya-Burgos JI, Stadelmann B, Pawlowski J (2006) Molecular phylogeny, evolutionary rates, and divergence timing of the symbiotic dinoflagellate genus Symbiodinium. *Molecular Phylogenetics and Evolution*, **38**, 20-30.

Pochon X, Pawlowski J (2006) Evolution of the soritids-Symbiodinium symbiosis. *Symbiosis*, **42**, 77-88.

Pochon X, Gates RD (2010) A new Symbiodinium clade (Dinophyceae) from soritid foraminifera in Hawai'i. *Molecular Phylogenetics and Evolution*, **56**, 492-497.

Pochon X, Putnam HM, Burki F, Gates RD (2012) Identifying and characterizing alternative molecular markers for the symbiotic and free-living dinoflagellate genus Symbiodinium. *PLoS One*, **7**, e29816.

Pochon X, Putnam HM, Gates RD (2014) Multi-gene analysis of Symbiodinium dinoflagellates: a perspective on rarity, symbiosis, and evolution. *PeerJ*, **2**, e394.

Poland DM, Mansfield JM, Hannes AR, Fairbank CL, Shearer TL, Connelly SJ, ..., Coffroth MA (2013) Variation in Symbiodinium communities in juvenile Briareum asbestinum (Cnidaria: Octocorallia) over temporal and spatial scales. *Marine Ecology Progress Series*, **476**, 23-37.

Pollard DA, Iyer VN, Moses AM, Eisen MB (2006) Widespread discordance of gene trees with species tree in Drosophila: evidence for incomplete lineage sorting. *PLoS Genetics*, **2**, e173.

Pollock FJ, Lamb JB, Field SN, Heron SF, Schaffelke B, Shedrawi G, ..., Willis BL (2014) Sediment and turbidity associated with offshore dredging increase coral disease prevalence on nearby reefs. *PLoS One*, **9**, e102498.

Pontasch S, Hill R, Deschaseaux E, Fisher PL, Davy SK, Scott A (2014) Photochemical efficiency and antioxidant capacity in relation to Symbiodinium genotype and host phenotype in a symbiotic cnidarian. *Marine Ecology Progress Series*, **516**, 195-208.

Pratchett MS, Munday P, Wilson SK, Graham NA, Cinner JE, Bellwood DR, ..., McClanahan TR (2008) Effects of climate-induced coral bleaching on coral-reef fishes - Ecological and economic consequences. *Oceanography and Marine Biology: Annual Review*, **46**, 251-296.

Pratchett MS, McCowan D, Maynard JA, Heron SF (2013) Changes in bleaching susceptibility among corals subject to ocean warming and recurrent bleaching in Moorea, French Polynesia. *PLoS One*, **8**, e70443.

Putnam HM, Stat M, Pochon X, Gates RD (2012) Endosymbiotic flexibility associates with environmental sensitivity in scleractinian corals. *Proceedings of the Royal Society of London B: Biological Sciences*, **279**, 4352-4361.

Quigley KM, Davies SW, Kenkel CD, Willis BL, Matz MV, Bay LK (2014) Deep-sequencing method for quantifying background abundances of Symbiodinium types: exploring the rare Symbiodinium biosphere in reef-building corals. *PloS One*, **9**, e94297.

Qvarnstrom Y, Sullivan JJ, Bishop HS, Hollingsworth R, da Silva AJ (2007) PCR-based detection of Angiostrongylus cantonensis in tissue and mucus secretions from molluscan hosts. *Applied and Environmental Microbiology*, **73**, 1415-1419.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. *Conservation Biology*, **17**, 230-237.

Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics*, **27**, 83-109.

Rieseberg LH, Whitton J, Linder CR (1996) Molecular marker incongruence in plant hybrid zones and phylogenetic trees. *Acta Botanica Neerlandica*, **45**, 243-262.

Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, ..., Lexer C (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, **301**, 1211-1216.

Robinson DA, Monk AB, Cooper JE, Feil EJ, Enright MC (2005) Evolutionary genetics of the accessory gene regulator (agr) locus in Staphylococcus aureus. *Journal of Bacteriology*, **187**, 8312-8321.

Rodriguez-Lanetty M (2003) Evolving lineages of Symbiodinium-like dinoflagellates based on ITS1 rDNA. *Molecular Phylogenetics and Evolution*, **28**, 152-168.

Rodriguez RJ, White JF, Arnold AE, Redman RS (2009) Fungal endophytes: diversity and functional roles. *New Phytologist*, **182**, 314-330.

Rohland N, Reich D (2012) Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, **22**, 939-946.

Rouzé H, Lecellier G, Saulnier D, Berteaux-Lecellier V (2016) Symbiodinium clades A and D differentially predispose Acropora cytherea to disease and Vibrio spp. colonization. *Ecology and Evolution*, **6**, 560-572.

Rouzé H, Lecellier GJ, Saulnier D, Planes S, Gueguen Y, Wirshing HH, Berteaux-Lecellier V (2017) An updated assessment of Symbiodinium spp. that associate with common scleractinian corals from Moorea (French Polynesia) reveals high diversity among background symbionts and a novel finding of clade B. *PeerJ*, **5**, e2856.

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

Rowan R, Powers DA (1991a) A molecular genetic classification of zooxanthellae and the evolution of animal-algal symbioses. *Science*, **251**, 1348-1351.

Rowan R, Powers DA (1991b) Molecular genetic identification of symbiotic dinoflagellates (zooxanthellae). *Marine Ecology Progress Series*, **71**, 65-73.

Rowan R, Powers DA (1992) Ribosomal RNA sequences and the diversity of symbiotic dinoflagellates (zooxanthellae). *Proceedings of the National Academy of Sciences*, **89**, 3639-3643.

Rowan R, Knowlton N (1995) Intraspecific diversity and ecological zonation in coral-algal symbiosis. *Proceedings of the National Academy of Sciences*, **92**, 2850-2853.

Rowan R, Knowlton N, Baker AC, Jara J (1997) Landscape ecology of algal symbionts creates variation in episodes of coral bleaching. *Nature*, **388**, 265-269.

Rybalka N, Wolf M, Andersen RA, Friedl T (2013) Congruence of chloroplast-and nuclear-encoded DNA sequence variations used to assess species boundaries in the soil microalga Heterococcus (Stramenopiles, Xanthophyceae). *BMC Evolutionary Biology*, **13**, 39.

Sachs JL, Wilcox TP (2006) A shift to parasitism in the jellyfish symbiont Symbiodinium microadriaticum. *Proceedings of the Royal Society of London Series B*, **273**, 425-429.

Sampayo EM, Ridgway T, Bongaerts P, Hoegh-Guldberg O (2008) Bleaching susceptibility and mortality of corals are determined by fine-scale differences in symbiont type. *Proceedings Natural Academy of Sciences*, **105**, 10444-10449.

Sampayo EM, Dove S, LaJeunesse TC (2009) Cohesive molecular genetic data delineate species diversity in the dinoflagellate genus Symbiodinium. *Molecular Ecology*, **18**, 500-519.

Sang T, Zhong Y (2000) Testing hybridization hypotheses based on incongruent gene trees. *Systematic Biology*, **49**, 422-434.

Santos SR, Taylor DJ, Coffroth MA (2001) Genetic comparisons of freshly isolated versus cultured symbiotic dinoflagellates: implications for extrapolating to the intact symbiosis. *Journal of Phycology*, **37**, 900-912.

Santos SR, Taylor DJ, Kinzie III RA, Hidaka M, Sakai K, Coffroth MA (2002) Molecular phylogeny of symbiotic dinoflagellates inferred from partial chloroplast large subunit (23S)-rDNA sequences. *Molecular Phylogenetics and Evolution*, **23**, 97-111.

Santos SR, Coffroth MA (2003) Molecular genetic evidence that dinoflagellates belonging to the genus Symbiodinium Freudenthal are haploid. *The Biological Bulletin*, **204**, 10-20.

Santos SR, Shearer TL, Hannes AR, Coffroth MA (2004) Fine-scale diversity and specificity in the most prevalent lineage of symbiotic dinoflagellates (Symbiodinium, Dinophyceae) of the Caribbean. *Molecular Ecology*, **13**, 459-469.

Santos SR (2014) Expanding the population genetic perspective of cnidarian-Symbiodinium symbiosis. *Molecular Ecology*, **23**, 4185-4187.

Scambary J (2009) Anatomy of a conflict: the 2006–2007 communal violence in East Timor. *Conflict, Security & Development*, **9**, 265-288.

Schardl CL, Leuchtmann A, Tsai HF, Collett MA, Watt DM, Scott DB (1994) Origin of a fungal symbiont of perennial ryegrass by interspecific hybridization of a mutualist with the ryegrass choke pathogen, Epichloe typhina. *Genetics*, **136**, 1307-1317.

Schardl CL, Craven KD (2003) Interspecific hybridization in plant-associated fungi and oomycetes: a review. *Molecular Ecology*, **12**, 2861-2873.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, ..., Sahl JW (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, **75**, 7537-7541.

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, ..., Miller AN (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences*, **109**, 6241-6246.

Schoenberg DA, Trench RK (1980a) Genetic variation in Symbiodinium (= Gymnodinium) microadriaticum Freudenthal, and specificity in its symbiosis with marine invertebrates. I. Isoenzyme and soluble protein patterns of axenic cultures of Symbiodinium microadriaticum. *Proceedings of the Royal Society of London B: Biological Sciences*, **207**, 405-427.

Schoenberg DA, Trench RK (1980b) Genetic variation in Symbiodinium (=Gymnodinium) microadriaticum Freudenthal, and specificity in its symbiosis with marine invertebrates. II. Morphological variation in S. microadriaticum. *Proceedings of the Royal Society of London B: Biological Sciences*, **207**, 429-444.

Schoenberg DA, Trench RK (1980c) Genetic variation in Symbiodinium (=Gymnodinium) microadriaticum Freudenthal, and specificity in its symbiosis with marine invertebrates. III. Specificity and infectivity of S. microadriaticum. *Proceedings of the Royal Society of London B: Biological Sciences*, **207**, 445-460.

Schwarz JA, Krupp DA, Weis VM (1999) Late larval development and onset of symbiosis in the scleractinian coral Fungia scutaria. *The Biological Bulletin*, **196**, 70-79.

Seewaldt E, Stackebrandt E (1982) Partial sequence of 16S ribosomal RNA and the phylogeny of Prochloron. *Nature*, **295**, 618-620.

Sheppard C, Davy SK, Pilling G, Graham N (2017) The biology of coral reefs (Second Edition). Oxford University Press, 384pp.

Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. *Systems Biology*, **51**, 492-508.

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, **16**, 1114-1116.

Shymanovich T, Charlton ND, Musso AM, Scheerer J, Cech NB, Faeth SH, Young CA (2017) Interspecific and intraspecific hybrid Epichloë species symbiotic with the North American native grass Poa alsodes. *Mycologia*, **109**, 459-474.

Silove D, Liddell B, Rees S, Chey T, Nickerson A, Tam N, ..., Steel Z (2014) Effects of recurrent violence on post-traumatic stress disorder and severe distress in conflict-affected Timor-Leste: a 6-year longitudinal study. *The Lancet Global Health*, **2**, e293-e300.

Silva-Lima AW, Walter JM, Garcia GD, Ramires N, Ank G, Meirelles PM, ..., Thompson CC (2015) Multiple Symbiodinium Strains Are Hosted by the Brazilian Endemic Corals Mussismilia spp. *Microbial Ecology*, **70**, 301-310.

Silverstein RN, Correa AMS, LaJeunesse TC, Baker AC (2011) Novel algal symbiont (Symbiodinium spp.) diversity in reef corals of Western Australia. *Marine Ecology Progress Series*, **422**, 63-75.

Silverstein RN, Correa AMS, Baker AC (2012) Specificity is rarely absolute in coral-algal symbiosis: implications for coral response to climate change. *Proceedings of the Royal Society B: Biological Sciences,* **279**, 2609-2618.

Silverstein RN, Cunning R, Baker AC (2015) Change in algal symbiont communities after bleaching, not prior heat exposure, increases heat tolerance of reef corals. *Global Change Biology*, **21**, 236-249.

Smith RT, Pinzon JH, LaJeunesse TC (2009) Symbiodinium (Dinophyta) diversity and stability in aquarium corals. *Journal of Phycology*, **45**, 1030-1036.

Som A (2014) Causes, consequences and solutions of phylogenetic incongruence. *Briefings in Bioinformatics*, **16**, 536-548.

Sota T, Vogler AP (2001) Incongruence of mitochondrial and nuclear gene trees in the carabid beetles Ohomopterus. *Systematic Biology*, **50**, 39-59.

Sprintall J, Wijffels S, Gordon AL, Ffield A, Molcard R, Susanto RD, ..., Aken HM (2004) INSTANT: A new international array to measure the Indonesian Throughflow. *Eos, Transactions American Geophysical Union,* **85**, 369-376.

Stat M, Carter D, Hoegh-Guldberg O (2006) The evolutionary history of Symbiodinium and scleractinian hosts—Symbiosis, diversity, and the effect of climate change. *Perspectives in Plant Ecology, Evolution and Systematics*, **8**, 23-43.

Stat M, Loh WKW, Hoegh-Guldberg O, Carter DA (2008) Symbiont acquisition strategy drives host–symbiont associations in the southern Great Barrier Reef. *Coral Reefs*, **27**, 763-772.

Stat M, Pochon X, Cowie RO, Gates RD (2009) Specificity in communities of Symbiodinium in corals from Johnston Atoll. *Marine Ecology Progress Series*, **386**, 83-96.

Stat M, Gates RD (2010) Clade D Symbiodinium in scleractinian corals: a "nugget" of hope, a selfish opportunist, an ominous sign, or all of the above?. *Journal of Marine Biology*, **9**.

Stat M, Pochon X, Franklin EC, Bruno JF, Casey KS, Selig ER, Gates RD (2013) The distribution of the thermally tolerant symbiont lineage (Symbiodinium clade D) in corals from Hawaii: correlations with host and the history of ocean thermal stress. *Ecology and Evolution*, **3**, 1317-1329.

Stat M, Yost DM, Gates RD (2015) Geographic structure and host specificity shape the community composition of symbiotic dinoflagellates in corals from the Northwestern Hawaiian Islands. *Coral Reefs*, **34**, 1075-1086.

Strimmer K, Rambaut A (2002) Inferring confidence sets of possibly misspecified gene trees. *Proceedings of the Royal Society of London B: Biological Sciences*, **269**, 137-142.

Suggett DJ, Warner ME, Smith DJ, Davey P, Hennige S, Baker NR (2008).Photosynthesis and production of hydrogen peroxide by Symbiodinium (Pyrrhophyta) phylotypes with different thermal tolerances. *Journal of Phycology*, **44**, 948-956.

Sun M, Soltis DE, Soltis PS, Zhu X, Burleigh JG, Chen Z (2015) Deep phylogenetic incongruence in the angiosperm clade Rosidae. *Molecular Phylogenetics and Evolution*, **83**, 156-166.

Swain TD, Chandler J, Backman V, Marcelino L (2017) Consensus thermotolerance ranking for 110 Symbiodinium phylotypes: an exemplar utilization of a novel iterative partial-rank aggregation tool with broad application potential. *Functional Ecology*, **31**, 172-183.

Swofford DL (2002) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0a159. Sinauer Associates, Sunderland, Massachusetts

Takabayashi M, Santos SR, Cook CB (2004) Mitochondrial DNA phylogeny of the symbiotic dinoflagellates (Symbiodinium, Dinophyta). *Journal of Phycology*, **40**, 160-164.

Takahashi S, Yoshioka-Nishimura M, Nanba D, Badger MR (2013) Thermal acclimation of the symbiotic alga Symbiodinium spp. alleviates photobleaching under heat stress. *Plant Physiology*, **161**, 477-485.

Takishita K, Ishikura M, Koike K, Maruyama T (2003) Comparison of phylogenies based on nuclear-encoded SSU rDNA and plastid-encoded psbA in the symbiotic dinoflagellate genus Symbiodinium. *Phycologia*, **42**, 285-291.

Tchernov D, Gorbunov MY, de Vargas C, Narayan Yadav S, Milligan AJ, Haggblom M, Falkowski PG (2004) Membrane lipids of symbiotic algae are diagnostic of sensitivity to thermal bleaching in corals. *Proceedings of the National Academy of Sciences USA*, **101**, 13531-13535.

Templeton AR (1983) Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution*, **37**, 221-244.

Thomas L, Kendrick GA, Kennington WJ, Richards ZT, Stat M (2014) Exploring Symbiodinium diversity and host specificity in Acropora corals from geographical extremes of Western Australia with 454 amplicon pyrosequencing. *Molecular Ecology*, **23**, 3113-3126.

Thornhill DJ, LaJeunesse TC, Kemp D, Fitt W, Schmidt G (2006) Multi-year, seasonal genotypic surveys of coral-algal symbioses reveal prevalent stability or post-bleaching reversion. *Marine Biology*, **148**, 711-722.

Thornhill DJ, LaJeunesse TC, Santos SR (2007) Measuring rDNA diversity in eukaryotic microbial systems: how intragenomic variation, pseudogenes, and PCR artifacts confound biodiversity estimates. *Molecular Ecology*, **16**, 5326-5340.

Thornhill DJ, Lewis AM, Wham DC, LaJeunesse TC (2014) Host-specialist lineages dominate the adaptive radiation of reef coral endosymbionts. *Evolution*, **68**, 352-367.

Thornhill DJ, Howells EJ, Wham DC, Steury TD, Santos SR (2017) Population genetics of reef coral endosymbionts (Symbiodinium, Dinophyceae). *Molecular Ecology*, **26**, 2640-2659.

Tonk L, Sampayo EM, Weeks S, Magno-Canto M, Hoegh-Guldberg O (2013) Host-specific interactions with environmental factors shape the distribution of Symbiodinium across the Great Barrier Reef. *PLoS One*, **8**, e68533.

Trainor CR, Soares T (2004) Birds of Atauro Island, Timor Leste (East Timor). *Forktail*, **20**, 41-48.

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

Trench RK (1993) Microalgal-invertebrate symbioses: a review. *Endocytobiosis Cell Research*, **9**, 135-175.

Trench RK (1997) Diversity of symbiotic dinoflagellates and the evolution of microalgal-invertebrate symbioses. In *Proceedings of the 8th International Coral Reef Symposium*, **2**, 1275-1286.

Trench RK, Blank RJ (1987) Symbiodinium microadriaticum Freudenthal, S. goreauii sp. nov., S. kawagutii sp. nov. and S. pilosum sp. nov.: gymnodinioid dinoflagellate symbionts of marine invertebrates. *Journal of Phycology*, **23**, 469-481.

Tsai HF, Liu JS, Staben C, Christensen MJ, Latch GC, Siegel MR, Schardl CL (1994) Evolutionary diversification of fungal endophytes of tall fescue grass by hybridization with Epichloë species. *Proceedings of the National Academy of Sciences*, **91**, 2542-2546.

Turak E, DeVantier L (2013) Reef-building corals in Timor-Leste. A rapid marine biological assessment of Timor-Leste. Timor Leste, Dili: Coral Triangle Support Partnership, Conservation International.

Ulstrup KE, van Oppen MJH (2003) Geographic and habitat partitioning of genetically distinct zooxanthellae (Symbiodinium) in Acropora corals on the Great Barrier Reef. *Molecular Ecology*, **12**, 3477-3484.

Ulstrup KE, Hill R, van Oppen MJH, Larkum AWD, Ralph PJ (2008) Seasonal variation in the photo-physiology of homogeneous and heterogeneous Symbiodinium consortia in two scleractinian corals. *Marine Ecology Progress Series*, **361**, 139-150.

van der Heijden MG, Bardgett RD, van Straalen NM (2008) The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology Letters*, **11**, 296-310.

van Oppen MJH (2001) In vitro establishment of symbiosis in Acropora millepora planulae. *Coral Reefs*, **20**, 200.

van Oppen MJH, Palstra FP, Piquet AMT, Miller DJ (2001a) Patterns of coral–dinoflagellate associations in Acropora: significance of local availability and physiology of Symbiodinium strains and host–symbiont selectivity. *Proceedings of the Royal Society of London B: Biological Sciences*, **268**, 1759-1767.

van Oppen MJH, McDonald BJ, Willis B, Miller DJ (2001b) The evolutionary history of the coral genus Acropora (Scleractinia, Cnidaria) based on a mitochondrial and a nuclear marker: reticulation, incomplete lineage sorting, or morphological convergence?. *Molecular Biology and Evolution*, **18**, 1315-1329.

van Oppen MJH, Wörheide G, Takabayashi M (2002) Nuclear markers in evolutionary and population genetic studies of scleractinian corals and sponges. In *Proceedings of the 9th International Coral Reef Symposium*, **1**, 131-138.

van Oppen MJH, Mieog JC, Sanchez CA, Fabricius KE (2005) Diversity of algal endosymbionts (zooxanthellae) in octocorals: the roles of geography and host relationships. *Molecular Ecology*, **14**, 2403-2417.

van Oppen MJH, Souter P, Howells EJ, Heyward A, Berkelmans R (2011a) Novel genetic diversity through somatic mutations: fuel for adaptation of reef corals? *Diversity*, **3**, 405-423.

van Oppen MJH, Bongaerts P, Underwood JN, Peplow LM, Cooper TF (2011b) The role of deep reefs in shallow reef recovery: an assessment of vertical connectivity in a brooding coral from west and east Australia. *Molecular Ecology*, **20**, 1647-1660.

van Oppen MJH, Oliver JK, Putnam HM, Gates RD (2015) Building coral reef resilience through assisted evolution. *Proceedings of the National Academy of Sciences*, **112**, 2307-2313.

Venn AA, Loram JE, Douglas AE (2008) Photosynthetic symbioses in animals. *Journal of Experimental Botany,* **59**, 1069-1080.

Veron JEN, Stafford-Smith M (2000) Corals of the World. Volumes 1-3. Australian Institute of Marine Science, Townsville, Australia. 1382pp

Wang J, Douglas AE (1998) Nitrogen recycling or nitrogen conservation in an alga-invertebrate symbiosis?. *Journal of Experimental Biology*, **201**, 2445-2453.

Warner ME, Fitt WK, Schmidt GW (1999) Damage to photosystem II in symbiotic dinoflagellates: a determinant of coral bleaching. *Proceedings of the National Academy of Sciences*, **96**, 8007-8012.

Watanabe T, Kii SI, Tanaka J, Takishita K, Maruyama T (2006) cDNA cloning and phylogenetic and expression analyses of actin in symbiotic dinoflagellates (Symbiodinium spp.). *Journal of Applied Phycology*, **18**, 219-225.

Wei NV, Wallace CC, Dai C, Pillay KRM, Chen CA (2006) Analyses of the ribosomal internal transcribed spacers (ITS) and the 5.8 S gene indicate that extremely high rDNA heterogeneity is a unique feature in the scleractinian coral genus Acropora (Scleractinia; Acroporidae). *Zoological Studies*, **45**, 404-418.

Weis VM (2008) Cellular mechanisms of Cnidarian bleaching: stress causes the collapse of symbiosis. *Journal of Experimental Biology*, **211**, 3059-3066.

Weis VM, Davy SK, Hoegh-Guldberg O, Rodriguez-Lanetty M, Pringle JR (2008) Cell biology in model systems as the key to understanding corals. *Trends in Ecology and Evolution*, **23**, 369-376.

Wendel JF, Stewart JM, Rettig JH (1991) Molecular evidence for homoploid reticulate evolution among Australian species of Gossypium. *Evolution*, **45**, 694-711.

Werren JH, Baldo L, Clark ME (2008) Wolbachia: master manipulators of invertebrate biology. *Nature Reviews Microbiology*, **6**, 741-751.

Wham DC, Pettay DT, LaJeunesse TC (2011) Microsatellite loci for the host-generalist "zooxanthella" Symbiodinium trenchi and other Clade D Symbiodinium. *Conservation Genetics Resources*, **3**, 541-544.

Wham DC, Carmichael M, LaJeunesse TC (2014) Microsatellite loci for Symbiodinium goreaui and other Clade C Symbiodinium. *Conservation Genetics Resources*, **6**, 127-129.

Wham DC, LaJeunesse TC (2016) Symbiodinium population genetics: testing for species boundaries and analysing samples with mixed genotypes. *Molecular Ecology*, **25**, 2699-2712.

Wham DC, Ning G, LaJeunesse TC (2017) Symbiodinium glynnii sp. nov., a species of stress-tolerant symbiotic dinoflagellates from pocilloporid and montiporid corals in the Pacific Ocean. *Phycologia,* **56**, 396-409.

Wilcox TP (1998) Large-subunit ribosomal RNA systematics of symbiotic dinoflagellates: morphology does not recapitulate phylogeny. *Molecular Phylogenetics and Evolution*, **10**, 436-448.

Wilkinson SP (2015) Intra-genomic variation in symbiotic dinoflagellates: Recent divergence or natural hybridization? Doctoral Thesis, Victoria University of Wellington, New Zealand.

Wilkinson SP, Fisher PL, van Oppen MJH, Davy SK (2015) Intra-genomic variation in symbiotic dinoflagellates: recent divergence or recombination between lineages?, *BMC Evolutionary Biology*, **15**, 1.

Wilkinson SP, Pontasch S, Fisher PL, Davy SK (2016) The distribution of intra-genomically variable dinoflagellate symbionts at Lord Howe Island, Australia. *Coral Reefs*, **35**, 565-576.

Williams GJ, Knapp IS, Maragos JE, Davy SK (2010) Modeling patterns of coral bleaching at a remote Central Pacific atoll. Marine Pollution Bulletin, 60, 1467-1476.

Willis BL, van Oppen MJH, Miller DJ, Vollmer SV, Ayre DJ (2006) The role of hybridization in the evolution of reef corals. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 489-517.

Yaakub SM, Bellwood DR, van Herwerden L (2007) A rare hybridization event in two common Caribbean wrasses (genus Halichoeres; family Labridae). *Coral Reefs*, **26**, 597-602.

Yamashita H, Suzuki G, Hayashibara T, Koike K (2011) Do corals select zooxanthellae by alternative discharge?. *Marine Biology*, **158**, 87-100.

Yamashita H, Suzuki G, Hayashibara T, Koike K (2013) Acropora recruits harbor "rare" Symbiodinium in the environmental pool. *Coral Reefs*, **32**, 355-366.

Yang SY, Keshavmurthy S, Obura D, Sheppard CR, Visram S, Chen CA (2012) Diversity and distribution of Symbiodinium associated with seven common coral species in the Chagos Archipelago, central Indian Ocean. *PLoS One*, **7**, e35836.

Yao H, Song J, Liu C, Luo K, Han J, Li Y, ..., Chen S (2010) Use of ITS2 region as the universal DNA barcode for plants and animals. *PloS One*, **5**, e13102.

Yellowlees D, Rees TAV, Leggat W (2008) Metabolic interactions between algal symbionts and invertebrate hosts. *Plant, Cell & Environment*, **31**, 679-694.

Yorifuji M, Takeshima H, Mabuchi K, Watanabe T, Nishida M (2015) Comparison of Symbiodinium dinoflagellate flora in sea slug populations of the Pteraeolidia ianthina complex. *Marine Ecology Progress Series*, **521**, 91-104.

Zapalski MK (2014) Evidence of photosymbiosis in Palaeozoic tabulate corals. *Proceedings of the Royal Society of London B: Biological Sciences*, **281**, 20132663.

Zhang H, Bhattacharya D, Lin S (2005) Phylogeny of dinoflagellates based on mitochondrial cytochrome B and nuclear small subunit rDNA sequence comparisons. *Journal of Phycology*, **41**, 411-420.

Zhou GW, Huang H (2011) Low genetic diversity of symbiotic dinoflagellates (Symbiodinium) in scleractinian corals from tropical reefs in southern Hainan Island, China. *Journal of Systematics and Evolution*, **49**, 598-605.

Ziegler M, Arif C, Burt JA, Dobretsov S, Roder C, LaJeunesse TC, Voolstra CR (2017a) Biogeography and molecular diversity of coral symbionts in the genus Symbiodinium around the Arabian Peninsula. *Journal of Biogeography*, **44**, 674-686.

Ziegler M, Eguíluz VM, Duarte CM, Voolstra CR (2017b) Rare symbionts may contribute to the resilience of coral–algal assemblages. *The ISME Journal*, **12**, 161.

Appendix A1: Appendices for Chapter 2

A1.1. Full details of all cycling conditions for PCR amplifications

A1.1.1. Amplification of the *cob* gene

Dinocob1F/Dinocob1R primers

Seventy-five samples were amplified using cycling conditions of denaturation for 3 min at 95ºC, followed by 35 cycles of 15 s at 95ºC, 15 s at 56ºC, and 10 s at 72ºC. Reactions consisted of 12.5 µL MyTaq HS Red Mix 2x (BioLine), ~20 ng sample DNA, 10 µg BSA, 0.4 µM of each primer and 7.5 µL PCRH$_2$O for a total reaction volume of 25 µL.

Eighty-four samples were amplified using cycling conditions of denaturation for 1min at 95ºC, followed by 40 cycles of 15 s at 95ºC, 20 s at 55ºC, and 30 s at 72ºC. Reactions consisted of 10 µL MyTaq HS Red Mix 2x (BioLine), ~20 ng sample DNA, 10 µg BSA, 0.5 µM of each primer and 5 µL PCRH$_2$O for a total reaction volume of 20 µL.

Ninety-three samples were amplified using the final optimised conditions described in the methods, leading to a total of 252 successful amplifications with the primer pair Dino*cob*1F/Dino*cob*1R.

Cob_f1/Cob_r1 primers

All reactions consisted of 10 µL MyTaq HS Red Mix 2x (BioLine), ~20 ng sample DNA, 10 µg BSA, 0.25 µM of each primer and 6 µL PCRH$_2$O for a total volume of 20 µL.

Ten samples were amplified using cycling conditions of denaturation for 2 min at 95ºC, followed by 35 cycles of 20 s at 95ºC, 20 s at 59ºC, 30 s at 72ºC, and a final extension for 7 min at 72ºC.

Eleven samples were amplified using cycling conditions of denaturation for 2 min at 95ºC, followed by 38 cycles of 20 s at 95ºC, 20 s at 58ºC, 30 s at 72ºC, and a final extension for 7 min at 72ºC.

Twelve samples were amplified using cycling conditions of denaturation for 2 min at 95ºC, followed by 35 cycles of 20 s at 95ºC, 20 s at 58ºC, 30 s at 72ºC, and a final extension for 7 min at 72ºC.

A total of 271 samples were amplified using the final optimised conditions described in the methods, leading to a total of 304 successful amplifications with the primer pair *Cob*_f1/*Cob*_r1.

A1.1.2. Amplification of the minicircle psbA^{ncr} (clade C samples)

Sixty-three samples were amplified using cycling conditions of denaturation for 3 min at 95ºC, followed by 38 cycles of 30 s at 95ºC, 30 s at 54ºC, and 30 s at 72ºC. Reactions consisted of 12.5 µL MyTaq HS Red Mix 2x (BioLine), ~20 ng sample DNA, 10 µg BSA, 0.4 µM of each primer and 7.5 µL PCRH$_2$O for a total reaction volume of 25µL.

The following methods all had reactions consisting of 10 µL MyTaq HS Red Mix 2x (BioLine), ~20 ng sample DNA, 10 µg BSA, 0.25 µM of each primer and 6 µL PCRH$_2$O for a total volume of 20 µL:

Six samples were amplified using cycling conditions of denaturation for 3 min at 95ºC, followed by 35 cycles of 20 s at 95ºC, 30 s at 57ºC, and 30 s at 72ºC.

Twelve samples were amplified using cycling conditions of denaturation for 3 min at 95ºC, followed by 37 cycles of 20 s at 95ºC, 30 s at 58ºC, 30 s at 72ºC, and a final extension for 7 min at 72ºC.

A total of 309 samples were amplified using the two sets of conditions outlined in the methods, leading to a total of 390 successful clade C amplifications with the primer pair 7.4-Forw/7.8-Rev.

A1.2. PCR clean-up process using MagNA solution

The purpose of this clean-up method is to match commercial kits in quality at a cost that is much more manageable for large numbers of samples (\$0.46 mL^{-1} *vs.* \$19 mL^{-1}). It also allows accurate size-selection of fragments during the clean-up process. The principle of using MagNA solution (0.1% carboxyl-modified Sera-Mag Magnetic Speed-beads [FisherSci, cat.#: 09-981-123] 18% w/v PEG-8000, 1 M NaCl, 10 mM Tris-HCl, 1 mM EDTA, 0.05% Tween 20, pH 8.0) and an initial protocol was developed by Rohland and Reich (2012). The preparation of the Speed-bead solution used in this study was done following Faircloth and Glenn (2014). The protocol used to clean up the reactions also follows Faircloth and Glenn (2014) and is copied below, but with several modifications that I made to enhance the process (indicated by italics).

1. Add 20 µL DNA template mixture to a volume of MagNA solution (the specific volume depends on amplicon size and PCR volume).

   For the cob *gene, I added 25 µL of MagNA to the 16 µL of PCR product. For the psbA^{ncr}, I added 30 µL to the 16 µL of PCR product. For the host ITS2 region, I added 35 µL to the 16 µL of PCR product. For the first clean-up of NGS ITS2 region I added 35 µL to the 25 µL of PCR product. For the second clean-up of NGS ITS2 region I added 30 µL to the 20 µL of PCR product.*

2. Incubate mixture 5 min at room temperature.

   It was noticed that the bead mixture and PCR product formed two distinct layers, so the mixture was vortexed and centrifuged for ~2 s each to homogenise the mixture.

3. Place on magnet stand.

4. Remove supernatant.

5. Add 500 µL 80% EtOH.

   70% EtOH was used as this is generally the recommended concentration for cleaning of DNA product, is the concentration recommended in the original Rohland and Reich (2012) protocol, and saves on reagent costs. Only 450 µL were used, as it was judged to be functionally the same and saved a considerable amount of ethanol given that over 1200 samples were cleaned up in this manner.

6. Incubate on stand for 1 min.

7. Remove supernatant.

8. Add 500 µL 80% EtOH.

   450 µL 70% EtOH were used (see step 5 above).

9. Incubate on stand for 1 min.

10. Remove supernatant.

11. Allow beads to sit until dry.

    Samples were dried using a heat block, as this sped up processing time. For the cob gene, psbA^{ncr} and host ITS2 region, samples were dried for ~7 min at 37ºC. For the two sets of clean-ups for the NGS of symbiont ITS2 sequences, samples were dried for ~5 min at 34ºC to avoid over-drying, as this can affect subsequent yields in NGS.

12. Rehydrate with 20 µL dH_2O.

    20 µL TE (10 mM Tris-HCl, 1 mM EDTA) was used instead, as it is better for longer-term storage of DNA. The sample was also left to sit for ~8 min to ensure all the DNA retained on the beads eluted into the TE.

13. Place on magnet stand.

14. Transfer supernatant to new tube.

    Following this step, samples were stored at 4ºC until all samples from a single gene had been cleaned up, at which point they were sent for sequencing.

A1.3. Amplification of clade D psbA^{ncr} sequences

Difficulties were encountered in amplifying clade D samples. Trial runs were conducted using the general primers psbAFor_1 (5'-GCA GCT CAT GGT TAT TTT GGT AGA C-3')/psbARev_1 (5'-AAT TCC CAT TCT CTA CCC ATC C-3'), which are non-specific primers designed to target the psbA minicircle of all clades (LaJeunesse and Thornhill 2011). Running the amplicons on a gel revealed highly non-specific products, with samples often having up to three bands (Figure A1.1). This could be due to the hyper-variable nature of the psbA^{ncr}, or the primers amplifying multiple minicircles, as symbiont chloroplast genomes are composed of many different minicircles (Barbrook *et al.* 2014). A novel set of more specific primers were trialled: SYMPSBANCRF (5'-AAT CGT GCT GAT CTA GGW ATG G-3')/SYMPSBANCRR (5'-GAG ACG ATT TGT TGT GGA TAG-3') (S. Wilkinson, unpublished). Trial runs revealed a greater degree of specificity, but multiple bands *per* sample were often still observed. In order to attempt to isolate single bands while still getting reasonable coverage, the approach described in the Methods was applied.
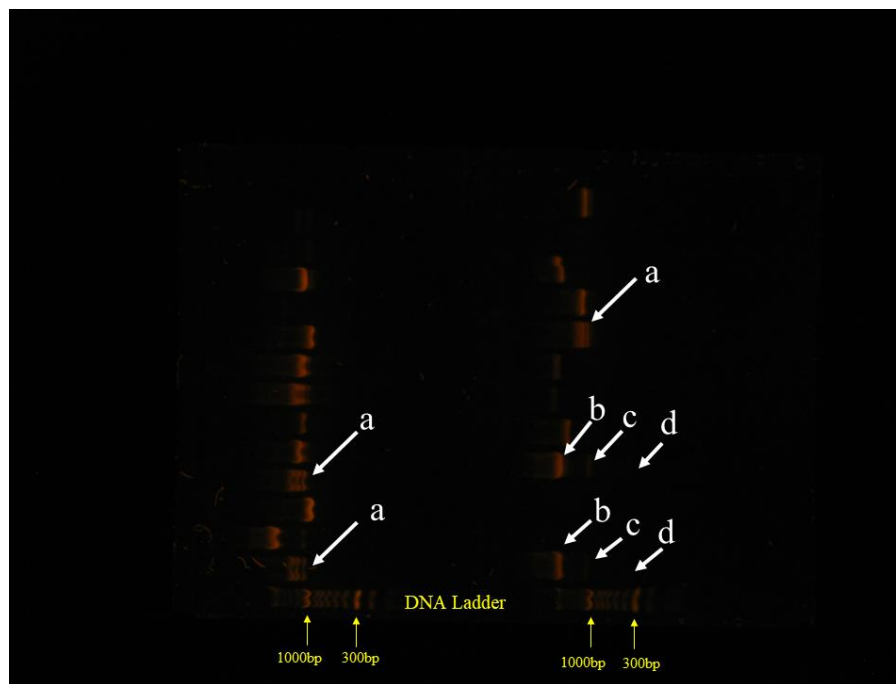
Figure A1.1: Example of a 1.5% agarose gel running clade D samples amplified with the primer pair psbAFor_1/psbARev_1, stained with SYBR-Safe DNA stain. Note two rows of samples are shown, each with a ladder in the first well. White arrows indicate typical banding patterns observed in a single sample, either three bands close together of comparable strength (a) or three bands separated by >500 bp, the first being strong (b) and the smaller fragments being weak (c, d).

A1.4. IDT primers for Next Generation Sequencing

Table A1.1: List of indexing primers used in Next Generation Sequencing of symbiont ITS2 region. The primers had the general sequence 5'- AAT GAT ACG GCG ACC ACC GAG ATC TAC ACX XXX XXX XTC GTC GGC AGC GTC-3' (forward) or 5'- CAA GCA GAA GAC GGC ATA CGA GAT XXX XXX XXG TCT CGT GGG CTC GG-3' (reverse), where the X's represent the specific index listed in the table.

Primer Name	Index Sequence (5'-3')
Forward	
S502	CTCTCTAT
S503	TATCCTCT
S505	GTAAGGAG
S506	ACTGCATA
S507	AAGGAGTA
S508	CTAAGCCT
S510	CGTCTAAT
S511	TCTCTCCG
S513	TCGACTAG
S515	TTCTAGCT
S516	CCTAGAGT
S517	GCGTAAGA
S518	CTATTAAG
S520	AAGGCTAT
S521	GAGCCTTA
S522	TTATGCGA
Reverse	
N701	TCGCCTTA
N702	CTAGTACG
N703	TTCTGCCT
N704	GCTCAGGA
N705	AGGAGTCC
N706	CATGCCTA
N707	GTAGAGAG
N710	CAGCCTCG
N711	TGCCTCTT
N712	TCCTCTAC
N714	TCATGAGC
N715	CCTGAGAT
N716	TAGCGAGT
N718	GTAGCTCC
N719	TACTACGC
N720	AGGCTCCG
N721	GCAGCGTA
N722	CTGCGCAT
N723	GAGCGCTA
N724	CGCTCAGT
N726	GTCTTAGG

Table A1.1 (cont.)

N727	`ACTGATCG`
N728	`TAGCTGCA`
N729	`GACGTCGA`

A1.5. Designation of all symbiont types found from sequencing that were the dominant type in at least one sample

A1.5.1. Designation of previously defined *cob* gene sequences

Table A1.2: Identification of *cob* sequences found in this study that match previously defined types listed in GenBank. Accession numbers are listed that have a 100% match to my sequence.

Name	Number found	GenBank Accession No.	Source paper and other notes
C3	209	FJ529535	Sampayo *et al.* 2009
C42a	72	FJ529541	Sampayo *et al.* 2009
C1/3	22	KP234522 AB971152 KF740692	Hume *et al.* 2015; Yorifuji *et al.* 2015; Jeong *et al.* 2014. Hume and colleagues ID as C3. Yorifuji and colleagues ID as C1. Jeong and colleagues ID as F (likely incorrect). This is despite their accession numbers giving identical sequences. To reflect ambiguity, has been named C1/3 here, though my results clearly place it as part of C3 radiation.
C1	7	FJ529534 JN557943 KF206028	Sampayo *et al.* 2009; Pochon *et al.* 2012; LaJeunesse and Parkinson (direct submission, 2013).
D1	46	KY131780 KF193520 JN557956	Wham *et al.* 2017; LaJeunesse *et al.* 2014; Pochon *et al.* 2012.

A1.5.2. Naming of *cob* gene sequences that did not match previously defined types:

Name: C15p (GenBank Accession Number: MH236749)
Number: 29
Sequence:
```
TACTATTTTATTACAAATTATATCTGGAATCTTCTTAGGTTTACATTATACATCAGATATTAATTCAGCATATTT
TAGTATTTTCTTTATTATTAGAGAAATATATTATGGATGGTGTTTACGTTATCTTCATTCTAATGGTTCATCATT
TGTCTTTCTTTTGATATTTCTACATCTTGGAAGAGCTATATCTTATGGTTCATATTTTTATAATCCAAATACTTG
GTTTTCTGGAATTATTATTATCTTCTTCTTAATGGGAACAGCATTTATGGGTTATGTGTTACCTTTAGGACAAAT
GAGTTTATGGGGGGGTTACAGTAATTACAAATTTATTATCTGCATTTCCATCTTTAATAGAATGGCTTTGTGGAGG
ACATTACATTTACAATCCTACATTTAAGAGGTTCTTTGTCTTTCATTTTCTATTTCCATTTCTTCTTTGTGGGTT
TCTTCTTTATCATATTTTTAATCTACATTTTCTATCTTCTAATAATCCTTTAAGGAATTCCACTAATAATAAAAT
AGCATTTTTCCCTTTCATTATTAGTAAAGATTTATATGGAAGATATTAATTCTCTATCTATATCTTCTTCAAAT
TCATTTCGGTTTCTCTTCTTTCTCACATCCAGATAATGCATTAGAAGCATGTGGATTACTTACTCCTTTACATAT
AGTACCTGAATGGTATTTCTTATGCCAATATGCTATGTTAAAAGCTGTGCCAAATAAAAATGCAGGATTCATTAT
ATTATTTACTTCTATCTTCATCTTATTTTACTTTATG
```

Characterisation and justification for name:
1 bp difference from type C1/3 (Table A1.2). Always in samples that matched ITS2 type C15. Letter 'p' chosen to avoid overlap with previously defined subtypes.

Name: C15q (GenBank Accession Number: MH236750)
Number: 13
Sequence:
```
TACTATTTTATTACAAATTATATCTGGAATCTTCTTAGGTTTACATTATACATCAGATATTAATTCAGCATATTT
TAGTATTTTCTTTATTATTAGAGAAATATATTATGGATGGTGTTTACGTTATCTTCATTCTAATGGTTCATCATT
TGTCTTTCTTTTGATATTTCTACATCTTGGAAGAGCTATATCTTATGGTTCATATTTTTATAATCCAAATACTTG
GTTTTCTGGAATTATTATTATCTTCTTCTTAATGGGAACAGCATTTATGGGTTATGTGTTACCTTTAGGACAAAT
GAGTTTATGGGGGGTTACAGTAATTACAAATTTATTATCTGCATTTCCATCTTTAATAGAATGGCTTTGTGGAGG
ACATTACATTTACAATCCTACATTTAAGAGGTTCTTTGTCTTTCATTTTCTATTTCCATTTCTTCTTTGTGGGTT
TCTTCTTTATCATATTTTTAATCTACATTTTCTATCTTCTAATAATCCTTTAAGGAATTCCACTAATAATAAAAT
AGCATTTTTCCCTTTCATTATTAGTAAAGATTTATATGGAAAGATATTAATTCTCTATCTATATCTTCTTCAAAT
TCATTTCGGTTTCTCTTCTTTCTCACATCCAGATAATGCATTAGAAGCATGTGGATTACTTACTCCTTTACATAT
AGTACCTGAATGGTATTTCTTATGCCAATATGCTATGTTAAAAGCTGTCCCAAATAAAAATGCAGGATTCATTAT
ATTATTTACTTCTATCTTCATCTTATTTTACTTTATG
```

Characterisation and justification for name:
2 bp difference from type C1/3 (Table A1.2), 1 bp different from C15p. Always in samples that matched ITS2 type C15. Letter 'q' chosen to avoid overlap with previously defined subtypes.

Name: C42p (GenBank Accession Number: MH236751)
Number: 9
Sequence:
```
TACTATTTTATTACAAATTATATCTGGAATCTTCTTAGGTTTACATTATACATCAGATATTAATTCAGCATATTT
TAGTATTTTCTTTATTATTAGAGAAATATATTATGGATGGTGTTTACGTTATCTTCATTCTAATGGTTCATCATT
TGTCTTTCTTTTGATATTTCTACATCTTGGAAGAGCTATATCTTATGGTTCATATTTTTATAATCCAAATACTTG
GTTTTCTGGAATTATTATTATCTTCTTCTTAATGGGAACAGCATTTATGGGTTATGTGTTACCTTTAGGACAAAT
GAGTTTATGGGGGAGTTACAGTAATTACAAATTTATTATCTGCATTTCCATCTTTAATAGAATGGCTTTGTGGAGG
ACATTACATTTACAATCCTACATTTAAGAGGTTCTTTGTCTTTCATTTTCTATTTCCATTTCTTCTTTGTGGTTT
TCTTCTTTATCATATTTTTAATCTACATTTTCTATCTTCTAATAATCCTTTAAGGAATCCCACTAATAATAAAAT
AGCATTTTTCCCTTTCATTATTAGTAAAGATTTATATGGAAAGATATTAATTCTCTATCTATATCTTCTTCAAAT
TCATTTCGGTTTCTCTTCTTTCTCACATCCAGATAATGCATTAGAAGCATGTGGATTACTTACTCCTTTACATAT
AGTACCTGAATGGTATTTCTTATGCCAATATGCTATGTTAAAAGCTGTACCAAATAAAAATGCAGGATTCATTAT
ATTATTTACTTCTATCTTCATCTTATTTTACTTTATG
```

Characterisation and justification for name:
1 bp different from type C42a (Table A1.2), no other close similarities. Letter 'p' chosen to avoid overlap with previously defined subtypes.

All other previously undefined sequences (13 clade C) only appeared once or twice and are not reproduced here. GenBank Accession Numbers: MH236752-MH236764

A1.5.3. Designation of previously defined ITS2 gene sequences

Table A1.3: Identification of ITS2 sequences found in this study that match previously defined types listed in GenBank. Accession numbers are listed that have a 100% match to my sequence; N/A indicates there was no exact match in GenBank and it was identified from the GeoSymbio database, where it had a 100% match (Franklin *et al.* 2012).

Name	Number found	GenBank Accession No.	Source paper and other notes
C40	130	N/A	Franklin *et al.* 2012
C15	40	JN558044	Pochon *et al.* 2012
C3d/C21	36	JN711493	Hume and Wiedenmann (unpublished)
C1	35	JN558041	Pochon *et al.* 2012
C3z	11	N/A	Franklin *et al.* 2012
C66	9	N/A	Franklin *et al.* 2012
C1d	5	KU197085	Kopp *et al.* 2016
C3	4	JN711498 EU828690	Hume and Wiedenmann (unpublished), Fay *et al.* 2009
C1#	2	JF298202	Franklin *et al.* 2012. This sequence was identified as being part of the C1 radiation by van Oppen *et al.* 2011b, hence its name. However, this appears to be an error – my Figure 2.4 groups it with the C3 radiation, and blasting the accession number they provide (JF298202) against C1 reveals that it is 6 bp different, not the 1 bp they claim in their paper.
C1b/C1e	2	N/A	Franklin *et al.* 2012
C35.2/C26.b1	1	N/A	Franklin *et al.* 2012
C1c/C45	1	N/A	Franklin *et al.* 2012
C42a	1	N/A	Franklin *et al.* 2012
C1j	1	KU197084	Kopp *et al.* 2016
C31	1	N/A	Franklin *et al.* 2012
C3u	1	N/A	Franklin *et al.* 2012
D1	32	N/A	Franklin *et al.* 2012
D5	4	N/A	Franklin *et al.* 2012
D6	1	N/A	Franklin *et al.* 2012
D3	1	N/A	Franklin *et al.* 2012

A1.5.4. Naming of ITS2 gene sequences that did not match previously defined types:

Name: C1x (GenBank Accession Number: MH236765)
Number: 28
Sequence:
```
GTGAATTGCAGAACTCCGTGAACCAATGGCCTCCTGAACGTGCGTTGCACTCTTGGGATTTCCTGAGAGTATGTC
TGCTTCAGTGCTTAACTTGCCCCCAACTTTGCAAGCAGGATGTGTTTCTGCCTTGCGTTCTTATGAGCTATTGCC
CTCTGAGCCAATGGCTTGTTAATTGCTTGGTTCTTGCAAAATGCTTTGCGCGCTGTTATTCAGGTTTCTACCTTC
GTGGTTTTACTTGAGTGACGCTGCTCATGCTTGCAACCGCTGGGATGCAGGTGCATGCCTCTAGCATGAAGTCAG
ACAAGTGAACCCGCTGAATTTAAGCATATAAGTAAGCGGAGG
```

Characterisation and justification for name:
1 bp different to C1 (Table A1.3), no other close similarities.

Name: C1z (GenBank Accession Number: MH236766)
Number: 9
Sequence:
GTGAATTGCAGAACTCCGTGAACCAATGGTCTCCTGAACGTGCGTTGCACTCTTGGGTTTCCTGAGAGTATGTCT
GCTTCAGTGCTTAACTTGCCCCAACTTTGCAAGCAGGATGTGTTTCTGCCTTGCGTTCTTATGAGCTATTGCCCT
CTGAGCCAATGGCTTGTTAATTGCTTGGTTCTTGCAAAATGCTTTGCGCGCTGTTATTCAGGTTTCTACCTTCGT
GGTTTTACTTGAGTGACGCTGCTCATGCTTGCAACCGCTGGGATGCAGGTGCATGCCTCTAGCATGAAGTCAGAC
AAGTGAACCCGCTGAATTTAAGCATATAAGTAAGCGGAGG

Characterisation and justification for name:
2 bp different to C1 (Table A1.3), 1 bp different to C1x, no other close similarities.

All other previously undefined sequences (13 clade C, 5 clade D) only appeared once or twice and are not reproduced here. Clade C Genbank Accession Numbers: MH236767-MH236779. Clade D GenBank Accession Numbers: MH236780-MH236784.

A1.5.5: Naming of psbA^{ncr} 95% groupings, based on consensus sequences.

This appendix will list all significant groups found for the psbA^{ncr} region, as represented by their consensus sequences. Because each one represents a group of closely-related sequences, there is often more than one GenBank Accession Number for a single group.

Name: C40n (GenBank Accession Number: MH329431)
Number: 10 (all identical)
Sequence:
CCCATATCCCCGAAGGGGATTAGAGGGCCCGCAGGAGCCCGCAAAGAAAAATCACGAAGTAGAACCGAGAAAAGG
CTGCCGCTAATTTTGGCCTAGATCACGATAATTTTGGCCAAAATTGGCCTTCGGGAATTCGCACACGCGCGTATA
AAAGATCGCGACTTATAATGCACCACCCTTCGGGTGGTGCATATGGGTGCCCTTTAGGGCACCCATTATAATGGG
TACCCCGTAGGGGTACCCATATGCACCCCTACGGGGTGCATTGGAGCCGGGAACGAGCCGAAGGCGAGTGGACGT
GCGACCACGAAAAGAAAAGAAATCGCGACCTATAATGGGTGCCCCGTAGGGGCACCCATATGCATGCCCCAAAG
GGGCATGCATTAATGGGCTGGGTGCCCTACCCAGCCCATATGGGCCCACGCTTCGCGGGGCCCATAACGCCCTTC
GGGCTTCAAAAATA

Characterisation and justification for name:
Always occurred in samples with ITS2 type C40. Also, 1 bp different to psbA^{ncr} sequence C40 (KF572359, Thornhill *et al.* 2014).

Name: C40o (GenBank Accession Numbers: MH329466-MH329472)
Number: 11
Consensus sequence:
CCCATATCCCCGAAGGGGATTAGAGGGCCCGCAGGAGCCCGCAAAGAAAAATCACGAAGTAGAACCGAGAAAAGG
CTGCCGCTAATTTTGGCCTAGATCACGATAATTTTGGCCAAAATTGGCCTTCGGGAATTCGCACACGCGCGTATA
AAAGATCGCGACTTATAATGCACCACCCTTCGGGTGGTGCATATGGGTGCCCTTTAGGGCACCCATTATAATGGG
TACCCCGTAGGGGTACCCATATGCACCCCTACGGGGTGCATTGGAGCCGGGAACGAGCCGAAGGCGAGTGGACGT
GCGACCACGAAAAGAAAAGAAATCGCGACCTATAATGGGTGCCCCGTAGGGGCACCCATATGCATGCCCCAAAG
GGGCATGCATTAATGGGCTGGGTGCCCTACCCAGCCCATATGGGCCCACGCTTCGCGGGGCCCATAACGCCCTTC
GGGCTTCAAAAATA

Characterisation and justification for name:
Always occurred in samples with ITS2 type C40. Also, 3 bp different to psbA^{ncr} sequence C40 (KF572359, Thornhill *et al.* 2014).

Name: C40c (GenBank Accession Numbers: MH329473-MH329486)

Number: 21

Consensus sequence:

CCCATATCCCCGAAGGGGATTAGAGGGCCCGCAGGAGCCCGCAAAGAAAAATCACGAAGTAGAACCGAGAAAAGG
CTGCCGCTAATTTTGGCCTAGATCACGATAATTTTGGCCAAAATTGGCCTTCGGGAATTCGCACACGCGCGTATA
AAAGATCGCGACTTATAATGCACCACCCGAAGGGTGGTGCATATGGGTGCCCTTTAGGGCACCCATTATAATGGG
TACCCCGTAGGGGTACCCATATGCACCCCTATGGGGTGCATTGGAGCCGGGAACGAGCCGAAGGCGAGTGGACGT
GCGACCATGAAAAGAAAAAGAAATCGCGACCTATAATGGGTGCCCCTACGGGGCACCCATATGCATGCCCCAAAG
GGGCATGCATTAATGGGCTGGGTGCCCTACCCAGCCCATATGGGCCCACGCTTCGCGGGGCCCATAACGCCCTTC
GGGCTTCAAAAAT

Characterisation and justification for name:

Always occurred in samples with ITS2 type C40. Also, 2 bp different to psbA[ncr] sequence C40 (KF572365, Thornhill *et al.* 2014).

Name: C40d (GenBank Accession Number: MH329487)

Number: 14 (all identical)

Sequence:

CCCATATCCCCGAAGGGGATTAGAGGGCCCGCAGGAGCCCGCAAAGAAAAATCACGAAGTAGAACCGAGAAAAGG
CTGCCGCTAATTTTGGCCTAGATCACGATAATTTTGGCCAAAATTGGCCTTCGGGAATTCGCACACGCGCGTATA
AAAGATCGCGACTTATAATGCACCACCCTTCGGGTGGTGCATATGGGTGCCCTTTAGGGCACCCATTATAATGGG
TACCCCGTAGGGGTACCCATATGCACCCTATGGGGTGCATTGGAGCCGGGAACGAGCCGAAGGCGAGTGGACGTG
CGACCATGAAAAGAAAAAGAAATCGCGACCTATAATGGGTGCCCCGTAGGGGCACCCATATGCATGCCCCAAAGG
GGCATGCATTAATGGGCTGGGTGCCCTACCCAGCCCATATGGGCCCACGCTTCGCGGGGCCCATAACGCCCTTCG
GGCTTCAAAAATA

Characterisation and justification for name:

Always occurred in samples with ITS2 type C40. Also, 2 bp different to psbA[ncr] sequence C40 (KF572359, Thornhill *et al.* 2014).

Name: C40e (GenBank Accession Number: MH329489)

Number: 7 (all identical)

Sequence:

CCCATATCCCCGAAGGGGATTAGAGGGCCCGCAGGAGCCCGCAAAGAAAAATCACGAAGTAGAACCGAGAAAAGG
CTGCCGCTAATTTTGGCCTAGATCACGATAATTTTGGCCAAAATTGGCCTTCGGGAATTCGCACACGCGCGTATA
AAAGATCGCGACTTATAATGCACCACCCTTCGGGTGGTGCATATGGGTGCCCTTTAGGGCACCCATTATAATGGG
TACCCCTACGGGGTACCCATATGCACCCCTATGGGGTGCATTGGAGCCGGGAACGAGCCGAAGGCGAGTGGACGT
GCGACCATGAAAAGAAAAAGAAATCGCGACCTATAATGGGTGCCCCTACGGGGCACCCATATGCATGCCCCAAAG
GGCATGCATTAATGGGCTGGGTGCCCTACCCAGCCCATATGGGCCCACGCTTCGCGGGGCCCATAACGCCCTTCG
GGCTTCAAAAATA

Characterisation and justification for name:

Always occurred in samples with ITS2 type C40. Also, 5 bp different to psbA[ncr] sequence C40 (KF572366, Thornhill *et al.* 2014).

Name: C40f (GenBank Accession Numbers: MH329490-MH329496)

Number: 8

Consensus sequence:

CCCATATCCCCGAAGGGGATTAGAGGGCCCGCAGGAGCCCGCAAAGAAAAATAACGAAGTAGAACCGAGAAAAGG
CTGCCGCTAATTTTGGCCTAGATCACGATAATTTTGGCCTAAATTGGCCTTCGGGAATTCGCACACGCGCGTATA
AAAGATCGCGACTTATAATGCACCACCCTTCGGGTGGTGCATATGGGTGCCCTTTAGGGCACCCATTATAATGGG
TACCCCGTAGGGGTACCCATATGCACCCCGTAGGGGTGCATTGGAGCCGGGAACGAGCCGAAGGCGAGTGGACGT
GCGACCACGAAAAGAAAAAGAAATCGCGACCTATAATGGGTGCCCCGTAGGGGCACCCATATGCATGCCCCTGT

GGGGCATGCATTAATGGGGGTGGGTGCCCTACCCACCCCATATGGGCCCACACTTCGTGGGGCCCATAACGCCCT
TCGGGCTTCAAAAAATA

Characterisation and justification for name:
Always occurred in samples with ITS2 type C40. Also, 5 bp different to psbA^{ncr} sequence C40 (KF572366, Thornhill *et al.* 2014).

Name: C40g (GenBank Accession Numbers: MH329497-MH329505)
Number: 9
Consensus sequence:
CCCATATCCCCGAAGGGGATTAGAGGGCCCGCAGGAGCCCGCAAAGAAAAATCACGAAGTAGAACCGAGAAAAGG
CTGCCGCTAATTTTGGCCTAGATCACGATAATTTTGGCCAAAATTGGCCTTCGGGAATTCGCACACGCGCGTATA
AAAGATCGCGACTTATAATGCACCACCCTTCGGGTGGTGCATATGGGTGCCCTTTAGGGCACCCATTATAATGGG
TACCCCTACGGGGTACCCATATGCACCCCTATGGGGTGCATTGGAGCCGGGAACGAGCCGAAGGCGAGTGGACGT
GCGACCATGAAAAGAAAAAGAAATCGCGACCTATAATGGGTGCCCCTACGGGGCACCCATATGCATGCCCCAAAG
GGGCATGCATTAATGGGCTGGGTGCCCTACCCAGCCCATATGGGCCCACGCTTCGCGGGGCCCATAACGCCCTTC
GGGCTTCAAAAATA

Characterisation and justification for name:
Always occurred in samples with ITS2 type C40. Also, 5 bp different to psbA^{ncr} sequence C40 (KF572366, Thornhill *et al.* 2014).

Name: C40h (GenBank Accession Numbers: MH329516-MH329521)
Number: 15
Consensus sequence:
CCCATATCCCCGAAGGGGATTAGAGGGCCCGCAGGAGCCCGCAAAGAAAAATCACGAAGTAGAACCGAGAAAAGG
CTGCCGCTAATTTTGGCCTAGATCACGATAATTTTGGCCAAAATTGGCCTTCGGGAATTCGCACACGCGCGTATA
AAAGATCGCGACTTATAATGCACCACCCTTCGGGTGGTGCATATGGGTGCCCTTTAGGGCACCCATTATAATGGG
TACCCCGTAGGGGTACCCATATGCACCCTATGGGGTGCATTGGAGCCGGGAACGAGCCGAAGGCGAGTGGACGTG
CGACCATGAAAAGAAAAAGAAATCGCGACCTATAATGGGTGCCCCGTAGGGGCACCCATATGCATGCCCCAAAGG
GGCATGCATTAATGGGCTGGGTGCCCTACCCAGCCCATATGGGCCCACGCTTCGCGGGGCCCATAACGCCCTTCG
GGCTTCAAAAATA

Characterisation and justification for name:
Always occurred in samples with ITS2 type C40. Also, 2 bp different to psbA^{ncr} sequence C40 (KF572359, Thornhill *et al.* 2014).

Name: C40i (GenBank Accession Numbers: MH329522-MH329526)
Number: 7
Consensus sequence:
CCCATATCCCCGAAGGGGATTAGAGGGCCCGCAGGAGCCCGCAAAGAAAAATCACGAAGTAGAACCGAGAAAAGG
CTGCCGCTAATTTTGGCCTAGATCACGATAATTTTGGCCAAAATTGGCCTTCGGGAATTCGCACACGCGCGTATA
AAAGATCGCGACTTATAATGCACCACCCTTCGGGTGGTGCATATGGGTGCCCTTTAGGGCACCCATTATAATGGG
TACCCCGTAGGGGTACCCATATGCACCCTATGGGGTGCATTGGAGCCGGGAACGAGCCGAAGGCGAGTGGACGTG
CGACCATGAAAAGAAAAAGAAATCGCGACCTATAATGGGTGCCCCTACGGGGCACCCATATGCATGCCCCAAAGG
GGCATGCATTAATGGGCTGGGTGCCCTACCCAGCCCATATGGGCCCACGCTTCGCGGGGCCCATAACGCCCTTCG
GGCTTCAAAAATA

Characterisation and justification for name:
Always occurred in samples with ITS2 type C40. Also, 3 bp different to psbA^{ncr} sequence C40 (KF572370, Thornhill *et al.* 2014).

Name: C40j (GenBank Accession Numbers: MH329527-MH329533)

Number: 9

Consensus sequence:

CCCATATCCCCGAAGGGGATTAGAGGGCCCGCAGGAGCCCGCAAAGAAAAATCACGAAGTAGAACCGAGAAAAGG
CTGCCGCTAATTTTGGCCTAGATCACGATAATTTTGGCCAAAATTGGCCTTCGGGAATTCGCACACGCGCGTATA
AAAGATCGCGACTTATAATGCACCACCCTTCGGGTGGTGCATATGGGTGCCCTTTAGGGCACCCATTATAATGGG
TACCCCGTAGGGGTACCCATATGCACCCCGTAGGGGTGCATTGGAGCCGGGAACGAGCCGAAGGCGAGTGGACGT
GCGACCACGAAAAGAAAAAGAAATCGCGACCTATAATGGGTGCCCCTACGGGGCACCCATATGCATGCCCCAAAG
GGGCATGCATTAATGGGCTGGGTGCCCTACCCAGCCCATATGGGCCCACGCTTCGCGGGGCCCATAACGCCCTTC
GGGCTTCAAAATA

Characterisation and justification for name:

Always occurred in samples with ITS2 type C40. Also, 2 bp different to psbA^{ncr} sequence C40 (KF572366/KF572364, Thornhill *et al.* 2014).

Name: C40k (GenBank Accession Numbers: MH329534-MH329535)

Number: 5

Consensus sequence:

CCCATATCCCCGAAGGGGATTAGAGGGCCCGCAGGAGCCCGCAAAGAAAAATCACGAAGTAGAACCGAGAAAAGG
CTGCCGCTAATTTTGGCCTAGATCACGATAATTTTGGCCAAAATTGGCCTTCGGGAATTCGCACACGCGCGTATA
AAAGATCGCGACTTATAATGCACCACCCTTCGGGTGGTGCATATGGGTGCCCTTTAGGGCACCCATTATAATGGG
TACCCCGTAGGGGTACCCATATGCACCCCTATGGGGTGCATTGGAGCCGGGAACGAGCCGAAGGCGAGTGGACGT
GCGACCATGAAAAGAAAAAGAAATCGCGACCTATAATGGGTGCCCCTACGGGGCACCCATATGCATGCCCCAAAG
GGGCATGCATTAATGGGCTGGGTGCCCTACCCAGCCCATATGGGCCCACGCTTCGCGGGGCCCATAACGCCCTTC
GGGCTTCAAAATA

Characterisation and justification for name:

Always occurred in samples with ITS2 type C40. Also, 3 bp different to psbA^{ncr} sequence C40 (KF572366/KF572364, Thornhill *et al.* 2014).

Name: C40l (GenBank Accession Numbers: MH329536-MH329540)

Number: 5

Consensus sequence:

CCCATATCCCCGAAGGGGATTAGAGGGCCCGCAGGAGCCCGCAAAGAAAAATCACGAAGTAGAACCGAGAAAAGG
CTGCCGCTAATTTTGGCCTAGATCACGATAATTTTGGCCAAAATTGGCCTTCGGGAATTCGCACACGCGCGTATA
AAAGATCGCGACTTATAATGCACCACCCTTCGGGTGGTGCATATGGGTGCCCTTTAGGGCACCCATTATAATGGG
TACCCCGTAGGGGTACCCATATGCACCCTACGGGTGCATTGGAGCCGGGAACGAGCCGAAGGCGAGTGGACGTGC
GACCACGAAAAGAAAAAGAAATCGCGACCTATAATGGGTGCCCCTACGGGGCACCCATATGCATGCCCCAAAGGG
GCATGCATTAATGGGCTGGGTGCCCTACCCAGCCCATATGGGCCCACGCTTCGCGGGGCCCATAACGCCCTTCGG
GCTTCAAAATA

Characterisation and justification for name:

Always occurred in samples with ITS2 type C40. Also, 2 bp different to psbA^{ncr} sequence C40 (KF572370, Thornhill *et al.* 2014).

Name: C40m (GenBank Accession Numbers: MH329506-MH329515)

Number: 12

Consensus sequence:

CCCATATCCCCGAAGGGGATTAGAGGGCCCGCAGGAGCCCGCAAAGAAAAATCACGAAGTAGAACCGAGAAAAGG
CTGCCGCTAATTTTGGCCTAGATCACGATAATTTTGGCCAAAATTGGCCTTCGGGAATTCGCACACGCGCGTATA
AAAGATCGCGACTTATAATGCACCACCCTTCGGGTGGTGCATATGGGTGCCCTTTAGGGCACCCATTATAATGGG
TACCCCGTAGGGGTACCCATATGCACCCCTACGGGGTGCATTGGAGCCGGGAACGAGCCGAAGGCGAGTGGACGT
GCGACCACGAAAAGAAAAAGAAATCGCGACCTATAATGGGTGCCCCGTAGGGGCACCCATATGCATGCCCCAAAG

```
GGGCATGCATTAATGGGCTGGGTGCCCTACCCAGCCCATATGGGCCCACGCTTCGCGGGGCCCATAACGCCCTTC
GGGCTTCAAAAATA
```

Characterisation and justification for name:
Always occurred in samples with ITS2 type C40. Also, 1 bp different to psbA[ncr] sequence C40 (KF572359, Thornhill *et al.* 2014).

Name: C3z (GenBank Accession Numbers: MH329432-MH329438)
Number: 7
Consensus sequence:
```
CCCATATGCACGAAGTGCATTAGAGGGCCCGCAGGAGCCCGCAAAGAAAAATCACGAAGTAGAACCGAGAAAAGG
CTGCCGCTAATTTTGGCCTAGATCACGATAATTTTGGCCAAAATTGCCCTTCGGGCATTCGCACACGCGCGTATA
AAAGATTGCGACTTATAATGCACCACCCTTCGGGTGGTGCATATGGGTGCCCTTTAGGGCACCCATTAGAGCCGG
GAACGAGCCGAAGGCGAGTGGACGTGCGACCACGAAAAAGAAAAATAAATCGCGACCTATAATGGGTGCCCCTAC
GGGGCACCCATATGCATGCCCCACAGGGGCATGCATTAATGGGGTGGGTGCCCTACCCACCCCATATGGGCCCAC
ACTTCGTGGGGCCCATAACGCCCTTCGGGCTTCAAAAAATA
```

Characterisation and justification for name:
Almost exclusively occurred with ITS2 type C3z (See Chapter 3 for exception).

Name: C1d (GenBank Accession Numbers: MH329439-MH329442)
Number: 5
Consensus sequence:
```
CCCATATGCCCTTCGGGCATTTAGGGCCCGCAGGAGCCCGCAGAAAAGAAAAGACCCTGCCGCTAATTTTGGCCT
AAATCGCGAGTAAAATGGCCTAAATTGCCCCTTCGGGGCTTCACGTACACGCGTACACGCGCACACGCGTTATAT
AATGTGCACCCCTTCGGGGTGCACATATGCACCACCCTTTGGGTGGTGCATTGGAGCCGGTAACGAGCCGAAGGC
GAGTGGACGTGCGACCACGAATAAAGAAAAATAAATCGCGACCTATAATGGGTGCCCCTGCGGGGCACCCATAT
GCACACCCCGAAGGGGTGTGCATTAATGGGCTGGGTGCCCTACCCAGCCCATATGGGCCCACGCTTCGCGGGGCC
CATAACGGCCCTTCGGGCCTTCAAAAAATA
```

Characterisation and justification for name:
Exclusively occurred with ITS2 type C1d. Also, 1 bp different from psbA[ncr] type C1d (HQ336236, Pinzon and LaJeunesse 2011).

Name: C1x (GenBank Accession Numbers: MH329443-MH329454)
Number: 17
Consensus sequence:
```
CCCATATGCCCGCAGGGCATTGAGGTCCACGAAGTGTGACCGTTAATTTTGGCCAAAAAAAGGGTGCCGCACAAA
ATGGCCTAAACTGCGACTATTTTGGCCAAAATTGCGACATGTGCCTCGCGCACACGCACACGCGTTATATAATGT
GCACCCCTTCGGGGTGCACATATGCACCACCCTTTGGGTGGTGCATTGGAGCCGGGAACGAGCCGAAGGCGAGTG
GACGTGCGACCACGAAGAAAGAAAATAAATCGCGACCTATAATGGGTGCCCCGTAGGGGCACCCATATGCACAC
CCCGAAGGGGTGTGCATTAATGGGCTGGGTGCCCTACCCAGCCCATATGGGCCCACGCTTCGCGGGGCCCATAAC
GGCCCTTCGGGCCTTCAAAAAATA
```

Characterisation and justification for name:
Occurred with ITS2 type C1 and C1x. Closest related match in GenBank was C1.

Name: C1u (GenBank Accession Numbers: MH329462-MH329463)
Number: 2
Consensus sequence:
```
CCCATATGCCCGCAGGGCATTGAGGTCCACGAAGTGTGACCGTTAATTTTGGCCAAAAAAAGGGTGCCGCTAATT
TTGGCCTAAACTGCGACTATTTTGGCCAAAATTGCGACATGTGTCCTCGCGCACACGCGCACACGCGCACACGTA
TTAAAAAGATCGCGACTTATAATGTGACGTAAATGGGTGCCCCTGTGGGGCACCCATATGCACCACCCTTTGGGT
```

GGTGCATTGGAGCCGGGAACGAGCCGAAGGCGAGTGGACGTGCGACCACGAAGAAAAGAAAAAGAAATCGCGACC
TATAATGGGTGCCCCTTTGGGGCACCCATATGCACACCCCGAAGGGGTGTGCATTAATGGGCTGGGTGCCCTACC
CAGCCCATATGGGCCCACGCTTCGCGGGGCCCATAACGGCCCTTCGGGCCTTCAAAAAATA

Characterisation and justification for name:
Occurred with ITS2 type C1. 100% match to C1 (KF572172, Thornhill *et al.* 2014), though the matched area missed 7% of my sequence.

Name: C1v (GenBank Accession Number: MH329488)
Number: 11 (all identical)
Sequence:
CCCATATGCCCGCAGGGCATTGAGGTCCACGAAGTGTGACCGTTAATTTTGGCCAAAAAAAGGGTGCCGCACAAA
ATGGCCTAAACTGCGACTATTTTGGCCAAAATTGCGACATGTGCCTCGCGCACACGCACACGCGTTATATAATGT
GCACCCCTTCGGGGTGCACATATGCACCACCCTTTGGGTGGTGCATTGGAGCCGGGAACGAGCCGAAGGCGAGTG
GACGTGCGACCACGAAGAAAAGAAAATAAATCGCGACCTATAATGGGTGCCCCGTAGGGGCACCCATATGCACAC
CCCGAAGGGGTGTGCATTAATGGGCTGGGTGCCCTACCCAGCCCATATGGGCCCACGCTTCGCGGGGCCCATAAC
GGCCCTTCGGGCCTTCAAAAAATA

Characterisation and justification for name:
Occurred with ITS2 type C1 and C1x. Closest related match in GenBank was C1.

Name: C3d/C21p (GenBank Accession Numbers: MH329455-MH329457)
Number: 3
Consensus sequence:
CCCATATGCACGAAGTGCATTAGAGGGCCCGGAGGAGCCCGCAAAGAAAAATCACGAAGTGGAACCGAGAAAAGG
CTGCCGCTAATTTTGGCCTAGATCGCGATAATTTTGGCCAAAATTGCCCTTTAGGGCATTCGCACACGCGCGTAT
AAAAGATCGCGACTTATAATGGGTACCCCTTAGGGGTACCCATATGCACCCCGTAGGGGTGCATTATAGGAGCCG
GGAACGAGCCGAAGGCGAGTGGACGAGCGACAAAGAAAAAGAAATCGCGACCTATAATGGGTGCCCCGTAGGG
GCACCCATATGCACACCCCGGAGGGGTGTGCATTAATGGGCTGGGTGCCCTACCCAGCCCATATGGGCCCACACT
TCGTGGGGCCCATAACGGAGCCCTTCGGGCTCCTTCAAAAAATA

Characterisation and justification for name:
Occurred exclusively with ITS2 type C3d/C21.

Name: C3d/C21q (GenBank Accession Numbers: MH329458-MH329461)
Number: 4
Consensus sequence:
CCCATATGCACGAAGTGCATTAGAGGGCCCGCAGGAGCCCGCAAGAAAAATCACGAAGTGGAACCGAGAAAAGGC
TGCCGCTAATTTTGGCCTAGATCACGATAATTTTGGCCAAAATTGTCCTTTAGAGCATTCGCACACGCGCGTATA
AAAGATCGCGACTTATAATGGGTACCCCTACGGGGTACCCATATGCACCCCGTAGGGGTGCATTATAGGAGCCGG
GAACGAGCCGAAGGCGAGTGGACGTGCGACCACGATAAAGAAAAATAAATCGCGACCTATAATGGGTGCCCCTAC
GGGGCACCCATATGCATGCCCCAAAGGGGCATGCATTAATGGGCTGGGTGCCCTACCCAGCCCATATGGGCCCAC
ACTTCGTGGGGCCCATAACGCCCTTCGGGCTTCAAAAAATA

Characterisation and justification for name:
Occurred exclusively with ITS2 type C3d/C21. Also very similar to C3d/C21p.

Name: Cva (GenBank Accession Numbers: MH329464-MH329465)
Number: 2
Consensus sequence:
CCCATATGCCCGCAGGGCATTGAGGTCCACGAAGTGTGACCGTTAATTTTGGCCAAAAAAAAGGGTGCCGCACAA
AATGGCCTAAACTGCGACTATTTTGGCCAAAATTGCGACATGTGCCTCGCGCACACGCGCACACGTATTAAAAAG
ATCGTGACTTATAAGGTGCACCCCTTCGGGGTGCACATATGGGCACCCTTCGGGTGCCCATTGGAGCCGGGAACG

124

AGCCGAAGGCGAGTGGACGTGCGACCACGAAAAAGAAAAATAAATCGCGACCTATAATGGGGTGCCCCTGCGGGG
CACCCATATGCACACCCCGAAGGGGTGTGCATTAATGGGCTGGGTGCCCTACCCAGCCCATATGGGCCCACGCTT
CGCGGGGCCCATAACGGCCCTTCGGGCCTTCAAAAATA

Characterisation and justification for name:
Occurred in different ITS2 types, hence has been given the designation 'v' (variable). No clear similarities were found in GenBank.

Name: Cvb (GenBank Accession Numbers: MH329541-MH329542)
Number: 2
Consensus sequence:
CCCATATGCCCGCAGGGCATTGAGGTCCACGAAGTGTGACCGTTAATTTTGGCCAAAAAAAGGGTGCCGGTAATT
TTGGCCTAAATCGCGAGTCTTCTGGCCTAAATTGCCCCTTCGGGGCTTCACGTACACGCGCACACGCGATAATGT
GCACCCCTTCGGGGTGCACATATGGGTGCCCTAAAGGGCACCCATTAGAGCCGGGAGCCCGAAGGGCGGACGAGC
GACAAAGAAAAAATAAATCGCGACCTATAATGGGTGCCCCTGCGGGGCACCCATATGCACACCCCGAAGGGGT
GTGCATTAATGGGCTGGGTGCCCTACCCAGCCCATATGGGCCCACGCTTCGCGGGGCCCATAACGGCCCTTCGGG
CCTTCAAAAAATA

Characterisation and justification for name:
Occurred in different ITS2 types, hence has been given the designation 'v' (variable). No clear similarities were found in GenBank.

For the clade D sequences, (x) indicates a deletion of x base pairs from the overall alignment.

Name: D1p (GenBank Accession Number: MH329568)
Number: 4 (all identical)
Sequence:
CACTGCTTTGGAGAGGTTGGACAGGTCTAATTATTAATAAATTTATTAATTAATTTATAGGATTGTATAGCCTCT
TGCTCCTCTTGCACCACTTTGGAGCAAGAGTAACATCTATAGC(4)TGCACCACTCCAACCTAAAATATATGTGT
CCTACAACTACAAATAAATAAACTAAAATATATTTTAGTTGGAGCAAGAGTAAAATGATTTACAGCTATACATCT
TACTGTTGCTCCTT(19)TGGGCTACAGGTAGAAACCTTCTACACACTAAACTATAACAGATCTTAAATAGTTTT
GGTATTCATTCTGTCAAATCAAAGAATTAATTAAGGTACACTTCTAGTTAATATGGCTTGTAGTATAAGATGACT
GATTTGATTTATGAGAAAGAGGTTTTAAAGCTGTGAGTTATAGCAGCTTGTCTTACTGTACATTATGAACAGGCT
TTTCACTTGTAAATAGCTGCTTCTATAGCTGCTATTTCTGGAGAATCAATAGCAATTTAAACCTCGTGATGGAAG
GACTGGTGAGATTGTTTGGAAGGGCTGGTGATAATTTATTTGTAGGTAGTACAGTGTACTACTGTACAGTAGTAC
TGTGCACTTAGGATTGTACTAGGGCCTTTCTAGTTGGCCAACTA

Characterisation and justification for name:
Occurred entirely with ITS2 type D1.

Name: D1q (GenBank Accession Number: MH329567)
Number: 4 (all identical)
Sequence:
CACTGCTTTGGAGAGGTTGGACAGGTCTAATTATTAATAAATTTATT(3)TATTTTAG(5)TTTAAAGCT(4)GC
TCCACCTCCA(2)ACT(10)AGTAAGATGTATAGCCTCTTGCACCACT(5)TTGGAGAAGATGGG(32)ATTTTA
GTTGGAGCAAGAGTAAAAT(177)GATGACTGATTTGATTTATGAGAAAGAGGTTTTAAAGCTGTGAGTTATAGC
AGCTTGTCTTACTGTACATTATGAACAGGCTTTTCACTTGTAAATAGCTGCTTCTATAGCTGCTATTTCTGGAGA
ATCAATAGCAATTTAAACCTCGTGATGGAAGGACTGGTGAGATTGTTTGGAAGGGCTGGTGATAATTTATTTGTA
GGTAGTACAGTGTACTACTGTACAGTAGTACTGTGCACTTAGGATTGTACTAGGGCCTTTCTAGTTGGCCAACTA
G

Characterisation and justification for name:
Occured entirely with ITS2 type D1.

Name: Dva (GenBank Accession Number: MH329569)

Number: 2 (identical)

Sequence:

```
CACTGCTTTGGAGAGGTTGGACAGGTCTAATTATTAATAAATTTATT(3)TATTTTAG(5)TTTAAAGCT(4)GC
TCCACCTCCA(2)ACT(10)AGTAAGATGTATAGCCTCTTGCACCACT(5)TTGGAGAAGATGGG(32)ATTTTA
GTTGGAGCAAGAGTAAAATGATTTACAGCTATACATCTTACTGTTGCTCCTTTGGGCAGCTAAGCTGGCCATGGG
CTACAGGTAGAAACCTTCTACACACTAAACTATAACAGATCTTAAATAGTTTTGGTATTCATTCTGTCAAATCAA
AGAATTAATTAAGGTACACTTCTAGTTAATATGGCTTGTAGTATAAGATGACTGATTTGATTTATGAGAAAGAGG
TTTTAAAGCTGTGAGTTATAGCAGCTTGTCTTACTGTACATTATGAACAGGCTTTTCACTTGTAAATAGCTGCTT
CTATAGCTGCTATTTCTGGAGAATCAATAGCAATTTAAACCTCGTGATGGAAGGACTGGTGAGATTGTTTGGAAG
GGCTGGTGATAATTTATTTGTAGGTAGTACAGTGTACTACTGTACAGTAGTACTGTGCACTTAGGATTGTACTAG
GGCCTTTCTAGTTGGCCAACTAG
```

Characterisation and justification for name:

Occurred with variable unique ITS2 clade D sequences, so has been given the designation 'v' (variable).

Name: Dvb (GenBank Accession Number: MH329570)

Number: 2 (identical)

Sequence:

```
CACTGCTTT(402)GGTTTTAAAGCTGTGAGTTATAGCAGCTTGTCTTACTGTACATTATGAACAGGCTTTTCAC
TTGTAAATAGCTGCTTCTATAGCTGCTATTTCTGGAGAATCAATAGCAATTTAAACCTCGTGATGGAAGGACTGG
TGAGATTGTTTGGAAGGGCTGGTGATAATTTATTTGTAGGTAGTACAGTGTACTACTGTACAGTAGTACTGTGCA
CTTAGGATTGTACTAGGGCCTTTCTAGTTGGCCAACTA
```

Characterisation and justification for name:

Occurred with variable unique ITS2 clade D sequences, so has been given the designation 'v' (variable).

Name: Dvc (GenBank Accession Number: MH329571)

Number: 2 (identical)

Sequence:

```
CACTGCTTTGGAGAGGTTGGACAGGTCTAATTATTAATAAATTTATT(3)TATTTTAG(5)TTTAAAGCT(4)GC
TCCACCTCCA(2)ACT(10)AGTAAGATGTATAGCCTCTTGCACCACT(5)TTGGAGAAGATGGG(32)ATTTTA
GTTGGAGCAAGAGTAAAAT(177)GATGACTGATTTGATTTATGAGAAAGAGGTTTTAAAGCTGTGAGTTATAGC
AGCTTGTCTTACTGTACATTATGAACAGGCTTTTCACTTGTAAATAGCTGCTTCTATAGCTGCTATTTCTGGAGA
ATCAATAGCAATTTAAACCTCGTGATGGAAGGACTGGTGAGATTGTTTGGAAGGGCTGGTGATAATTTATTTGTA
G(23)GTAGTACTGTGCACTTAGGATTGTACTAGGGCCTTTCTAGTTGGCCAACTAG
```

Characterisation and justification for name:

Occurred with variable unique ITS2 clade D sequences, so has been given the designation 'v' (variable).

All other previously undefined sequences (24 clade C) only appeared once or twice and are not reproduced here. GenBank Accession Numbers: MH329543-MH329566

A1.6. Statistical tables for all results

A1.6.1. Results for the *cob* gene

ANOVA result comparing overall richness between Atauro Island and Timor sites:

Variation	df	Sum of squares	Mean square	F	p-value
Region	1	6.5019	6.5019	8.5035	0.03316
Residuals	5	3.8231	0.7646		

PERMANOVA result comparing symbiont type composition between Atauro Island and Timor sites:

Variation	df	Sum of squares	Mean square	F	p-value
Site	1	0.12548	0.12548	5.4416	0.002
Residuals	33	0.76097	0.02306		
Total	34	0.88645			
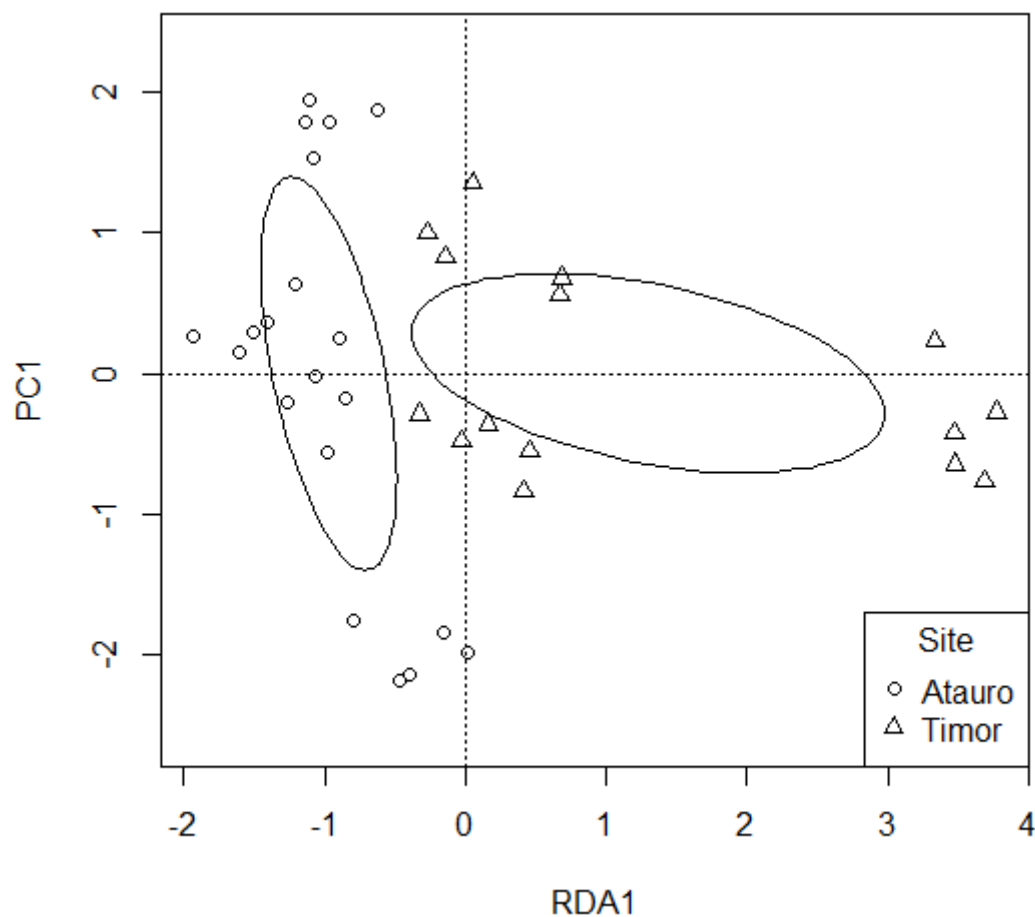
Figure A1.2: Principal Components Plot comparing Atauro and Timor sites in terms of *cob* gene symbiont composition, drawn using the statistical model implemented in the Atauro *vs*. Timor PERMANOVA test.

AMOVA result comparing symbiont type composition between Atauro Island and Timor sites:

$\Phi_{ST} = 0.03$, p = 0.014

Variation	df	Sum of squares	σ^2	% variation
Among groups	1	269.804	1.0124	2.39888
Among pops	5	283.974	0.6066	0.60086
Within pops	414	17150.274	41.426	97.00026
Total	420	17704.052	42.707	

A1.6.2. Results for the ITS2 gene region

ANOVA result comparing overall richness between Atauro Island and Timor sites:

Variation	df	Sum of squares	Mean square	F	p-value
Region	1	18.678	18.678	6.7091	0.04883
Residuals	5	13.920	2.784		

PERMANOVA result comparing symbiont type composition between Atauro Island and Timor sites:

Variation	df	Sum of squares	Mean square	F	p-value
Site	1	0.68604	0.68604	19.814	0.001
Residuals	33	1.14257	0.03462		
Total	34	1.82861			
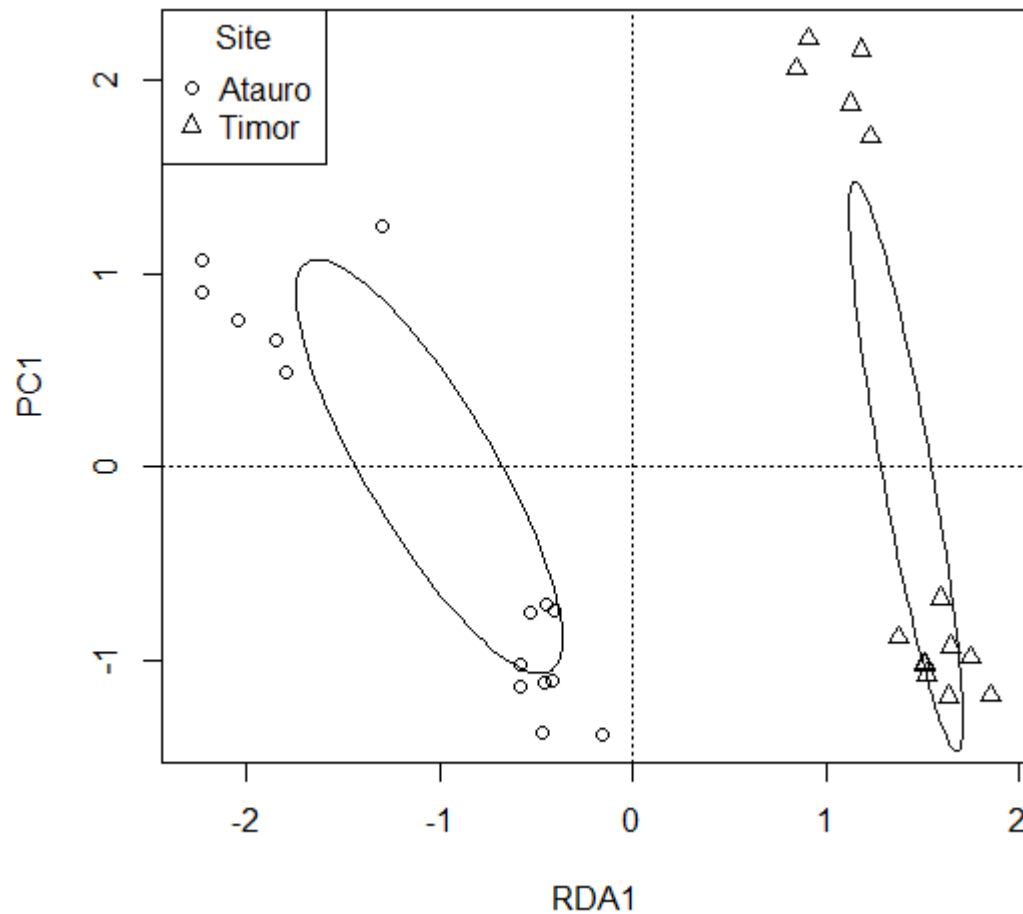
Figure A1.3: Principal Components Plot comparing Atauro and Timor sites in terms of ITS2 gene region symbiont composition, drawn using the statistical model implemented in the Atauro vs. Timor PERMANOVA test. The tight groupings correspond to specific sites within each region, showing that there are also some community differences within individual sites.

AMOVA result comparing symbiont type composition between Atauro Island and Timor sites:

$\Phi_{ST} = 0.02291$, p = 0.037

Variation	df	Sum of squares	σ^2	% variation
Among groups	1	1114.006	1.636	0.40328
Among pops	5	4046.607	7.658	1.88775
Within pops	371	147050.083	396.361	97.70898
Total	377	152210.696	405.655	

A1.6.3. Results for the psbA^{ncr} gene region

ANOVA result comparing overall richness between Atauro Island and Timor sites:

Variation	df	Sum of squares	Mean square	F	p-value
Region	1	8.743	8.743	0.2515	0.6373
Residuals	5	173.782	34.756		
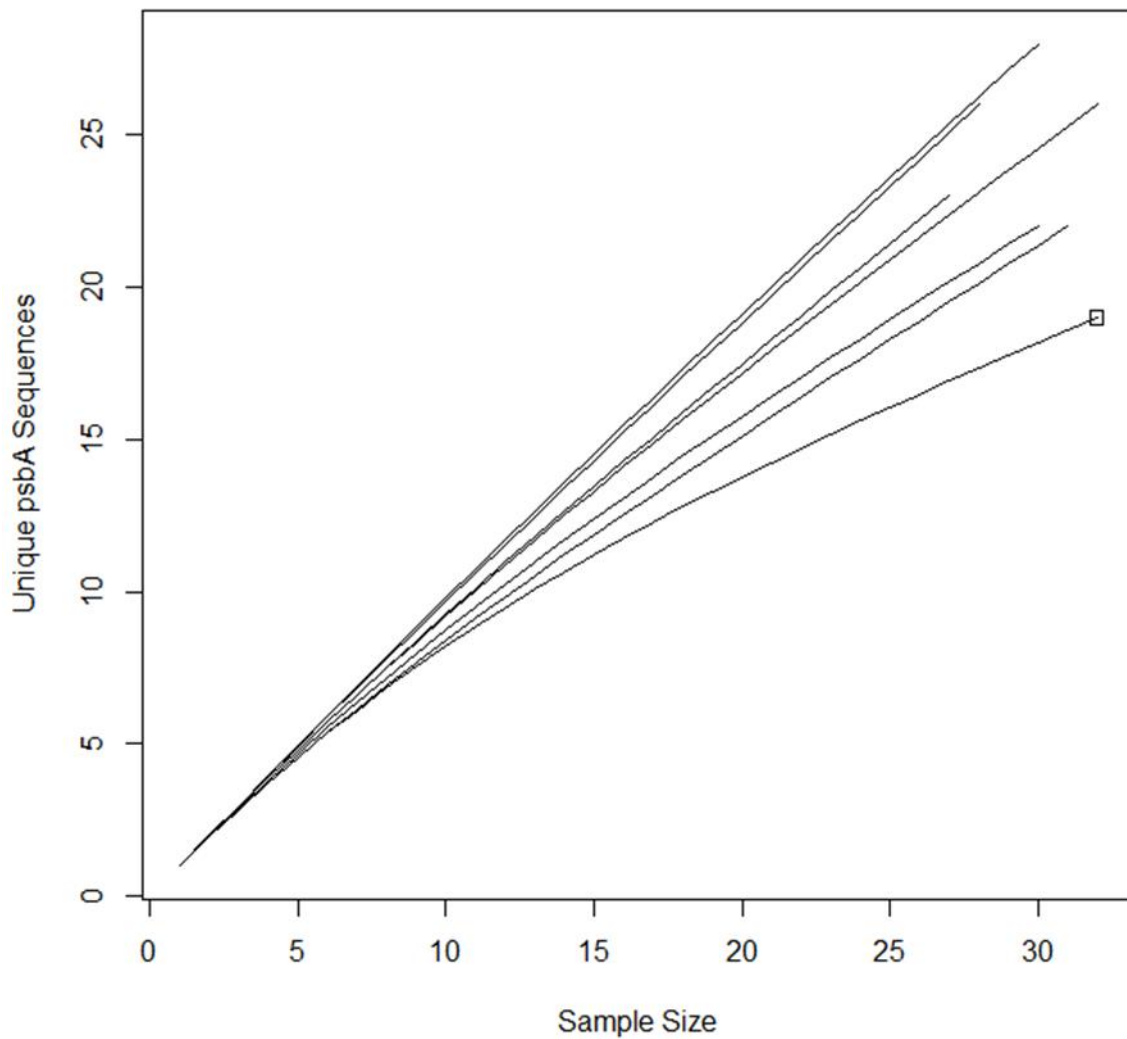
Figure A1.4: Rarefaction curves for the psbA^{ncr} region for all seven sites (single line each), showing how a sampling saturation point was not reached *i.e.* there was still a near-linear discovery rate of new sequences. Therefore, differences between sites and regions are potentially more likely due to chance than to there being genuine diversity differences (as the point where true diversity can be assessed has not yet been reached).

A1.6.4. Results from the OTU analyses

Kruskall-Wallis test of whether the two regions (Atauro Island, Timor) have different numbers of OTUs:

Test Statistic: 0.76142

Degrees of Freedom: 1

p = 0.3829

ANOVA results comparing richness between Atauro Island and Timor sites for the five most common genera:

Acropora:

Variation	df	Sum of squares	Mean square	F	p-value
Region	1	55.59	55.59	21.934	<0.0001
Residuals	56	141.93	2.534		

Porites:

Variation	df	Sum of squares	Mean square	F	p-value
Region	1	21.873	21.873	4.1008	0.05101
Residuals	33	176.013	5.3337		

Pocillopora:

Variation	df	Sum of squares	Mean square	F	p-value
Region	1	12.162	12.162	2.2248	0.1507
Residuals	21	114.795	5.4664		

Platygyra:

Variation	df	Sum of squares	Mean square	F	p-value
Region	1	9.8017	9.8017	9.6365	0.007254
Residuals	15	15.2571	1.0171		

Montipora:

Variation	df	Sum of squares	Mean square	F	p-value
Region	1	9.4554	9.4554	1.5933	0.2261
Residuals	15	89.015	5.9343		
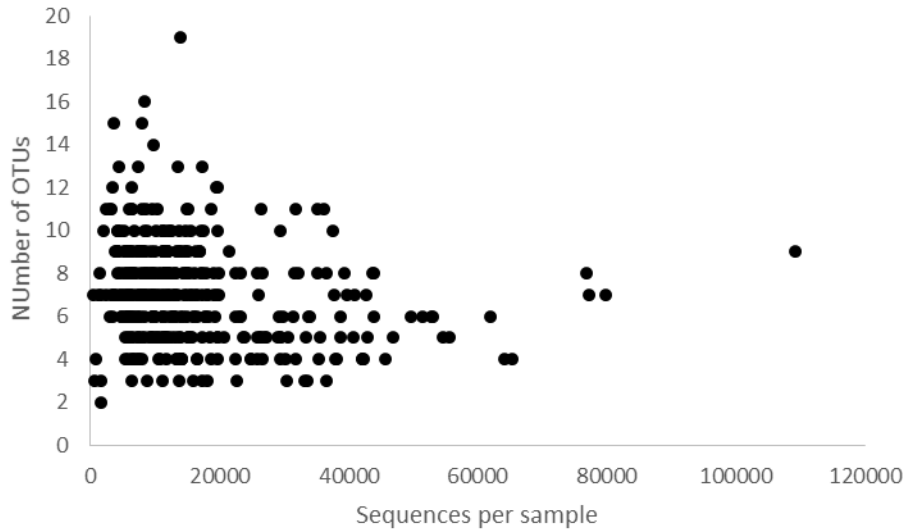
Figure A1.5: Scatterplot showing the lack of relationship between sequencing depth and number of OTUs recovered.

A1.7. Summary of host and symbiont diversity data

Table A1.4: List of all host genera sampled by site.

Site	BBR	BHB	BLS	BSP	HEW	LIE	LIW	TOTAL
Genus	Scleractinians							
Acropora	8	10	13	11	19	10	5	**76**
Porites	3	4	6	4	9	3	8	**37**
Pocillopora	2	2	2	3	13	0	5	**27**
Platygyra	2	0	4	5	5	1	4	**21**
Montipora	2	6	1	3	1	6	0	**19**
Echinopora	3	4	6	1	0	3	1	**18**
Favia	3	1	0	3	5	3	1	**16**
Favites	2	2	1	2	4	2	3	**16**
Galaxea	2	2	3	1	1	2	2	**13**
Goniastrea	4	3	1	2	0	2	1	**13**
Fungia	2	1	1	2	0	4	2	**12**
Hydnophora	2	1	2	3	0	1	2	**11**
Symphyllia	4	0	0	5	0	2	0	**11**
Goniopora	0	2	0	2	1	1	4	**10**
Ctenactis	4	2	2	0	0	1	0	**9**
Lobophyllia	1	2	1	1	1	0	3	**9**
Stylophora	1	2	2	0	1	2	1	**9**
Astreopora	0	1	0	3	0	2	2	**8**
Sandalolitha	3	2	1	1	0	0	1	**8**
Pectinia	2	0	1	1	0	1	2	**7**
Seriatopora	2	3	1	0	0	1	0	**7**
Diploastrea	0	0	1	2	0	2	1	**6**
Turbinaria	2	1	0	1	0	0	2	**6**
Pavona	0	3	0	0	0	2	0	**5**
Merulina	1	0	2	0	0	1	0	**4**
Cyphastrea	0	0	1	0	0	1	1	**3**
Mycedium	0	0	0	2	0	1	0	**3**
Oulophyllia	0	0	1	0	0	2	0	**3**
Polyphyllia	0	0	0	2	0	0	1	**3**

Table A1.4 (cont.)

Acanthastrea	0	0	0	0	1	0	1	**2**
Caulastrea	0	0	1	0	0	1	0	**2**
Coscinaraea	0	0	0	0	0	1	1	**2**
Leptastrea	0	0	0	0	0	1	1	**2**
Echinophyllia	0	0	0	0	0	0	1	**1**
Euphyllia	0	0	0	1	0	0	0	**1**
Halomitra	0	0	0	1	0	0	0	**1**
Heliofungia	0	0	0	0	0	1	0	**1**
Herpolitha	0	0	0	1	0	0	0	**1**
Oxypora	0	0	0	0	0	0	1	**1**
Pachyseris	0	0	1	0	0	0	0	**1**
Plerogyra	0	0	1	0	0	0	0	**1**
Psammocora	1	0	0	0	0	0	0	**1**
Scapophyllia	0	0	1	0	0	0	0	**1**
Genus	**Non-Scleractinians**							
Heliopora	2	0	2	1	1	0	1	**7**
Millepora	1	1	0	0	2	0	0	**4**
Rhodactis	0	0	0	0	0	0	2	**2**
Clavularia	1	0	0	0	0	0	0	**1**
Ricordea	0	0	0	0	1	0	0	**1**
Tubipora	0	0	1	0	0	0	0	**1**
TOTAL GENERA PER SITE	**25**	**21**	**27**	**26**	**15**	**28**	**28**	
TOTAL GENERA PER REGION	**40**				**39**			

Table A1.5: Complete list of symbiont types hosted by each coral genus (ordered as *per* Table A1.4). Note this only includes dominant types. Because some samples only had successful sequences for one or two of the gene regions, there may appear to be some discrepancies between the loci. Cu = unique (unnamed) clade C type. Du = unique (unnamed) clade D type.

Genus	*cob*	ITS2	psbA[ncr]
Acropora	C3, D1	C40, C3z, C3d/C21, D1	C40c, C40d, C40e, C40f, C40g, C40i, C40j, C40m, C40n, C3d/C21p, C3z, D1p
Porites	C15p, C15q, C3, Cu	C15, C40, Cu	C40i, Cu
Pocillopora	C42a, D1	C1d, C42a Cu, D1, Du	C1d, Cvb, D1q, Dva, Dvb, Dvc
Platygyra	C3	C40, C1#, C3z	C40c, C40d, C40g, C40h, C40j, C40m, C40o, C3z
Montipora	C3, C15q, C1/3, Cu	C40, C15, C35.2/C26.b1, C3z, C66, C31, Cu	C40o, C3z, Cu
Echinopora	C3, C42a, C1/3	C1, C1x, C40, C66, C3d/C21, D1	C40c, C40h, C40j, C40m, C1x, C1v, Cu
Favia	C42a, C3, C1, C1/3	C1, C3, C40, C3d/C21	C40d, C40e, C40f, C40h, C40m, C40o, C1x, C1v, Cu
Favites	C3, D1	C40, Cu, D1	C40g, C40h, C40i, C40j, C40k, C40l, C40m, C3z
Galaxea	C42a, D1	C1, C1c/C45, D1, D5	C1x, Cva
Goniastrea	C3, C42a, C1/3	C40, C3z, C1	C1x, C40c, C40d, C40k
Fungia	C3, C42a	C40, C3d/C21, C3z, C1, C1x	C40k, C40n, C40o, C3d/C21p, C1x, C1v
Hydnophora	C3, C42a, Cu	C40, C1, C3d/C21	C40c, C40h, C40i, C1v, Cu

Table A1.5 (cont.)

Symphyllia	C3	C40, C3u	C40d, C40e, C40g, C40h, C40l, C40n
Goniopora	C1/3, C42a, D1	C1, C1b/C1e, C1j, Cu, D1	C1d, C1x, Cu
Ctenactis	C3, C42a	C40, C1x	C40c, C40e, C40m, C1x
Lobophyllia	C3	C40, C3d/C21	C40c, C40e, C40g, C40h, C40n, C40o
Stylophora	C42p	C1z	Cu
Astreopora	C3, C1	C40, C1	C40n, C1u
Sandalolitha	C42a	C1, C1x	C1x, C1v
Pectinia	C3, C1/3	C40, C3d/C21, C66	C40a
Seriatopora	C42a, D1	Cu, D1, D6	Cu, D1p
Diploastrea	C3	C40	C40c, C40d, C40h, C40l
Turbinaria	C3, C42a, D1	C40, C1x, Cu, D1, D5	C40g
Pavona	C42a, D1	C1#, C1x, D1	C1v, Cu
Merulina	C3	C40	C40i, C40k
Cyphastrea	C42a	C1, C1x	C1v
Mycedium	C3	C3d/C21, Cu	C3d/C21q
Oulophyllia	C3	C40	C40g, C40o
Polyphyllia	C3, C42a	C1x	-
Acanthastrea	C3	C40	C40f, C40n
Caulastrea	C3	C40, C3d/C21	C40o
Coscinaraea	C42a	C1x	C1x
Leptastrea	C42a	C1x	C1x
Echinophyllia	C42a	C1x	-
Euphyllia	C3	C3z	-
Halomitra	C3	C3d/C21	C3d/C21q
Heliofungia	D1	D1	-
Herpolitha	Cu	C1x	-
Oxypora	C3	C40	C40n
Pachyseris	C42a	C1x	-
Plerogyra	C1/3	C3	Cu
Psammocora	C42a	C1	C1v
Scapophyllia	C3	C40	C40k
Heliopora	C1/3, Cu	C3, Cu	Cu
Millepora	C1/3	C66	-
Rhodactis	D1	D5	-
Clavularia	Du	D3	-
Ricordea	C42a	C1	C1x
Tubipora	Cu	Cu	Cu

Appendix A2: Appendices for Chapter 3

A2.1. Summaries of PAUP* output

Table A2.1: Evolutionary models selected by PAUP* for each dataset for building maximum likelihood trees, for the SH/AU tests. HKY = Hasegawa, Kishino and Yano (1985); GTR = Generalised Time Reversible (Tavaré 1986); F81 = Felsenstein (1981); JC = Jukes and Cantor (1969); K80 = Kimura (1980); K81 = Kimura (1981). G = gamma coefficient, I = invariant sites.

Dataset	Gene Region	Model	I	G
BBR	*cob*	HKY	-	-
	ITS2	GTR+G	-	0.032
	psbA[ncr]	GTR+G	-	0.206
BHB	*cob*	HKY	-	-
	ITS2	HKY	-	-
	psbA[ncr]	F81+G	-	0.315
BLS	*cob*	HKY	-	-
	ITS2	GTR+G	-	0.0002
	psbA[ncr]	JC+G	-	0.182
BSP	*cob*	HKY+I	0.833	-
	ITS2	F81	-	-
	psbA[ncr]	GTR+G	-	0.267
HEW	*cob*	HKY	-	-
	ITS2	HKY	-	-
	psbA[ncr]	GTR+G	-	0.278
LIE	*cob*	HKY	-	-
	ITS2	K81	-	-
	psbA[ncr]	GTR+G	-	0.419
LIW	*cob*	HKY+I	0.875	-
	ITS2	F81	-	-
	psbA[ncr]	JC+G	-	0.128
Atauro	*cob*	HKY+I	0.805	-
	ITS2	JC+I	0.851	-
	psbA[ncr]	JC+I+G	0.443	0.591
Timor	*cob*	HKY+I	0.874	-
	ITS2	K81	-	-
	psbA[ncr]	JC+I+G	0.484	0.862

Table A2.2: Results of the Incongruence Length Differences tests with all three gene regions concatenated. See Methods for details. Conclusions were drawn at α = 0.05. Bolded values indicate significance.

Dataset	p-value
BBR	0.99
BHB	0.066
BLS	0.99
BSP	**0.001**
HEW	1
LIE	1
LIW	1
Atauro	0.07
Timor	1

Table A2.3: Results of the Shimodaira-Hasegawa and related tests. p-values presented are whether a candidate tree is statistically differentiable from the best tree. Conclusions were drawn at Bonferroni-corrected α = 0.0085. Bolded values indicate potential true significance, while significance that is more likely due to type I error is designated by * (see Discussion). For three tests (psbA^{ncr} tests for LIE, Atauro and Timor datasets) p-values were unable to be computed, due an error with PAUP* optimising the likelihoods for these trees (D. Swofford, *pers. comm.*). SH = Shimodaira-Hasegawa; WSH = Weighted Shimodaira-Hasegawa; AU = Approximately Unbiased.

Dataset	Gene region used for test	Best tree	Tree to compare with best tree	SH p-value	WSH p-value	AU p-value
BBR	*cob*	*cob*	ITS2	0.2195	0.0476	**<0.0001**
			psbA^{ncr}	0.6699	0.5011	0.1226
	ITS2	ITS2	*cob*	0.6252	0.7203	0.4951
			psbA^{ncr}	0.6139	0.5801	0.2189
	psbA^{ncr}	psbA^{ncr}	*cob*	0.0767	0.0013*	<0.0001*
			ITS2	<0.0001*	<0.0001*	<0.0001*
BHB	*cob*	*cob*	ITS2	0.5083	0.0522	**<0.0001**
			psbA^{ncr}	0.6926	0.5164	0.1096
	ITS2	ITS2	*cob*	0.1502	0.009	**<0.0001**
			psbA^{ncr}	0.1502	0.009	**<0.0001**
	psbA^{ncr}	psbA^{ncr}	*cob*	0.1203	<0.0001*	<0.0001*
			ITS2	0.0001*	<0.0001*	<0.0001*
BLS	*cob*	*cob*	ITS2	0.6844	0.5215	0.1746
			psbA^{ncr}	0.6844	0.5215	0.1746
	ITS2	ITS2	*cob*	0.475	0.0789	**0.0083**
			psbA^{ncr}	0.475	0.0789	**0.0083**
	psbA^{ncr}	psbA^{ncr}	*cob*	0.1436	0.0001*	<0.0001*
			ITS2	0.1383	0.0002*	<0.0001*
BSP	*cob*	*cob*	ITS2	0.5347	0.4603	**0.0084**
			psbA^{ncr}	0.6482	0.6911	0.4410
	ITS2	ITS2	*cob*	0.0209	0.0518	**<0.0001**
			psbA^{ncr}	0.3394	0.0137	**<0.0001**
	psbA^{ncr}	psbA^{ncr}	*cob*	<0.0001*	<0.0001*	<0.0001*
			ITS2	<0.0001*	<0.0001*	<0.0001*
HEW	*cob*	*cob*	ITS2	0.6651	0.4869	0.2519
			psbA^{ncr}	0.8284	0.8284	0.7864
	ITS2	ITS2	*cob*	0.6908	0.6908	0.5018
			psbA^{ncr}	0.6908	0.6908	0.5018
	psbA^{ncr}	psbA^{ncr}	*cob*	<0.0001*	0.0002*	<0.0001*
			ITS2	<0.0001*	0.0004*	<0.0001*
LIE	*cob*	*cob*	ITS2	0.6440	0.5125	0.1152
			psbA^{ncr}	0.8433	0.8432	0.7155
	ITS2	ITS2	*cob*	1	1	0.4913
			psbA^{ncr}	0.6639	0.6639	0.4913
	psbA^{ncr}	psbA^{ncr}	*cob*	-	-	-
			ITS2	-	-	-
LIW	*cob*	*cob*	ITS2	0.6145	0.6140	0.5028
			psbA^{ncr}	0.6129	0.6120	0.5028
	ITS2	ITS2	*cob*	0.5276	0.3304	0.1629
			psbA^{ncr}	0.5276	0.3304	0.1629
	psbA^{ncr}	psbA^{ncr}	*cob*	0.0003*	<0.0001*	<0.0001*
			ITS2	0.0004*	0.0001*	<0.0001*

Table A2.3 (cont.)

Atauro	*cob*	psbA[ncr]	*cob*	1	0.9267	0.0367
			ITS2	0.9781	0.0918	**<0.0001**
	ITS2	ITS2	*cob*	**0.005**	**0.0062**	**<0.0001**
			psbA[ncr]	0.2361	**0.0043**	**<0.0001**
	psbA[ncr]	psbA[ncr]	*cob*	-	-	-
			ITS2	-	-	-
Timor	*cob*	ITS2	*cob*	0.5555	0.1400	0.0133
			psbA[ncr]	0.5555	0.1400	0.0133
	ITS2	ITS2	*cob*	0.5360	0.1648	0.046
			psbA[ncr]	0.5452	0.2020	0.0221
	psbA[ncr]	psbA	*cob*	-	-	-
			ITS2	-	-	-

A2.2. Outgroup selection

To make maximum likelihood trees, PAUP* requires an outgroup; if this is not specified by the user, it will pick a random taxon to use. The choice of an appropriate outgroup is a subject of rich debate. However, it appears less critical in the current case. For the ML trees created by PAUP*, it does not incorporate branch lengths, unless explicitly told to do so. Given that the Shimodaira-Hasegawa tests do not consider branch lengths, only explicit topology, this was not done in this case. As such, all branches are treated as the same length, and the genetic distance from the outgroup is largely irrelevant and does not affect the structure of the tree. An example is given below: the following code simply reads in ten samples with identical sequences and builds a maximum parsimony tree using a heuristic search (my commands in bold).

```
P A U P *
Version 4.0a (build 159) for 32-bit Microsoft Windows (built on Nov 18 2017 at 06:28:51)
Mon Jan 15 08:56:27 2018

paup> execute 'outgroup exemplar.nex';

Processing of file "F:\Masters Part Two\PAUP\outgroup exemplar.nex" begins...

Data read in DNA format

Data matrix has 10 taxa, 787 characters
Valid character-state symbols: ACGT
Missing data identified by '?'
Gaps identified by '-'
"Equate" macros in effect:
        R,r ==> {AG}
        Y,y ==> {CT}
        M,m ==> {AC}
        K,k ==> {GT}
        S,s ==> {CG}
        W,w ==> {AT}
        H,h ==> {ACT}
```

B,b ==> {CGT}
V,v ==> {ACG}
D,d ==> {AGT}
N,n ==> {ACGT}

Processing of input file "outgroup exemplar.nex" completed.

paup> **cstatus;**

Character-status summary:
Current optimality criterion = parsimony
No characters are excluded
Of 787 total characters:
All characters are of type 'unord'
All characters have equal weight
787 characters are constant
Number of parsimony-informative characters = 0

paup> **hs;**

    Heuristic search settings:
     Optimality criterion = parsimony
      Character-status summary:
       Of 787 total characters:
        All characters are of type 'unord'
        All characters have equal weight
        787 characters are constant
        Number of parsimony-informative characters = 0
      Gaps are treated as "missing"
    Starting tree(s) obtained via stepwise addition
     Addition sequence: simple (reference taxon = HEW1)
     Number of trees held at each step = 1
    Branch-swapping algorithm: tree-bisection-reconnection (TBR) with reconnection limit = 8
     Steepest descent option not in effect
    Initial 'Maxtrees' setting = 100
    Branches collapsed (creating polytomies) if maximum branch length is zero
    'MulTrees' option in effect
    No topological constraints in effect
    Trees are unrooted

   Heuristic search completed
    Total number of rearrangements tried = 274
    Score of best tree(s) found = 0
    Number of trees retained = 1
    Time used = 0.00 sec (CPU time = 0.00 sec)

paup> **showtrees;**

Note: No outgroup has been defined; tree is (arbitrarily) rooted at first taxon.

Tree 1 (rooted using default outgroup)

```
/-------------------------------------------------------------------- HEW1
|
|                              /----------------------------------------- HEW45
|                              |
|                              +----------------------------------------- BLS91
|                              |
|                              +----------------------------------------- LIE504
|                              |
|                              +----------------------------------------- LIE565
|                              |
\-----------------------------+----------------------------------------- LIE502
                              |
                              +----------------------------------------- BBR277
                              |
                              +----------------------------------------- BHB118
                              |
                              +----------------------------------------- BHB121
                              |
                              \----------------------------------------- BHB163
```

paup> **log stop;**

Without any additional prompting, PAUP* selects one of the taxa as an outgroup and uses it to root the rest of the tree. This is despite the outgroup being an identical sequence. Because of this, sample BHB151 was chosen as an outgroup. Given its potentially contaminated status (see Chapter 2) it cannot be used in the actual analyses, but this way it can still be used as an outgroup without affecting the structure of the tree or the results. This is why BHB151 was chosen. As it is, the tree-based tests carried out are reasonably robust to variation in rooting, even for decisions like choosing to root at the midpoint (van Oppen *et al.* 2001b; Huson and Linz 2016).

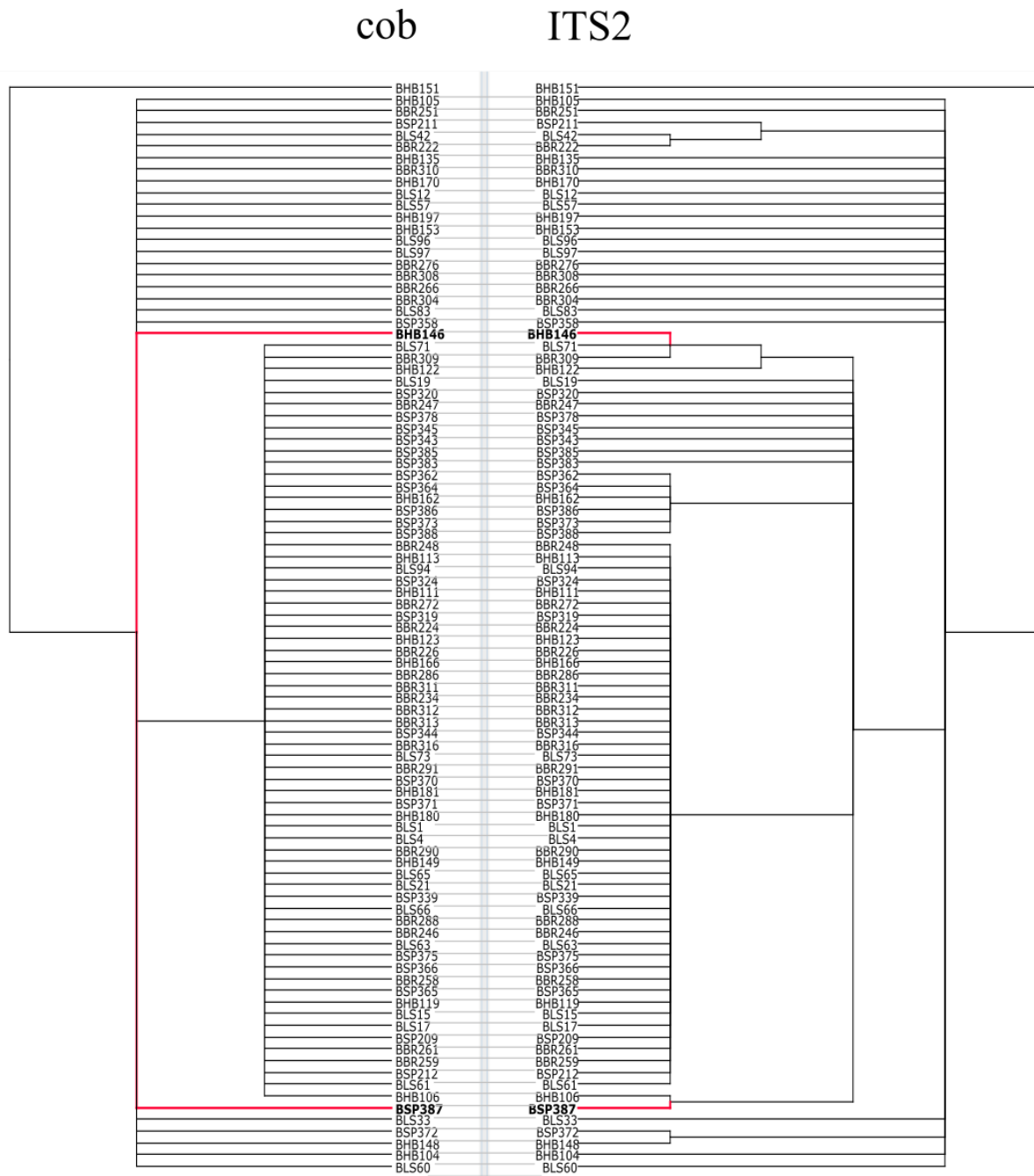## A2.3. Tanglegrams for Atauro Island

Figure A2.1: Tanglegram comparing *cob* and ITS2 gene regions. Red lines and bolded names indicate samples that were identified as incongruent by the tree hybridisation analysis.
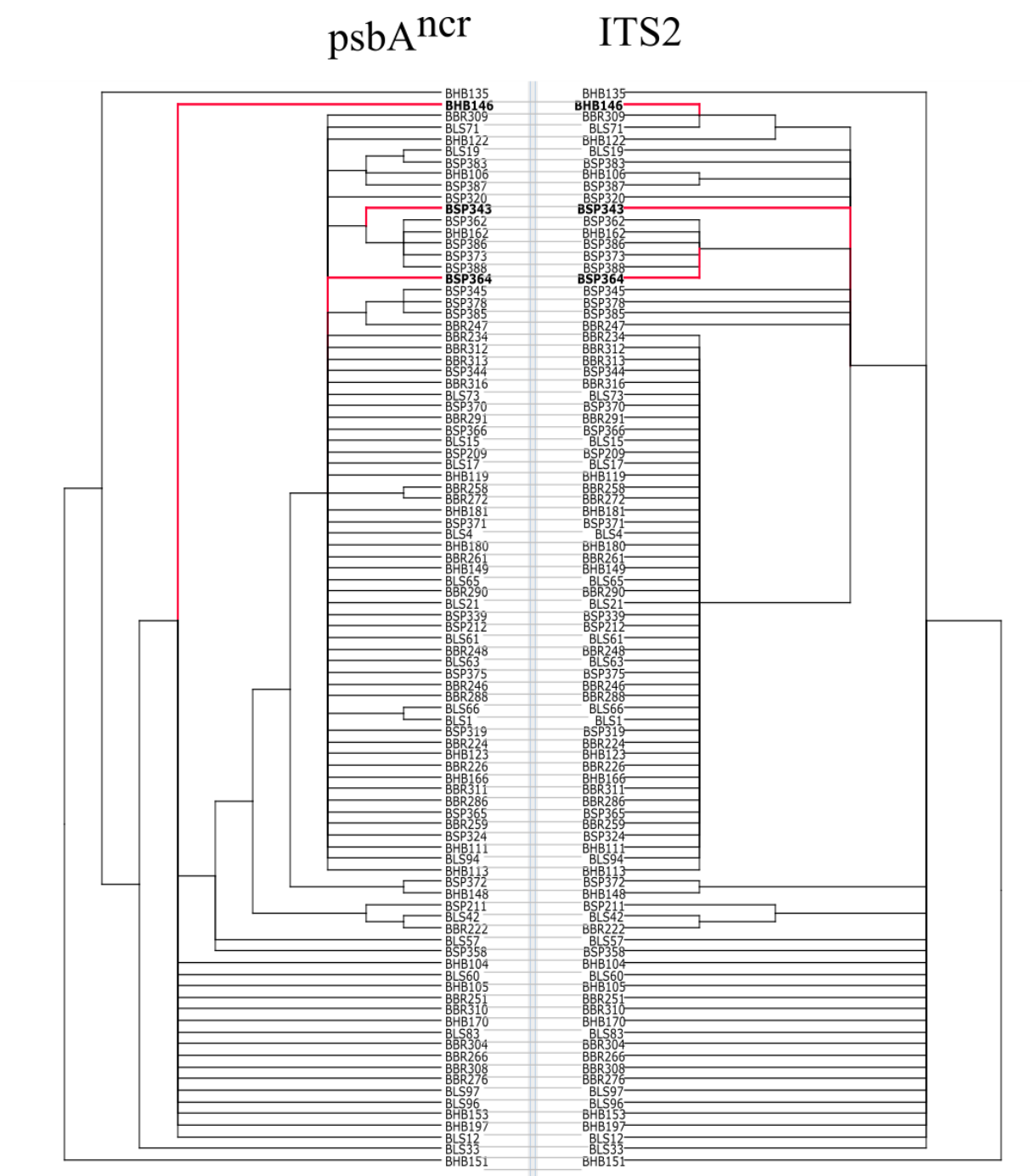
Figure A2.2: Tanglegram comparing psbA^{ncr} and ITS2 gene regions. Red lines and bolded names indicate samples that were identified as incongruent by the tree hybridisation analysis.
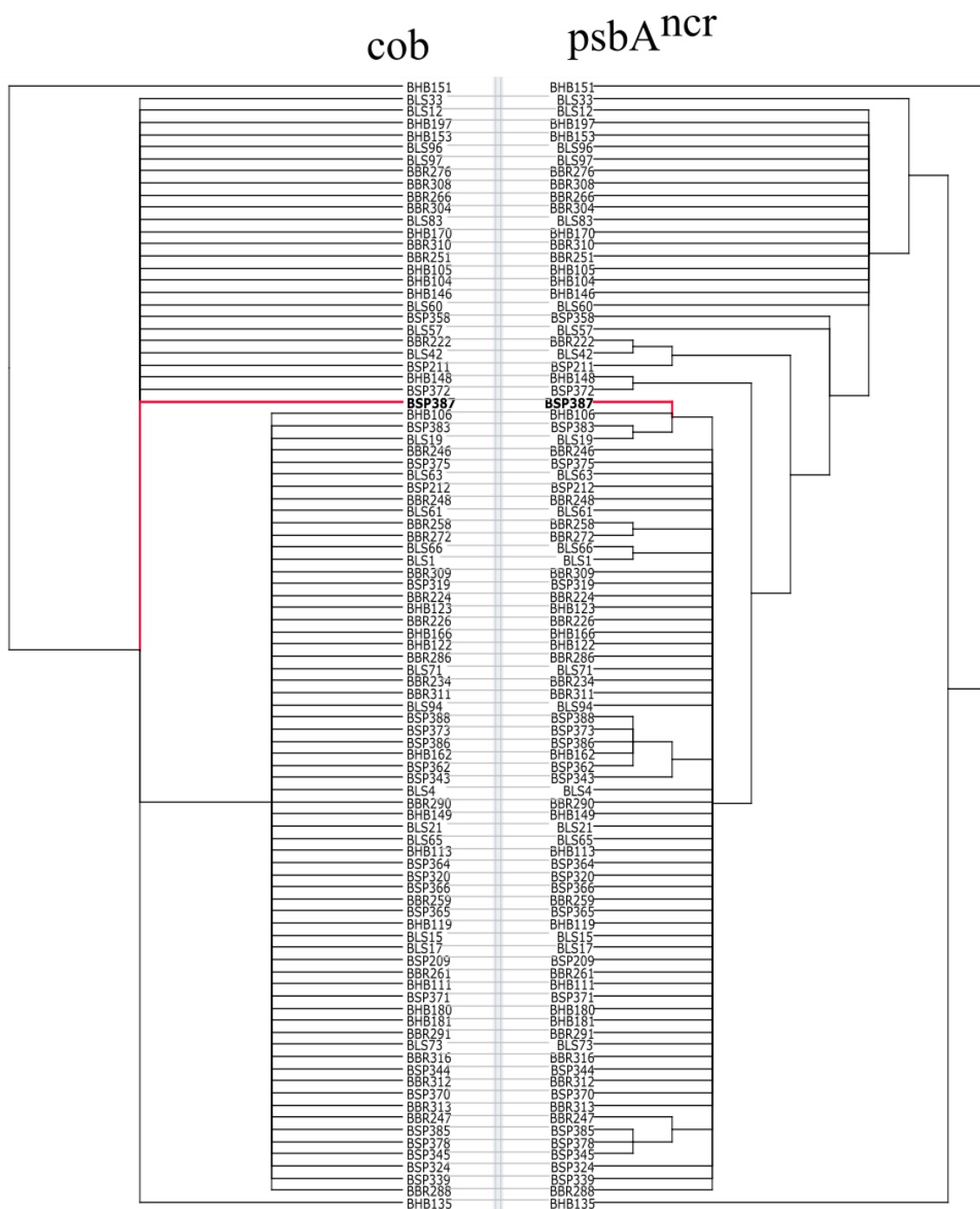
Figure A2.3: Tanglegram comparing *cob* and psbA^{ncr} gene regions. Red lines and bolded names indicate samples that were identified as incongruent by the tree hybridisation analysis.