



HAL
open science

Indexing coral health status by an artificial intelligence approach

Megan Clampitt

► **To cite this version:**

Megan Clampitt. Indexing coral health status by an artificial intelligence approach. *Vegetal Biology*. Université Côte d'Azur, 2023. English. NNT : 2023COAZ6019 . tel-04326159

HAL Id: tel-04326159

<https://theses.hal.science/tel-04326159>

Submitted on 6 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

L'indexation de l'état de santé des coraux par
une approche basée sur l'intelligence artificielle

Megan CLAMPITT

Institut de Recherche sur le Cancer et le Vieillessement, Nice (IRCAN)

**Présentée en vue de l'obtention
du grade de docteur en biologie des
interactions et écologie
d'Université Côte d'Azur**
Dirigée par : Eric Röttinger & Eric Gilson
Co-encadrée par : Marco Milanese
Soutenue le : 4 septembre 2023

Devant le jury, composé de :
Oren Levy, Professeur, Bar-Ilan University
Fanny Houlbreque, Chargée de recherche,
Université de la Réunion
Tries Razak, Chercheuse, National
Research, and Innovation Agency (BRIN)
Jean-Olivier Irisson, Maître de
Conférences, Sorbonne Université
Stephanie Lopez, Ingénieure de
Recherche, Université Côte D'Azur
Sylvie Tambutte, Directrice de Recherche,
Centre Scientifique de Monaco

L'indexation de l'état de santé des coraux par une approche basée sur l'intelligence artificielle

Indexing coral health status by an artificial intelligence approach

Jury :

Rapporteurs :

Pr. Oren Levy, Professeur, Bar-Ilan University

Dr. Fanny Houlbreque, Chargée de recherche, Université de la Réunion

Examineurs :

Dr. Tries Razak, Chercheuse, National Research, and Innovation Agency (BRIN)

Dr. Jean-Olivier Irisson, Maître de Conférence, Sorbonne Université

Invitées :

Dr. Stephanie Lopez, Ingénieure de Recherche, Université Côte D'Azur

Dr. Sylvie Tambutte, Directrice de Recherche, Centre Scientifique de Monaco

Résumé

Les récifs coralliens se détériorent à une vitesse surprenante et le développement de schémas de surveillance rapides et efficaces pouvant évaluer la santé des coraux et ne se concentrant pas uniquement sur l'absence ou la présence de maladies ou de blanchissement est essentiel. La recherche de ma thèse vise à combiner les domaines de la biologie corallienne, de l'informatique et de la conservation marine. La question principale de ma thèse étant : **comment les outils d'intelligence artificielle peuvent-ils être utilisés pour évaluer l'état de santé des coraux à partir de photographies de colonies ?** Étant donné que l'évaluation de l'état de santé des colonies coralliennes individuelles reste mal définie, notre approche consiste à utiliser des outils d'IA pour évaluer les indices visuels tels que les conditions physiques dommageables (organismes perforateurs et prédateurs), le contact avec d'autres organismes (algues, sédiments) et les changements de couleur qui pourraient être corrélés avec l'état de santé. Ceci a été réalisé en utilisant les données photographiques de l'expédition Tara Pacific pour construire la première version de machines d'IA capable de reconnaître automatiquement ces repères visuels, puis en appliquant cet outil à deux types d'études de terrain i). Une étude mise en place à Moorea, en Polynésie française, visant à étudier la santé des coraux au cours du temps. ii). Une étude comparative entre les sites endommagés, vierges et restaurés à Raja Ampat, en Indonésie. L'objectif de ces études est d'extraire les repères visuels qui distinguent les coraux en bonne santé des coraux présentant un risque de mortalité plus élevé. Ainsi, j'ai pu créer un modèle d'IA capable d'annoter automatiquement les photographies de colonies de coraux avec des repères visuels pertinents pour évaluer l'état de santé de la colonie.

Mots – clés : Coraux, Intelligence Artificielle, Santé

Summary

Coral reefs are deteriorating at a startling rate and the development of fast and efficient monitoring schemas that attempt to evaluate coral health without only focusing on the absence or presence of disease or bleaching is essential. My Ph.D. research aims to combine the fields of Coral Biology, Computer Science, and Marine Conservation with the main question of my thesis being: how can artificial intelligence tools be used to assess coral health states from colony photographs? Since the assessment of individual coral colony health state remains poorly defined, our approach is to use AI tools to assess visual cues such as physically damaging conditions (boring organisms & predation), contact with other organisms (algae, sediment), and color changes that could correlate with health states. This was achieved by utilizing photographic data from the Tara Pacific Expedition to build the first version of AI machines capable of automatically recognizing these visual cues and then applying this tool to two types of field studies i). A longitudinal study set up in Moorea, French Polynesia aimed to investigate coral health as assessed by mortality/partial mortality events. ii). A comparative study between damaged, pristine, and restoration sites in Raja Ampat, Indonesia. The objective of these studies is to extract the visual cues that distinguish healthy from unhealthy corals. Thus, I was able to create an AI Model capable of automatically annotating coral colony photographs for visual cues relevant to the current health state of the colony.

Keywords: Corals, Artificial Intelligence, Health

Table of Contents

ACKNOWLEDGEMENTS	6
ABBREVIATIONS	8
1 GENERAL INTRODUCTION	9
1.1 CORAL REEFS	9
1.1.1 <i>Importance of Coral Reefs</i>	9
1.1.1.1 Economic value.....	9
1.1.1.2 Ecological Value.....	11
1.1.1.3 Social/Cultural Value	12
1.1.2 <i>Monitoring Coral Reefs</i>	13
1.1.3 <i>Curent State of Coral Reefs</i>	14
1.2 CORAL PHYSIOLOGY.....	15
1.2.1 <i>The Coral Animal</i>	15
1.2.2 <i>Microbiota</i>	21
1.2.3 <i>Predator Prey Interactions</i>	22
1.2.4 <i>Towards Defining Coral Health</i>	26
1.3 TECHNOLOGICAL ADVANCES.....	28
1.3.1 <i>Scuba Diving</i>	29
1.3.2 <i>Photography</i>	31
1.3.3 <i>Artificial Intelligence</i>	32
1.3.3.1 Machine Learning	33
1.3.3.2 Deep Learning	33
1.3.3.3 Convolutional Neural Networks	34
1.3.3.4 Artificial Intelligence in the Medical Field	34
1.3.3.5 Artificial Intelligence in Marine Research.....	35
1.3.4 <i>The Combination of Scuba Diving, Photography, and AI for Coral Reef Conservation</i>	36
1.4 PHD OBJECTIVES	36
1.5 STUDIED REEF SITES	38
1.5.1.1 The Tara Pacific Expedition.....	38
1.5.1.2 Indonesia	39
1.5.1.3 French Polynesia.....	42
1.5.1.4 Site Conclusions.....	45
2 ARTICLE 1: OPEN SCIENCE RESOURCES FROM THE TARA PACIFIC EXPEDITION ACROSS CORAL REEF AND SURFACE OCEAN ECOSYSTEMS (2023)	46
2.1 CHAPTER 2 FOREWORD	46

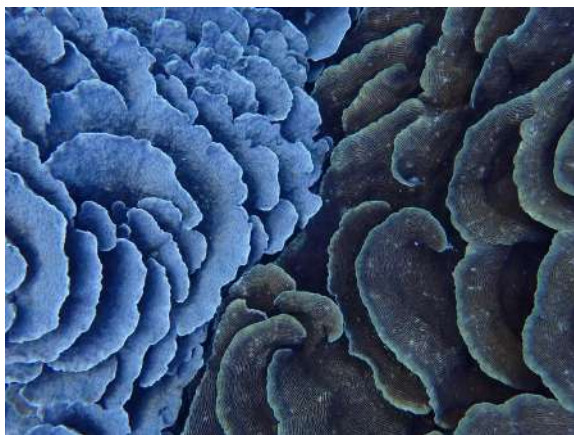
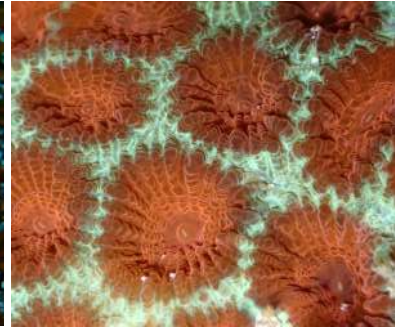
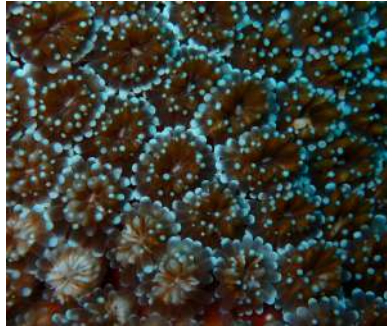
2.2	OPEN SCIENCE RESOURCES FROM THE TARA PACIFIC EXPEDITION ACROSS CORAL REEF AND SURFACE OCEAN ECOSYSTEMS (LOMBARD ET AL., 2023)	47
3	ARTICLE 2: UTILIZING ARTIFICIAL INTELLIGENCE TOOLS TO AUTOMATICALLY ANNOTATE VISUAL CUES RELATED TO REEF-BUILDING CORAL HEALTH (IN PREP)	72
3.1	CHAPTER 3 FOREWORD	72
3.2	UTILIZING ARTIFICIAL INTELLIGENCE TOOLS TO AUTOMATICALLY ANNOTATE VISUAL CUES RELATED TO REEF-BUILDING CORAL HEALTH.	73
3.3	SUPPLEMENTARY INFORMATION	102
4	ARTICLE 3: EVALUATION OF SITE TYPE BASED ON BENTHIC COMPOSITION, HARD CORAL DIVERSITY, AND CORAL PHYSIOLOGICAL ATTRIBUTES (IN PREP)	118
4.1	CHAPTER 4 FOREWORD	118
4.2	EVALUATION OF SITE TYPE BASED ON BENTHIC COMPOSITION AND HARD CORAL DIVERSITY	120
4.3	SUPPLEMENTARY INFORMATION	148
4.4	CHAPTER 4 ANNEX	155
5	CONCLUSIONS & PERSPECTIVES.....	167
5.1	CONCLUSIONS	167
5.2	DISCUSSION & PERSPECTIVES	169
6	REFERENCES	177

*To the coral reefs,
who are
forever and always
my muse
my happy place
and a living breathing example
of extraordinary beauty*



One way to open your eyes to unnoticed beauty is to ask yourself, what if I had never seen this before? What if I knew I would never see it again”?

-Rachel Carson



Acknowledgements

Aux Eric – je vous remercie tous les deux d’avoir m’offrir l’opportunité de faire partie de ce beau projet et de m’avoir confiée cette thèse. Merci de m’avoir accompagné tout au long de cette aventure où on a parcouru le monde 😊. Je vais me souvenir pour toujours de nos séjours en Indonésie, les plongées, les récifs magnifiques, et la beauté qui était partout autour de nous. Je vous remercie tous les deux pour vos points de vue intéressants pendant nos échanges pour avancer ce projet et pour vos corrections de ce manuscrit qui l’a rendu bien plus clair et intéressant. Merci. (**Bapak**) **Éric** : un merci supplémentaire pour ton partage de la culture et la langue Indonésien avec moi que j’ai vraiment beaucoup apprécié (le durian un peu moins 😊). **Eric (muda)** – A special thanks for your comments and suggestions relevant to the biological aspects and the science behind this PhD, you really pushed me to think above and beyond p.s.#snake man/snack mana/snake mahna forever 😊

To **Marco**, who not to be dramatic but without which none of this would have been possible. What a wild ride it has been! Me coming with absolutely NO background in computer science/coding/AI you name it. You taught me how to code, and (literally) everything I know about AI. Always with such patience, always available to help, and always full of weird Italian sayings that we don’t have in English (my personal favorite: you can’t squeeze blood out of a potato 😊). I am the padawan to your Yoda, the karate kid to your Mr. Miyagi, the Arthur to your Mrs. Frizzle. Name a great teacher, and I show you Marco Milanesio in return. Thank you so much - for your time, for our weekly meetings, for dealing with my sh***y code 😊, my millions of excels, and my really well thought out photo naming process 🐱. Beyond that, thank you for all the emotional support-I’m not 100% sure I would have finished this PhD, if you hadn’t been involved. Grazie mille.

À **Florentin!** Mais toi.... Toi, tu étais le joueur qui est venu de nulle part pour sauver le match dans la dernière minute de la quatrième partie quand c’était sûr qu’on allait perdre. Je te dédie le quatrième chapitre qui aurait été tellement différent (lis pire, horriblement pire 😊) sans toi. Je te remercie également pour ton soutien, et tous tes sages conseils autour de la rédaction et la fin de thèse (et pour les jolis couleurs accordés sur mes graphs 🐱)

I would like to thank all the jury members for taking their time to participate in my doctoral defense to **Fanny Houlbrequé** and **Oren Levy** for reading and judging this manuscript along with **Tries Razak**, **Jean-Olivier Irisson**, **Stephanie Lopez**, & **Sylvie Tambutte** for kindly accepting to be part of my PhD defense.

A mes équipes ! Pour tous les moments qu’on s’est passé ensemble dans et hors le labo : le sport avec JB, les apéros plages, les soirées jeux, les pauses café, les pauses midi avec les discussions marrants, etc. Je vous remercie tous pour votre soutien (surtout pendant la période où je crois que je suis venue au labo chaque vendredi essentiellement que pour pleurer 😊), vos perspectives scientifiques, vos questions pendant mes présentations, et tous qui vous avez fait pour m’aider à améliorer et finir ce travail. Shoutout à **Clara**, ma gym Buddy 🏋️ qui était toujours là pour non seulement aller se défouler avec le sport mais aussi pour me soutenir émotionnellement. Ta présence m’a beaucoup aidé pendant cette thèse ❤️. À **Steph** - tes conseils et tes mots d’encouragement valait beaucoup pour moi et m’a vraiment poussé à continuer dans les moments difficiles. Je les chérissent. À **Romane**, Moorea aurait été complètement différent sans toi ! J’ai adoré nos petits voyages ensemble, nos moments à la voltige 😊, nos pauses café, et nos plongées (toujours rapide et efficace 😊) #vivelepoissonrougedeclement. A **Marta**, qui était une autre actrice clé dans ma survie ces derniers mois. À **Rita** – ma déesse des memes 🙏. À **Nico** – pour les 80 centimes que tu m’as donnés pour un café l’autre jour 😊. Non, Je plaisante - mais vraiment merci pour tes conseils sur la fin de thèse et la rédaction et tes petits check-ins de comment ça allait 🤔. A **Eli** et les balistes qui sèment la terreur. À **Alex**, toujours de bonne humeur Et finalement à notre chère **Christian** qui était si présent dans le labo, et surtout autour du mon bureau 🐱.

Thank you to the members of my CST, **Christian Voolstra & Jean-Olivier Irisson**, for taking the time to follow my progress and for the interesting discussions during our meetings as well as your input and guidance which helped move this research forward.

À **Noémie**, sans qui, je n'aurais jamais connu Misool qui est vraiment le dernier paradis. Je te remercie pour tous que t'as fait pour ce projet. To **Virly**, who was integral to the success of the field visits and is an absolute pleasure to talk to. Thank you for all your help as well as all the nice moments we shared and the interesting discussions. I hope our paths will cross again. Terima Kasih Banyak to everyone that I had the pleasure of meeting from the **Misool Foundation** and the **Misool Resort** with a big shoutout to **Sabine, Sue, Mark, Andy, Irvan, Nanda, and Kiki** who were integral players during my first visit and helped enormously in determining the sites and providing information about the area. To all of the members of the ranger patrol: thank you so much for your hospitality, your kindness and your incredible cooking skills (I will forever be craving that peanut sauce...). Thank you for sharing your culture, your island, and your laughter with me. My time at Kalig was something I will truly treasure forever ❤️. A special shout to **Sahlan, Lagai, Ganti, Muid, & Awan** who were essential during data collection and diving and of course **Ayub**. Ayub who not only was essential to organizing our field work and data collection but who is never without a smile and always ready to help. Thank you so much for your warm welcome. To **Paul Boli, Ridwan Sala** and the team at **UNIPA**. À tout le monde que j'ai croisé au **CRIOBE**/membres du **CRIOBE** – merci pour votre accueil. À **Émilie, Gaëlle, Guillaume & Clément** pour votre aide avec les échantillonnages. Et aux boys (**Clément, Hugo, et Jules**) – j'ai vraiment adoré nos trajets de scooter, les repas de famille, les petits apéros plages – en gros notre temps ensemble. À **Valérie, Nicolas et Caroline**, je vous remercie pour votre aide avec les voyages qui était toujours si compliqués.

To the members of **Tara Pacific Consortium** for providing the dataset that is the foundation of this project and for the interesting and thought-provoking discussions during our meetings. To the **MSI**, above all **Stephane et Valeria**. To **LIA- Ropse**.

Et finalement, à mes amis qui était là tout au long de cette aventure ❤️. Surtout au fantastic four - mon petit soleil **Manu**, Mon little merMAN **François**, et mon jacuzzi gang, **MC et Juliette**. Je pense toujours à une citation d'un de mes livres préférés quand je pense à vous - "They were the best sort of Friends. The sort everyone hopes for but no one deserves". Je ne sais pas ce que je ferais sans vous. Je vous aime. To **Kristen**, who as someone with a PhD herself, has always been a listening ear and a source of guidance throughout this journey. I will also honestly forever be grateful for your delivery of enough food to feed a small village during my final weekend of writing 🍕. To **Jenny, Kirsten, Johnny, Nick, Frazer, Erin, Kasen, Nina, Alice, Laura**.

To my entire **family** who have been there since Day 1 and have offered a listening ear and support throughout this PhD above all **George, Emily, Tess, Mom, Stevie, Ilona, and Klaus**. I love you. Thank you. For little **Case**, I can't wait to share the ocean with you ❤️.

Abbreviations

- AI – Artificial intelligence
- AIMS - Australian Institute of Marine Science
- CCA -Crustose Coralline Algae
- CNNs -Convolutional Neural Networks
- CRIOBE – Center for Insular Research and Environmental Observatory
- COTS – Crown of Thorns Starfish
- CT – Coral triangle
- FP – French Polynesia
- GCRMN - Global Coral Reef Monitoring Network
- ICRI - International Coral Reef Initiative
- IPCC – International Panel on Climate Change
- LIT – Line Intercept Transect
- MPA – Marine Protected Areas
- NOAA- National Oceanic and Atmospheric Administration
- PLS – Pink Line Syndrome
- PSU – Practical Salinity Unit
- TP/TPE – Tara Pacific Expedition
- WHO – World Health Organization

1 General Introduction

1.1 Coral Reefs

1.1.1 Importance of Coral Reefs

Coral reefs only occupy .09 percent of the world's oceans but are incredibly important, biodiverse ecosystems with significant economic, ecological, and social value (Spalding et al., 2001) (Gonzalez-Rivero et al., 2014). They support a quarter of all marine life comprising around two million different species while additionally providing economic wealth in terms of tourism and fishing, an essential source of protein, a natural barrier offering protection from storms and erosion, cultural and aesthetic value, as well as being a treasure-trove of compounds for medicinal research (Gómez-Rios et al., 2018) (Neal et al., 2015) (Haas et al., 2015). Furthermore, the study of these ecosystems allows for a better understanding of our planet as a whole. Through the continued study of coral reefs, we can amass important information about global warming and water pollution levels while also gaining a better understanding of past climactic events due to the long lifespans of corals (Gómez-Rios et al., 2018).

1.1.1.1 Economic value

Constanza & colleagues (1997) described ecosystem services as “the benefits human populations derive directly or indirectly from ecosystem function”. In coral reefs this can be represented by goods *i.e.*, fish for nourishment and services such as coastal protection. The economic value of many of these services cannot be measured directly indicating that estimates for various economic values of coral reefs are undervalued, however the loss of irreplaceable services provided by coral reef ecosystems would result in huge economic changes (Constanza et al., 1997). For example, the shift from healthy coral reefs to degraded reefs in the coral triangle would result in a 2.4\$ billion USD loss per year as opposed to a 4.1\$ billion USD increase under a healthy reef scenario (UN Environment et al., 2018). A report jointly created by The Prince of Wales' International Sustainability Unit (ISU), the United Nations Environment Programme (UN Environment), and the International Coral Reef Initiative (ICRI) stated “Proactive policies to protect and restore the health of the world's coral reefs could generate a substantial economic gain, provide important societal benefits, including to local communities, and help deliver the UN Sustainable Development Goals”. Thus, highlighting that coral reefs are one of the most economically valuable and heavily utilized ecosystems for humankind (Spalding et al., 2001).

Coral reefs are estimated to provide almost US\$3 trillion per year in goods and services with US\$36 billion focused on coral reef tourism (Souter et al., 2020). Economically, tourism and recreation provide the most monetary value from coral reefs followed by coastal protection and fisheries (Costanza et al., 1997). Coral reefs provide direct values such as resources (food, construction, pharmaceuticals', fish, and coral for the aquarium trade), tourism and recreation, and educational and scientific interests, as well as indirect values such as biological support, coastal protection, global heritage and known and unknown future uses (Souter et al., 2020) (Moberg & Folke, 1999).

In 2017, direct and indirect economic returns from commercial fisheries were estimated to account for \$480 million USD for the Mesoamerican reef and \$5,850 million USD for the Coral Triangle Region with 240\$ million and 2,925\$ million accounting for direct and indirect returns respectively (UN Environment et al., 2018). This becomes more significant when accounting for the fact that the demand for fish has doubled in the last 50 years and is expected to double again around 2030 (Miththapala, 2008). Aquaculture and the aquarium trade is another sector benefiting from reefs with fish, shrimps, pearl oysters, clams, algae, mollusks, and sponges all being harvested (Spalding et al., 2001). Aquaria is increasing in popularity with approximately 2 million people involved and profits around \$200-330 million USD annually (Wabnitz et al., 2003)

In terms of medicinal benefits, the possibility of discovering a new drug is 300 to 400 times more likely to be found amongst coral reef species than terrestrial sources due to extreme phylogenetic diversity (Bruckner, 2002). Marine species produce natural products that can have many attractive properties for human use such as antimicrobial, antiviral, antitumor, antioxidative, antihypertensive, antiatherosclerosis, anticoagulant, immunomodulatory, analgesic, anxiolytic, anti-diabetic, appetite suppressing, and neuroprotective making them noteworthy prospects for pharmaceutical companies (Cheung et al., 2015). Many marine species have been used in traditional Chinese medicine and are now being used in allopathic medicines as well. (Miththapala, 2008). Some examples of drugs isolated from coral reef species include a cancer therapy derived from algae, a painkiller isolated from cone snails (Figure 1), and the antiviral drugs (Ara-A and AZT) derived from Caribbean sponges (Bruckner, 2002). Furthermore, marine algae have not only medicinal properties (antimicrobial, antifungal, antiviral) but can also contain trace minerals and vitamins (Bharathi, 2021). Thus,

highlighting the medicinal treasure trove of coral reefs and the discovery potential for future drugs and compounds.

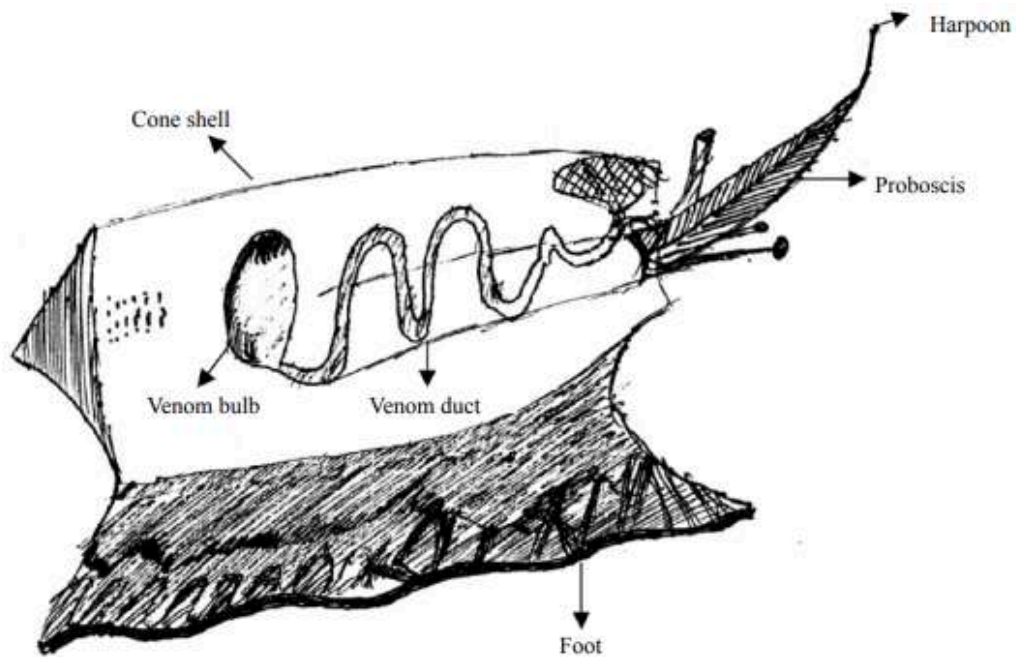


Figure 1: “*Conus magnus* is a snail from the Indo-Pacific region. Its toxin is an ω -conopeptide with analgesic characteristics. Prialt® is a synthetic copy of this toxin generally called ziconotide.” (Mironidou-Tzouveleki & Dokos, 2007)

1.1.1.2 Ecological Value

Corals are the most important bio constructors on the planet and the 3-D structure they provide plays a huge role not only in the diversity and density of other organisms such as fish and algae but also in coastal protection and tourism (Graham & Nash, 2013). Corals create a physical barrier which can diminish energy from waves, tides and extreme weather events such as cyclones thus preventing coastal erosion, flooding, and loss of infrastructure (Miththapala, 2008). The complex structure they create provides habitats for other marine organisms and acts as an important nursery ground. Rugosity, such as that created by corals, was found to increase fish species richness (Gratwicke & Speight, 2005). Likewise, structural complexity has been linked to increased fish biomass and abundance while also making an ecosystem more able to host various size classes of fish (Rogers et al., 2014).

There are more species per unit area of coral reefs than any other ecosystem with over one third of all marine fish species found on coral reefs (Spalding et al., 2001) (Paulay, 1997). This enormous diversity is beneficial to the ecosystem. For example, it has been shown that greater coral diversity is linked to higher fish abundance (Komyakova et al., 2013). The coral triangle region (Philippines, Indonesia, Malaysia, Timor Leste, Solomon Islands, and Papua New Guinea) harbors the highest diversity of marine species with a marked decrease in diversity

happening across the Pacific from west to east with entire classes and orders of organisms disappearing (Spalding et al., 2001) (Paulay, 1997). However, even though coral reefs harbor the greatest biodiversity of all the habitats on earth, up to 74% of coral species remain undiscovered, meaning that the actual diversity is not truly known (Paulay, 1997) (Sheppard et al., 2017) (Souter et al., 2020).

1.1.1.3 Social/Cultural Value

Finally, an underrepresented field of study of coral reefs is cultural ecosystem services which arise from the interactions between humans and ecosystems and results in non-material benefits (capabilities and experiences) obtained from the ecosystem (Retka et al., 2019). Traditionally, coral reefs have helped humans fulfill their basic needs of food and shelter by providing them with a place to live with easy access to sustenance and building materials whereas nowadays a heavier focus is on tourism and recreation (Wilkinson, 1996). In terms of food, coral reef fisheries provide a vital source of protein, lipids and micronutrients which might otherwise be inaccessible to islanders and inhabitants of developing countries while also playing a crucial role in the livelihood of millions of people (Cinner, 2014). In fact, the fish and invertebrates harbored by one km² of actively growing reef could support the protein needs of more than 300 people when no other protein sources are available (Jennings & Polunin, 1996). Reef fisheries can also contribute to one's cultural or personal identity through their role in ceremonies, feasts, social standing, networking, exchanges of goods and services, and the perpetuation of traditional skills and practices (Cinner, 2014) (Vaughan & Vitousek, 2013).

Another example of an important cultural ecosystem service is aesthetic value or artistic inspiration (Moberg & Folke, 1999). Overall, very few studies evaluate the aesthetic value of coral reefs even though it can be directly linked to touristic values (Haas et al., 2015) as well as environmental values such as biological diversity (Becken et al., 2017). For example, when using non-experts to assess beauty in the marine environment, water clarity, fish abundance and coral topography were all reliable predictors of aesthetic response ratings (Pert et al., 2020). Additionally, aesthetic values could prove to be a useful proxy for coral reef health with the added benefits of fostering public engagement by involving non-experts and citizen scientists as well as acting as a possible alternative when ecological monitoring is not possible or too expensive (Pert et al., 2020).

1.1.2 Monitoring Coral Reefs

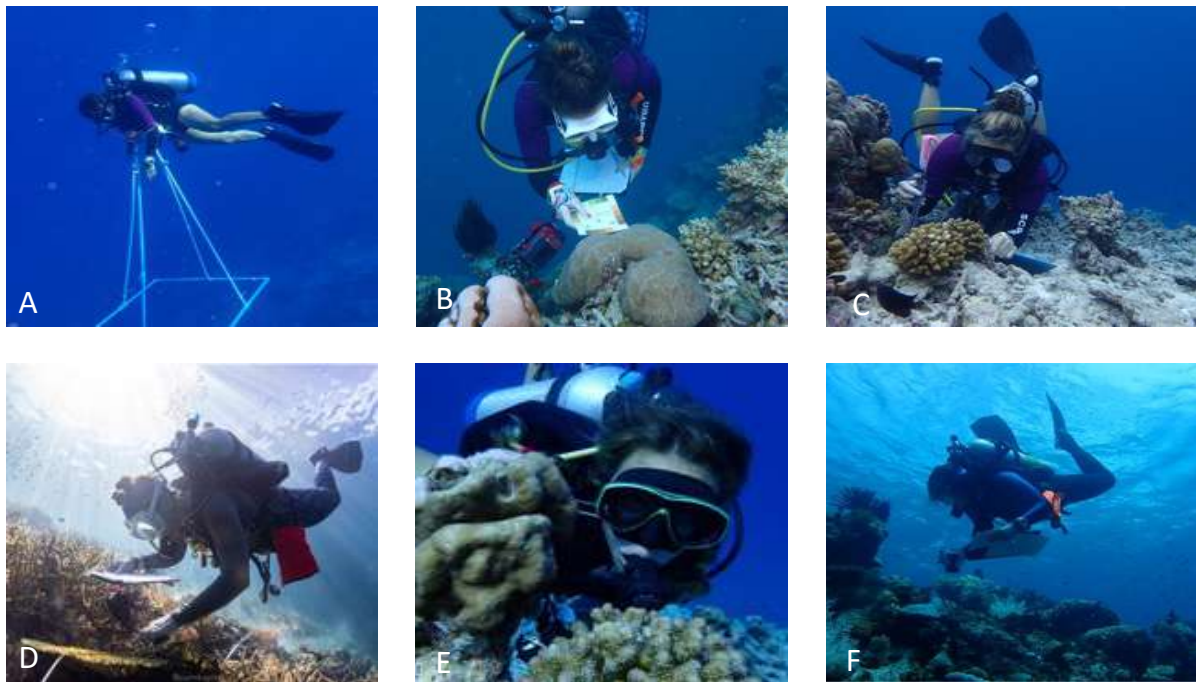


Figure 2: Examples of some different monitoring strategies with different tools utilized (A) Quadrat for photographic monitoring (B) Coral Watch Health Card for bleaching monitoring (C) Calipers for measuring growth (D) Measuring tape for benthic surveys (E) Observation (F) Camera for colony photographs (Picture D taken by Putro Pambajeng and Picture B taken by Eric Rottinger)

Surveying ecosystems gives important information on the distribution and abundance of the various communities present and helps determine the status of an ecosystem at both regional and global scales (English et al., 1997). Through monitoring, we gain significant information on the distribution and abundance of the various communities present and can compare this information between and within sites. This allows scientists to evaluate how ecosystems change over time to better understand ecosystem functioning as well as how ecosystems react to stressors, disturbance events and anthropogenic influences. Furthermore, monitoring is an important tool to inform and guide management decisions and conservation practices. Marine spatial planning and marine management cannot occur without effective monitoring that allows for evaluation and adaptation (Day, 2008). Many countries have their own monitoring programs such as the National Oceanic and Atmospheric Administration (NOAA) from the USA or the Australian Institute of Marine Science (AIMS) from Australia. Apart from country specific organization, there are many organizations that operate on a more global level such as the Global Coral Reef Monitoring Network (GCRMN) and the International Coral Reef Initiative (ICRI). Current standard coral reef monitoring strategies rely on divers and photographs to collect information such as coral cover, fish abundance, species assemblages,

and health-related visual signs (*i.e.*, the presence of lesions) (Aprill et al., 2023) (Figure 2). Obura and colleagues (2019) suggest that the combination of hard-coral cover and composition, macro-algal canopy cover, and fish diversity and abundance offer a robust description of reef status. Typically, one or a combination of the above-mentioned metrics are used to evaluate overall reef health. Globally, reef monitoring has increased substantially since 1978, but is not keeping up with the pace of reef decline (Souter et al., 2020). Standardization and consistent monitoring and reporting are critical for the conservation and management of coral reefs (Tun et al., 2005). Continued, collaborative, efficient, and rapid monitoring efforts are needed on a global level to keep pace with worldwide reef degradation.

1.1.3 Current State of Coral Reefs

Despite providing important ecosystem services and societal benefits, coral reefs are declining at an alarming rate (Neal et al., 2015) (Bellwood et al., 2004). In a ten-year period, threatened reefs increased by 30 percent with increases seen across all threat categories and all regions (Burke et al. 2011). The collapse of coral reef ecosystems is associated with loss of sensitive species or large colonies, loss of coral cover, or transitions from coral-dominant to non-calcifier-dominant (macroalgae, sponge, corallimorph, soft coral) states (McClanahan et al., 2022). Currently, over 60% of the world's reefs are facing immediate local threats with that number increasing to 75% when considering thermal stressors (Burke et al., 2011). In 2018, it was calculated that 14% of the world's coral reefs had been lost equating to all the coral found on Australian reefs (Souter et al., 2020). The main culprit for this decrease was an increase in frequency and distribution of mass bleaching events giving coral colonies little time to recover with reefs also experiencing additional compounding stressors such as disease, predator outbreaks, storms, overfishing, and pollution (Souter et al., 2020). Anthropogenic influences are a major contributor to reef decline with overfishing negatively affecting over half of all the world's reefs and increases in human populations leading to increases in sediment and nutrient levels which can both detrimentally affect reef functioning (Burke et al., 2011) (Hallock, 2001). As coral coverage decreases, algal coverage has been increasing indicating a shift towards algal-dominant reefs which lack the complex 3-D topography necessary to provide many of the ecosystem goods and services provided by coral reefs of which humans are dependent on (Souter et al., 2020). Current predictions are grim, stating that 99% of reefs could be lost by 2050 without drastic intervention (IPCC, 2018). The degradation of these ecosystems negatively impacts coastal communities who are highly dependent on reefs in terms of food

and income (Gonzalez-Rivero et al., 2014) while also adversely affecting water quality and abundance and diversity of reef inhabitants (Haas et al., 2015).

1.2 Coral Physiology

1.2.1 The Coral Animal

When thinking of corals, one might imagine a travel brochure picturing white sand beaches and crystal-clear tropical waters where the dazzling bright colors of coral reefs are peeking through from below the surface. Today, there are roughly 1,300 living species of Scleractinia or reef building corals with almost half of them found in clear, shallow, tropical waters (Budd et al., 2010) (Campoy et al., 2020). Here, I will focus on tropical zooxanthellate corals. When first discovered, corals were thought to be plants due to their sessile nature and the inability of scientists to observe them up close and for extended periods of time in their natural environment (Goreau et al., 1979). However, corals are animals who belong to the phylum Cnidaria which is distributed worldwide and includes sea anemones, sea fans, sea pens, jellyfish, and hydroids (Figure 3).

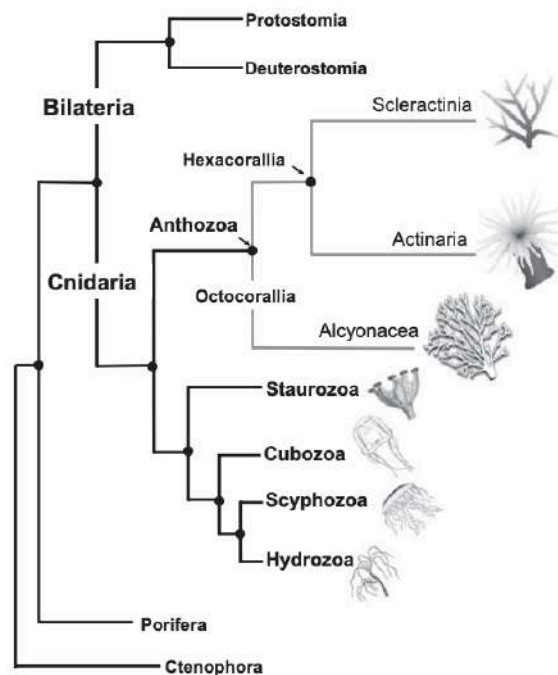


Figure 3: Depiction of the cnidarian phylogenetic tree. Extracted from Palmer & Traylor-Knowles (2018)

Cnidarians are relatively simple animals composed of two primary cell layers (epidermis and endodermis) which are typically separated by a mesoglea, they contain cnidocytes or “stinging cells”, and have a rudimentary nervous system (Spalding et al., 2001). Most corals are colonial and are made up of distinct, genetically identical polyps. Polyps can be thought of as transparent sacs with a slit like mouth surrounded by a ring of tentacles (

Figure 4). Although made up of individual polyps, the colony functions as a whole as can be seen by linked nervous system responses of adjacent polyps and nutritional exchanges throughout the colony (Sorokin, 2013).

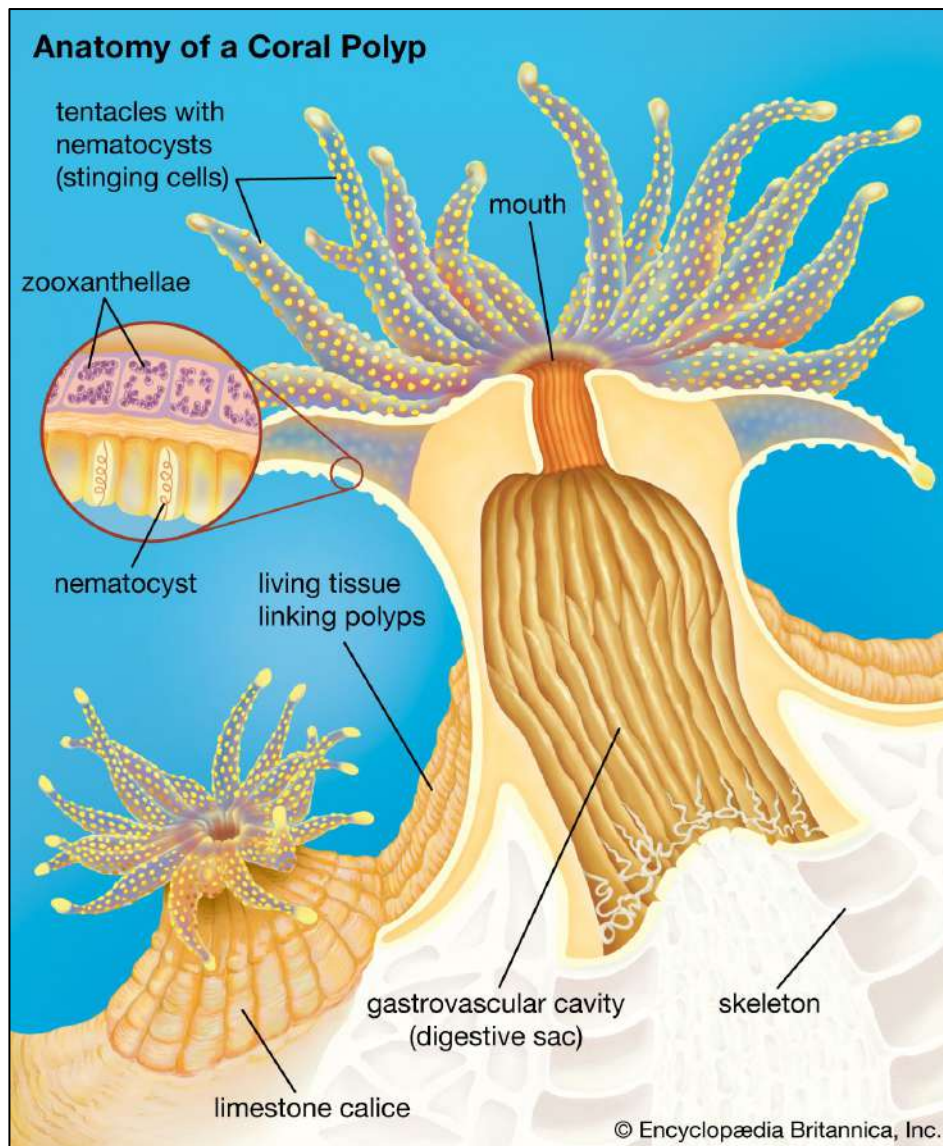


Figure 4: Depiction of the anatomy of a coral polyp (Encyclopedia Britannica, Inc.)

Clear waters and warm temperatures, two characteristics that make these ecosystems so appealing to travelers, are also a necessity for the survival of the coral animal and explain why the distribution of certain types of corals is limited to the tropics (Muir & Pichon, 2019). Although coral reefs are thriving ecosystems teeming with life, they are found in clear waters resulting from a lack of nutrients. This paradox was first noted by Charles Darwin (1842) and can be explained by the fact that the coral animal is nourished by a symbiotic relationship with a dinoflagellate algae, zooxanthellae, that lives in their endodermal tissue and provides them with up to 95% of their energy through photosynthesis in exchange for inorganic nutrients and

shelter (Campoy et al., 2020) (Dubinsky & Falkowski, 2011)(Figure 5). As photosynthesis is key to their energetic demands, this highlights the importance of sunlight and favors the tropics as a habitat. The increase in energy received from the zooxanthellae helps accelerate the calcification process and allows corals to deposit large amounts of calcium carbonate (CaCO_3) thus building large colonies and forming massive reef structures (Muir & Pichon, 2019) (Sorokin, 2013). This symbiotic algae is also responsible for giving coral colonies most of their color and thus the loss of these organisms leads to coral bleaching.

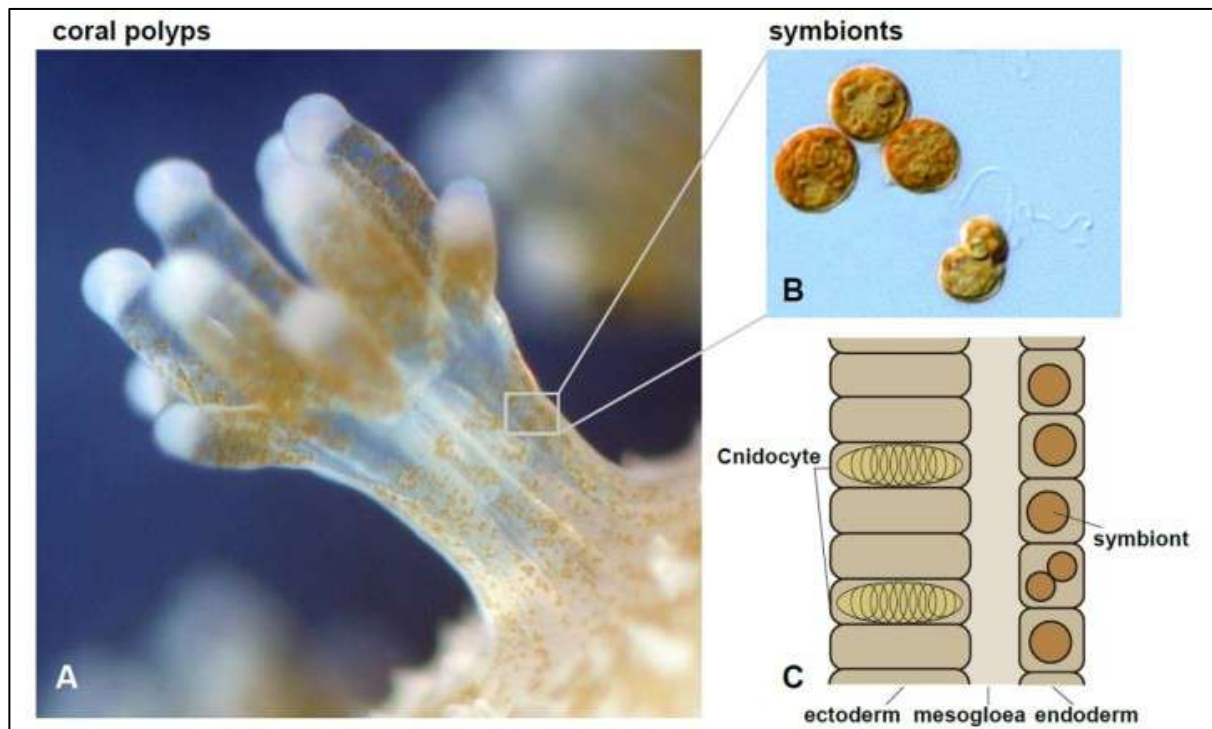


Figure 5: (A) Depiction of coral polyp with its symbionts (B) Microscopic view of symbionts found within the coral polyp (C) Drawing depicting the location of the symbionts in the endoderm (Okinawa Institute of Science and Technology Graduate University, 沖縄科学技術大学院大学).

The phenomenon referred to as “bleaching” occurs when the pigmented algae are no longer present in the animal tissue, exposing the white coral skeleton as the tissue is now transparent. Bleaching is a stress response resulting from numerous factors, most notably high sea surface temperatures but also UV irradiation, bacterial infection, lowered salinity, and pollution (Stat et al., 2006). Other aspects of coral biology such as settlement, growth, and reproduction are affected by environmental factors such as but not limited to temperature, light, water flow, depth, salinity, sediment, and substratum (Wells, 1957) (Figure 6). Below, I will briefly detail how various biotic and abiotic factors affect coral physiology.

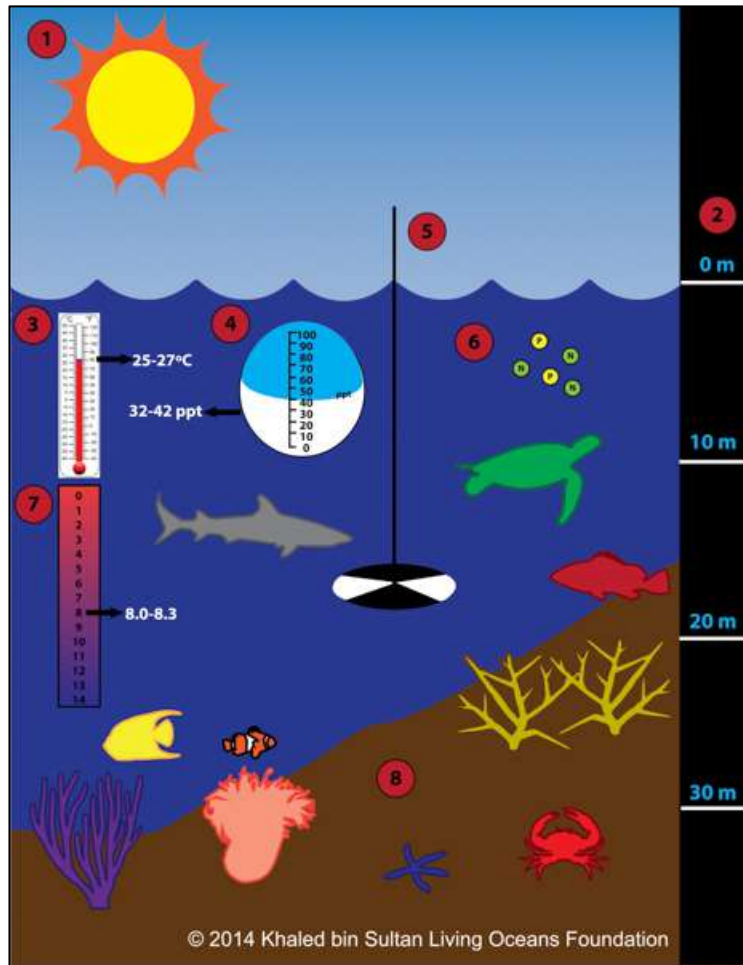


Figure 6: Diagram depicting different abiotic factors that affect coral reefs (1)Light (2)Depth (3)Water Temperature (4)Salinity (5)Turbidity (6)Nutrients (7)pH (8)Substrate (Khaled bin Sultan Living Oceans Foundation (2014))

1. Substratum: Hallock (2001) defines coral reefs as “rigid skeletal structures in which stony corals are major framework constituents”. Reefs are made up of calcium carbonate with corals being the dominant calcifying organisms on coral reefs (Fagerstrom, 1987). Other organisms such as coralline algae, oysters, and serpulid worms play a role in reef construction, and framework but corals are most notable for the fabrication and maintenance of the reef ecosystem (Hallock, 2001). Substratum is an essential component of reef bioconstruction with coastlines composed of fine mud or unstable sediment proving unsuitable for coral larval settlement (Spalding et al., 2001). Coral larvae rely on several cues such as chemicals, light, color, and sound for choosing settlement locations with physical cues such as surface complexity and roughness also used as indicators (Vermeij et al., 2010) (Mason et al., 2011) (Heyward & Negri, 1999) (Morse et al., 1988) (Praeger et al., 2012) (Whalan et al., 2015). This highlights the importance of the 3-D structural complexity provided by corals and other reef builders and helps explain why certain human activities such as dynamite fishing

which transforms stable, living coral colonies into unstable, shifting rubble fragments are so detrimental to coral reef ecosystems (this is discussed in more detail in Chapter 4). Additionally, one of the most well-known settlement inducing cues is the presence of Crustose Coralline Algae (CCA) (Tebben et al., 2015). Overall, substratum is a key factor in coral larval settlement and reef formation and changes in substratum can reflect on the overall status of the reef environment.

2. **Water flow:** Different hydro physical factors such as waves, tides, and currents affect reef biota, growth, and structure. Tides have been shown to have the greatest effect in terms of hydrodynamic stress and can greatly impact reef formation and distribution as well as the function of reef benthos (Sorokin, 2013). For instance, wave action has been shown to influence diversity with exposed sites showing the highest biodiversity and the lowest coral cover (Huston, 1985). Waterflow, in terms of cold or warm water upwellings, can affect reef development and larval dispersal while also playing a role in prey capture with decreases in zooplankton capture rates at low flow speeds (Spalding et al., 2001) (Sebens et al., 1998). Furthermore, it has been shown that corals experiencing higher waterflow may recover better from bleaching than colonies in areas of low water flow (Finelli et al., 2006). Thus, waterflow proves to be an important factor in the functioning of the reef ecosystem.
3. **Salinity:** Corals are osmo-conformers meaning that their internal cellular environment is not constant and changes in conjunction with their external environment *i.e.*, they gain water in hypotonic conditions and lose water in hypertonic conditions (True, 2012). Even though they are often exposed to fluctuations in salinity levels through freshwater runoff, precipitation, and desalination during droughts, corals are believed to have a limited ability to adjust to these changes (Röthig et al., 2016). The normal marine environment salinity ranges from 33-37PSU (Practical Salinity Unit) with prolonged exposure below 30PSU or above 40PSU negatively affecting coral functioning (Hallock, 2001). Reduced salinity, for example, has been shown to lead to reduced fertilization, larval survival and development as well as increased signs of stress such as more mucus production and decreased feeding activity (True, 2012). Thus, salinity is an important environmental factor in the survival of the coral animal.
4. **Temperature:** Tropical corals are typically found in latitudes between 30°N and 30°S as most reef corals require temperature between 16-30°C (Coles & Riegl, 2013) (Spalding et al., 2001). Their optimum temperature ranges from 23-29°C with many

species dying when temperatures reach below 14°C or above 31°C (Hallock, 2001). Increased temperature has been linked to coral bleaching, reduced reproductive potential, and increased incidence of disease (McWilliams et al., 2005) (Lesser, 1997) (Jokiel & Guinther, 1978) (Miller & Richardson, 2015). Although, it has been shown that corals who experience temperature variability or are more frequently exposed to warming events may prove to be more resilient in future warming scenarios (Carilli et al., 2012). Temperature has proven to be a key factor in coral survival and with the current global warming predictions, effects of temperature change on marine animals, such as corals, has become a major concern.

5. Light & Depth: As discussed above, the coral symbiont, zooxanthellae, requires light for photosynthesis thus limiting the depth range for zooxanthellate corals. Therefore, most reef building corals are found at less than 100m depth within the photic zone with most colonies found from the surface to 30m. It has been shown that coral species diversity increases with depth up to 30m and drops off thereafter with diversity trends at the surface being more variable due to increased incidence of disturbance events (Huston, 1985). Light and depth also play a role in coral morphological presentation with deeper colonies tending towards flat horizontal shelves allowing for more exposed surface area for light capture while shallower zones tend to favor branching colonies (Huston, 1975). Light attenuation can also be affected by other factors such as sediments suspended in the water column or algal blooms which limit light penetration thereby reducing or preventing coral growth (Hallock, 2001) (Spalding et al., 2001). As photosynthesis via their endosymbiont is largely responsible for the energetic needs of corals, light and therefore depth as well are important to survival.
6. Nutrients: As mentioned above, coral reefs are typically found in warm, oligotrophic waters. Excess nutrients can be detrimental to reef accretion by reducing water transparency, inhibiting calcification, and increasing rates of bioerosion (Hallock, 2001). Eutrophication can often lead to increased macroalgal growth which then compete with corals for limited benthic space (Larned, 1998). Increased algal growth can lead to a macroalgal dominant state which can greatly change reef functioning. Typically, reefs shifting from coral dominant to algal dominant are viewed as less healthy (Kramer, 2003)

The Coral Holobiont

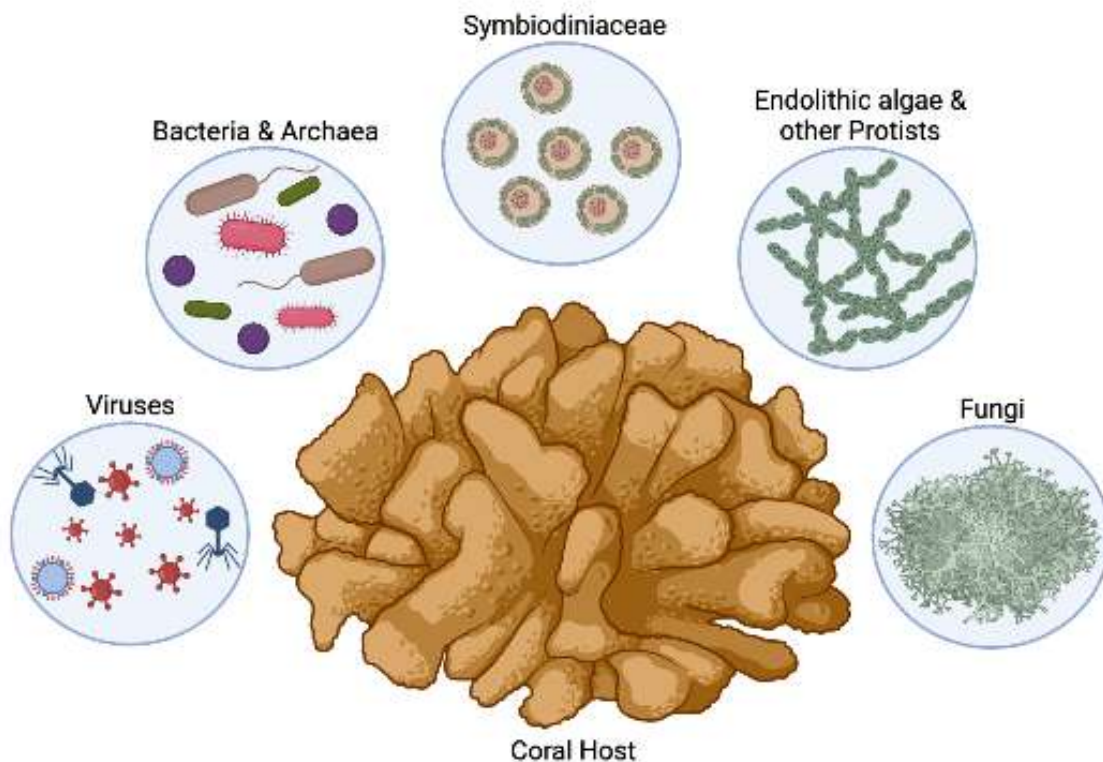


Figure 7: Depiction of the Coral Holobiont showing the coral host with its associated microbiome (inspired by Pogoreutz et al., 2020)

1.2.2 Microbiota

Another important aspect of coral biology is the coral holobiont which refers to not only the coral host but also its symbiotic algae (zooxanthellae) and its internal and external microbiota consisting of bacteria, archaea, viruses, and fungi (Figure 7). Corals have one of the most phylogenetically diverse microbiomes of all animals with the coral-algal symbiosis being the most well studied part of the coral holobiont (Huggett & Apprill, 2019). The composition of microorganism's changes within a single colony, with different assemblages found in the tissue, mucus and skeleton (Rosenberg et al., 2007) (Toledo-Hernández & Ruiz-Diaz, 2014). The coral holobiont plays an important role in the survival and resilience of corals. For example, certain bacterial species are involved in the exchange of essential metabolites, nutrient recycling, and protection from pathogens (Mohamed et al., 2023). Inversely, microorganisms can cause disease and bleaching under certain environmental conditions (Rosenberg et al., 2007). When exposed to stressors such as increased temperature, reduced pH, or elevated nutrient levels, scientists recorded an increase in virulence genes and a shift in the composition of the coral microbiota to one resembling a diseased colony (Thurber et al., 2009). However, research has shown that corals can alter the composition of their holobiont, notably the

photosymbiont, to adapt to changing environmental conditions leading to the coral probiotic hypothesis which states that a dynamic relationship exists between symbiotic microorganisms and environmental conditions, resulting in selection of the most advantageous coral holobiont (Reshef et al., 2006). Thus, allowing for a certain resilience and adaptability in coral populations. Therefore, when evaluating coral health, the coral holobiont must be considered.



Figure 8: Coral colony photographed at night showing individual polyps with tentacles extended ready to feed

1.2.3 Predator Prey Interactions

Although most of their energy comes from their symbiotic algae, corals are carnivorous and feed on tiny zooplankton in the water column. Most species feed at night and certain species only open their tentacles after sunset with one study showing that individual polyps can ingest .5 to 2 prey items per hour (Sebens et al., 1996) (Houlbreque & Ferrier-Pagès, 2008) (Figure 8). Corals rely on what they ingest (prey items, dissolved matter or particulate compounds) to supplement certain nutrients such as nitrogen and phosphorus which are not adequately supplied by their zooxanthellae (Goldberg, 2018). This food capture accounts for roughly 10% of their daily energy needs and typically is used for reproduction and growth (Hallock, 2001). Studies suggest that colony morphology and polyp size does not affect feeding rates but rather that feeding is limited by effective surface area, capture mechanisms such as nematocytes and/or presence of tentacles, and feeding effort (Palardy et al., 2005) (Sebens et al., 1996) (Houlbreque & Ferrier-Pagès, 2008). Furthermore, prey selection is not affected by

species, depth, or season but feeding rate does increase in relation to zooplankton abundance, depth, and temperature (Houlbreque & Ferrier-Pagès, 2008).

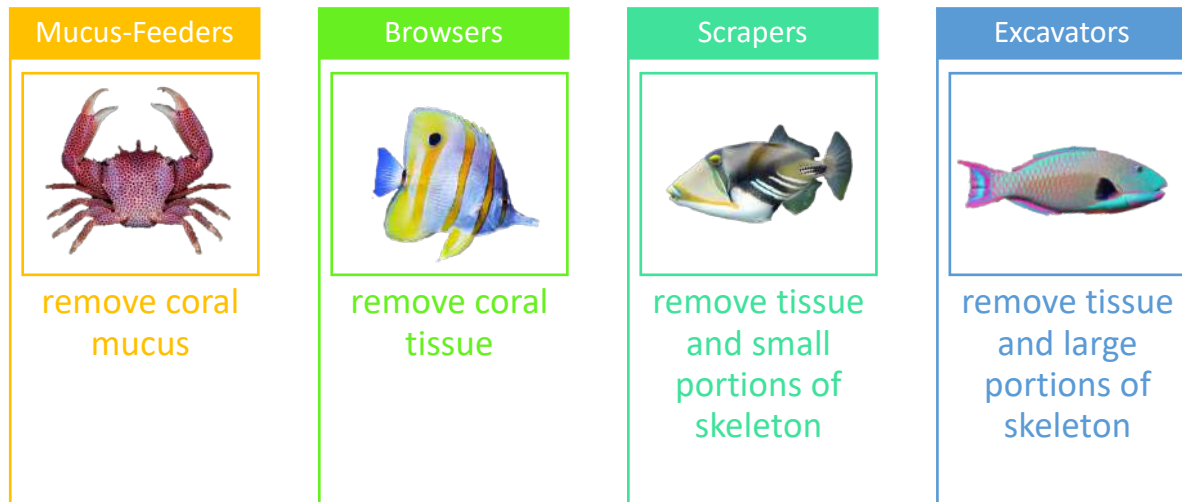


Figure 9: Examples of different types of corallivores (Mucus-Feeders, Browsers, Scrapers, Excavators) which feed on different parts of the coral animal such as the mucus, tissue and/or skeleton Photos from left to right: Crab Database, Encyclopedia Britannica, Marinelifephotograph, Egypt Divers,

In terms of predators, animals that eat coral are referred to as corallivores and they are an important part of coral reef ecosystems. Corallivores can change reef structure and community composition and in certain cases can greatly threaten reef functioning (Al-Horani et al., 2011). Some of the main corallivores are fish, gastropods, and echinoderms with *Acanthastar planci* (L.), or the crown of thorns starfish (COTS), being one of the most infamous coral predators. Corallivores can be facultative or obligate and may consume the coral skeleton, tissue and/or mucus. Corallivory can inhibit growth, sexual reproduction, and cause widespread mortality with corallivores also being linked to disease transmission (Rice et al., 2019) (Nicolet et al., 2018). In recent years, there have been increases in COTS outbreaks which may be due to anthropogenic influences such as overfishing and increased nutrient runoff (Miththapala, 2008). COTS preferentially feed on tabular corals such as *Acropora sp.* but during an outbreak they become less selective and target most species with instances of outbreaks recorded that have decimated corals and reduced coral cover to less than 1% (Harriott et al., 2006) (Figure 10A). In terms of fish, species of butterflyfish (family Chaetodontidae) are the most prominent coral predators accounting for over half of all 128 species of corallivorous reef fish with other notable families including Labridae, Tetraodontidae, Monacathidae, Pomacentridae, Scaridae and Balistidae (Cole et al., 2008). In the Indo-Pacific, the Bumphead parrotfish (*Bolbometopon muricatum*, family scaridae) is an important corallivore and has been shown to influence coral growth, mortality, and reproductive fitness (Rotjan & Lewis, 2008). Cole and colleagues (2011)

found that the quantity of tissue removed by corallivorous fish was much higher than original estimates indicating high energetic costs for the coral in terms of wound healing or repair (Cole et al., 2011) (Figure 10). Likewise, gastropods such as the corallivorous snail, *Corallophila violace*, were believed to have little effect on their coral host but Clements & Hay (2018) showed that *Corallophila* predation could reduce colony growth by 18-43%. Similarly, gastropods of the genus *Drupella* have been shown to have effects akin to COTS in terms of coral mortality (Turner, 1994) (Figure 10). In the eastern Pacific, three corallivores (Jenner's Pustulate cowry, pufferfish, and hermit crabs) were estimated to cause almost 7 tons of coral erosion per hectare per year equating to 1/3 of the annual production (Glynn, 1974). Thus, showing that corallivory can not only be an important factor in evaluating the fitness of a coral colony but also in the overall dynamics of the reef ecosystem.

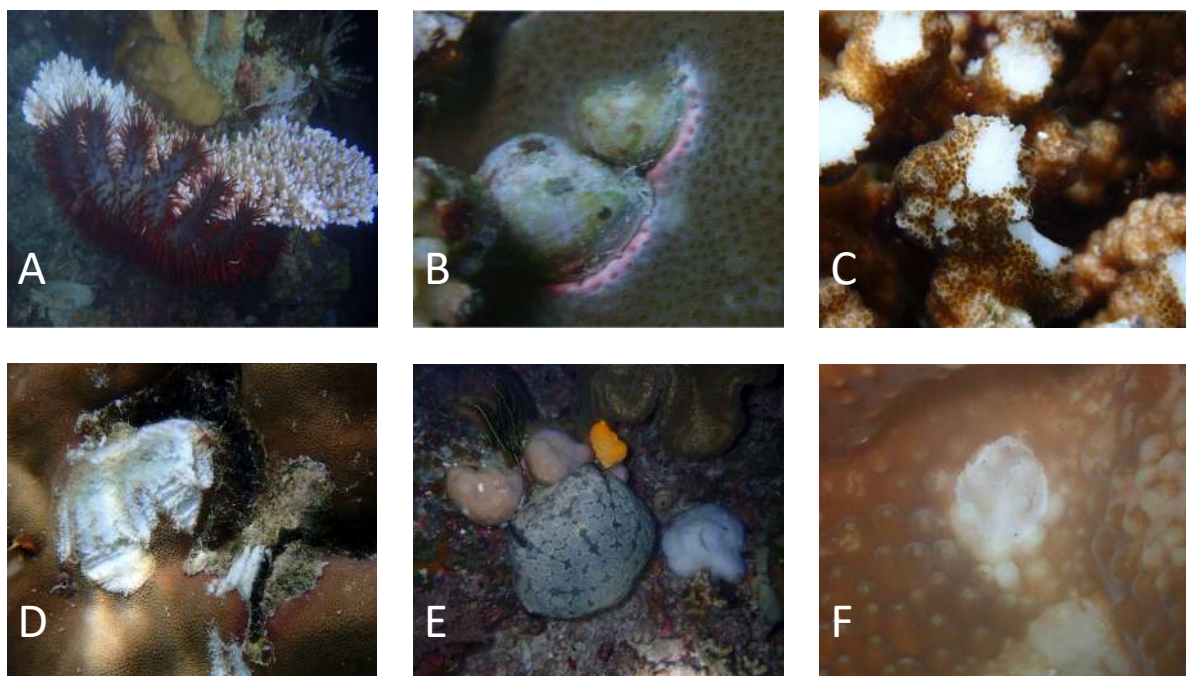


Figure 10: Examples of different coral predation scars and predators (A)COTS actively feeding on an *Acropora* colony (B)Snails actively feeding on a *Porites* colony (C)Fish predation marks on a *Pocillopora* colony (D) Parrotfish predation scars on a *Porites* colony (E) Sea star actively feeding on a *Porites* colony (F) Snail predation scar

Being sessile, corals cannot escape from predators, storms, anthropogenic disturbances, or environmental stressors meaning that their ability to respond to injury and mount an immune response is critical to their survival. Lacking the adaptive immune system of vertebrates, corals rely on innate immunity with physical barriers, such as coral mucus, acting as one of the first lines of defense (Cooper, 2010) (Toledo-Hernández & Ruiz-Diaz, 2014). Coral mucus is multifunctional and can be involved in alimentation, solar protection, and pathogen defense (Toledo-Hernández & Ruiz-Diaz, 2014). When mounting an immune response, corals maintain

microbial homeostasis and prevent bacterial infection through increased expression of immune relevant genes (van de Water et al., 2015). If the coral receives an injury, wound healing occurs which has been summarized in four main phases: 1. degranulation and plug formation, 2. infiltration of immune cells, 3. cell proliferation, 4. wound maturation (Palmer and colleagues, 2011). A common wound inflicted on corals is lesions resulting from predation as mentioned above (Figure 10). Tissue regeneration is dependent on several factors including species, depth, colony size, size of wound, type of wound (*i.e.*, tissue damage only vs. tissue and skeletal damage) and environmental factors such as sedimentation and water flow (Counsell et al., 2019) (Cróquer et al., 2002) (Traylor-Knowles, 2016) (Sabine et al., 2015). Meesters and colleagues (1994) found that lesion repair was more dependent on the amount of tissue bordering the lesion than the size of the lesion with larger lesions often showing incomplete recovery.

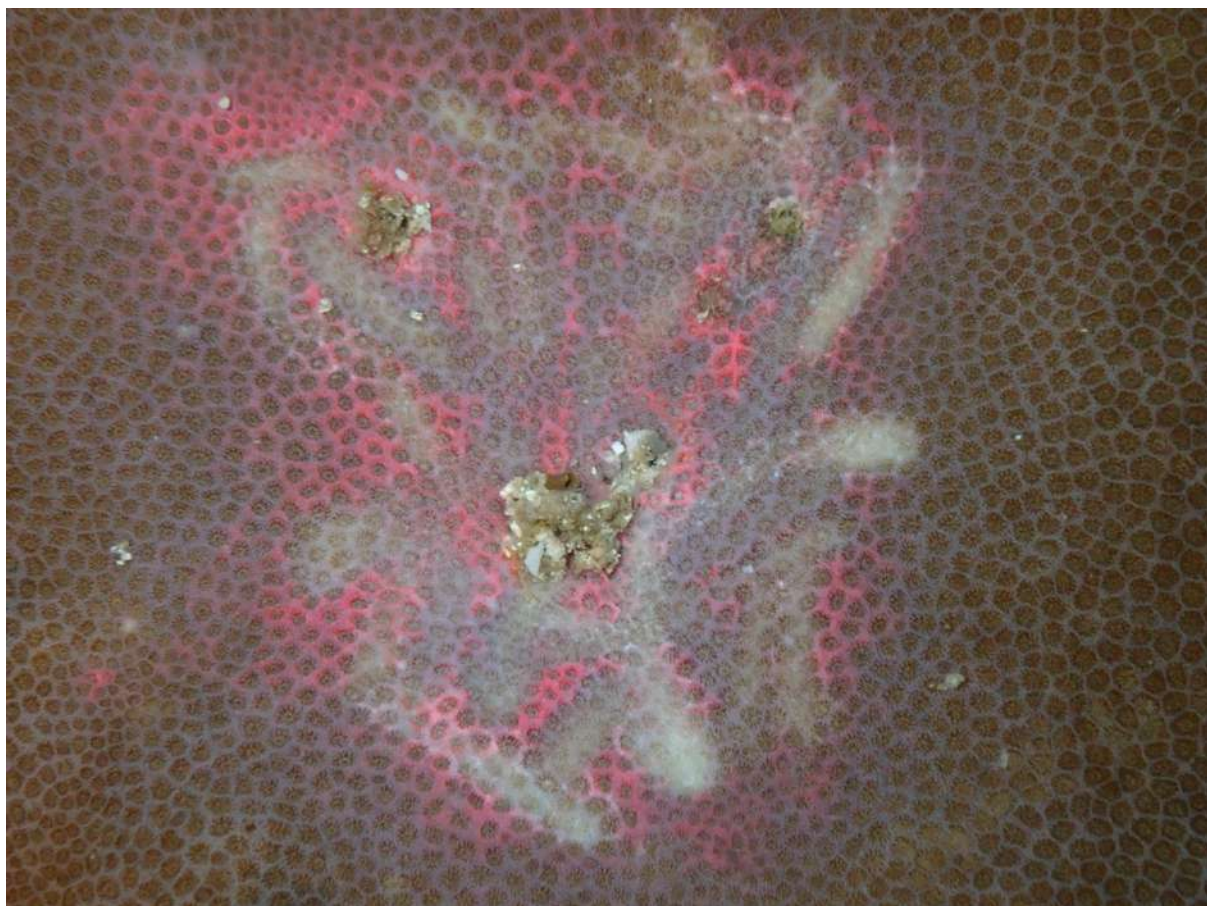


Figure 11: Pigmentation response(pink) displayed by a *Porites* colony

Another visual immune response mounted by the coral is a pigmentation response which has been described as an inflammatory immune response (Palmer et al., 2008). Originally, this pigmentation response was thought to be a disease called pink line syndrome (PLS) but now is seen as the colony's own immune response which may offer a defensive function against stressors (Kubomura et al., 2021). A pigmentation response usually displays as pink or purple

tissue and is often seen around wounds such as fish bites, or encircling holes created by boring organisms and/or delineating competition boundaries between species (personal observation) (Figure 11). In *Porites sp.*, pink tissue pigmentation can sometimes be the result of infection by a parasitic trematode which has been shown to greatly reduce colony growth (Aeby, 2003) (Figure 12). When evaluating coral colonies, visually examining the tissue for wounds/lesions and/or a pigmentation response may give a first idea on whether the coral is initiating an immune response and can help to evaluate the current state of a colony or indicate potential stressors.



Figure 12: Photograph showing infection by a parasitic trematode in a Porites colony as shown by the raised pink bumps in the center of the photograph

At first glance, corals appear to be relatively simple but end up being a bit of a conundrum. An animal that looks like a rock but has plant-like characteristics and requires precise environmental conditions. As discussed above, many complex interactions are happening not only within the coral animal (i.e., the coral holobiont) but also within its environment (abiotic factors such as temperature, depth, salinity, and biotic factors such as predator/prey interactions). All these interactions play a role in the survival of the coral animal and the ecosystem functioning. Given the complexity of these multiple interactions, how does one evaluate the health of an individual coral?

1.2.4 Towards Defining Coral Health

Defining health remains a complicated task even when it comes to humans. When defining human health, oftentimes health and its definition are linked to disease. For example, the world health organization (WHO) defines health as “a state of complete physical, mental and social

well-being and not merely the absence of disease or infirmity”. Likewise, Bircher (2005) jointly defined the terms as “health is a dynamic state of wellbeing characterized by a physical, mental and social potential, which satisfies the demands of a life commensurate with age, culture, and personal responsibility. If the potential is insufficient to satisfy these demands, the state is disease.” Both definitions reflect the fact that health is multifactorial and should consider more than just physical symptoms and the absence or presence of disease. As compared to the WHO definition, Bircher adds slightly more elements by considering age, culture, and personal responsibility. If we focus on the age component, she states that the past and future of an individual needs to be considered when evaluating health which includes not only genetics but also personal potential (Bircher, 2005). This idea highlights the importance of a temporal component when evaluating health. However, when contemplating the definition for health, Ereshefsky (2007) states that rather than debating the correct definitions of health and disease, we should use normative claims or “claims about states we value or disvalue” and state descriptions or “descriptions of physiological or psychological states”.

Therefore, by utilizing the above-described multifactorial definitions of human health and incorporating a temporal component we can move towards a more comprehensive definition of coral health all while emphasizing the use of careful language to specify evaluation criteria rather than using the vague blanket term of health. Likening coral health assessments to human health evaluations is not a novel idea and has been discussed at length in Downs and colleagues (2005) where researchers suggest likening reef health to clinical assessments, including both a historical review as well as a current examination. They highlight the necessity to monitor both the targeted biological system and known pressures (environmental and contaminant factors) in order to understand the causes and consequences of stressors and pressures and therefore combat them. In much of the literature one aspect of health, most typically bleaching or disease, is the sole criterion used to define a coral colony as healthy or unhealthy (Ruiz-Moreno et al., 2012) (Haapkylä et al., 2007) (Sweet & Bulling, 2017). This vagueness is also illustrated when defining reef health with studies listing improving reef health as their main goal while “health” remains undefined and unclear (as described in Downs et al., 2005). Thus, instead of referring to corals (and/or reefs) as “healthy” or “unhealthy”, state descriptions followed by normative claims should be used such as mentioned above by Ereshefsky (2007). For example, the coral colony is experiencing bleaching as a result of increased water temperature and this (bleached) state is viewed negatively by researchers and society. Such language would lead to clearer discussions about the current state of an organism or the reef as a whole.

Thus, we can use human models, of defining health, such as those defined above by the WHO and Bircher, as a basis to define the health of corals. However, in moving towards a definition of coral health, it is important to note that with corals being sessile animals, their inability to change or leave an inconducive environment must be considered more so than it would for a mobile organism such as humans. Accordingly, when evaluating coral health, it is necessary to evaluate the physiological state (physical), the environment (environmental), and interactions with other organisms (social) while also considering past and future experiences (temporal) such as genetics or adaptability (personal potential). To illustrate this point, corals are temperature sensitive, and their natural habitat may play a role in how the individual colony responds to change and/or how it could respond to future change. For example, it has been documented that corals in the Persian Gulf, an area that experiences the largest temperature variation of reef coral regions, can withstand temperatures of up to 36°C (Coles & Riegl, 2013). Thus, for these corals their environment has altered the expected ability of these organisms to respond to stressors and their state may have changed from one timepoint to another. This is confirmed by Baker & Colleagues (2004) who showed that corals can adapt in response to higher temperature by shifting their symbiont community composition across timepoints. Therefore, highlighting the importance of including and evaluating environment and temporal components when determining coral health.

Overall, coral reefs are dynamic environments which experience different stressors and environmental conditions across time. Moreover, the coral animal is influenced by the abiotic and biotic factors of its surroundings, as demonstrated above. All of which should be considered when evaluating the current state of the animal through the use of precise vocabulary as opposed to the blanket terms of unhealthy and healthy.

1.3 Technological Advances

Technological advances have revolutionized science across all domains but especially in the field of marine biology. This is specifically relevant and visible in terms of this research project through advances in scuba diving, camera technologies and computer science.

1.3.1 Scuba Diving

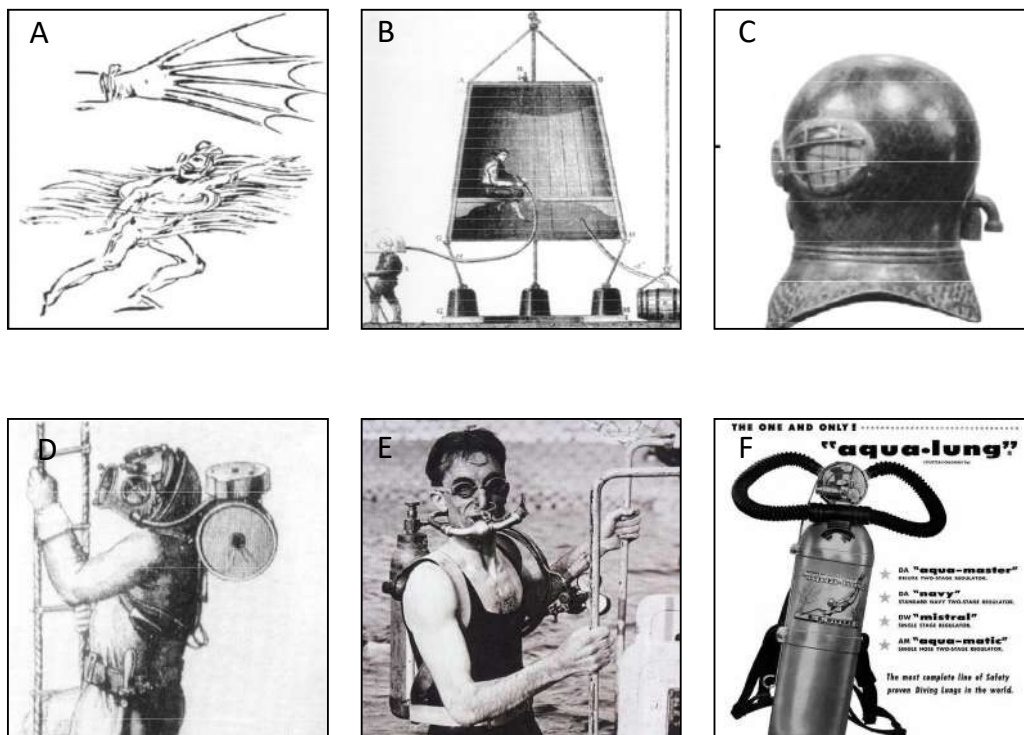


Figure 13: Compressed history of Scuba diving through photographs and drawings (A) Drawing of a diver by Leonardo Da Vinci (B) Depiction of the diving bell (C) Drawing of a diving helmet (D) Rouquayrol-Deanyrouse diving apparatus (E) Aqualung (Images A-D adapted from (Zurcher, 2002), Image E source: allthingsdiving.com Image F source: Diversdirect.com)

Exploring the underwater kingdom has captured imagination since ancient times and across disciplines. The ancient coastal communities of Greece, Mesopotamia and China used freediving as a means of food gathering, commerce and warfare, philosophers as early as Pliny and Aristotle described heroes exploring the oceans, and a drawing from the famed Leonardo Da Vinci depicted the first scuba diver long before scuba diving was possible (Chateaufreynaud & Lapierre, 2019) (Kocak & Caimi, 2005) (Zürcher, 2002) (Figure 13). In order to investigate this new realm, an understanding of pressurization and the effects of different gases on the human body was required, leading to numerous inventions spanning centuries such as the diving bell in 1692, the diving helmet in 1819, and finally the Rouquayrol-Deanyrouse diving apparatus in 1864 which was the first tank-based system and a precursor to modern diving (Zurcher, 2002) (Figure 13). SCUBA (Self-Contained Underwater Breathing Apparatus) was officially invented in 1936 by Commander Yves LePrieur and his equipment included a back mounted tank, a regulator, and a face mask (Rebikoff, 1967) (Figure 13). In 1943, the development of an autonomous diving system (named Aqualung) by Jacques Cousteau revolutionized scuba diving and made recreational diving a possibility through the simplicity

of this system and the technical support offered (Richardson, 1999)(Figure 13). Scuba diving was then popularized through various media outlets most notably Jacques Cousteau's film *Le Monde du Silence*, Jules Verne's *20,000 Leagues Under the Sea* and articles on underwater exploration featured in National Geographic (Zürcher, 2002) (Richardson, 1999). As a result of the development of safe and affordable diving equipment, easy access to remote scuba diving sites, and a growing interest in the natural world, scuba diving now plays a huge role in the multibillion-dollar marine tourism industry evidenced by more than 28 million people certified in 2021 compared to only 2.5 million in 1988 (Ong & Musa, 2012) (Musa & Dimmock, 2012) (PADI, 2021). Scuba diving has now become the most popular adventure sport in the world (Spalding et al., 2001).

Apart from tourism and recreation, modern day scuba diving is used by various industries such as the media, engineering, construction, archeology, and most notably science (Sayer & Barrington, 2005). Over 50% of scientific papers cited using scuba diving for sampling purposes with survey and quantitative observations accounting for approximately 25% of papers published; other uses include but are not limited to: in-situ measurements, impact assessments, ecological studies, deployments and retrieval, tag/recapture, biotechnology and/or pharmacology (Sayer, 2007).

In certain instances, scuba diving is the only means available for data collection, providing a non-destructive, cost-effective sampling approach which can produce samples of higher quality than those collected by other means (Sayer, 2007). Scuba diving transformed marine science by providing direct access to previously unexplored underwater habitats which allowed scientists to further their understanding of marine organisms and their environment (Witman et al., 2013). As stated in Lang et al. (2019) "The history of marine research has provided numerous examples of mysteries that would still be unsolved and findings that would have been misinterpreted with confidence if not for direct observation on scuba". Some concrete examples include life history strategies of marine invertebrates, discovery and identification of numerous species, reef fish ecology, and the field of marine chemical ecology (Lang et al., 2019). Additionally, data generated through scuba has been instrumental in conservation efforts such as developing Marine Protected Areas (MPA's) and restricting destructive fishing gear (Witman et al., 2013). In terms of corals, scuba diving prompted the understanding of fundamental information about coral reef biology such as coral heterotrophy, taxonomy, morphology, reproductive biology, UV protection systems, & larval behavior as dispersal as

well as the coral-algal symbiosis (Sebens et al., 2012) (Veron, 1995). It is safe to say that much of our knowledge about the marine world would remain a mystery without the invention of scuba diving.

However, like any innovation scuba has its limitations. The main one being bottom time or time spent at the site of interest which is limited by air consumption and nitrogen saturation (Sebens et al., 2012). Furthermore, divers alone are only able to cover small areas of the reef for surveying and data collection, further limited by the skill and knowledge level of the researcher and offering no physical data for verification or reanalysis (Bejibom et al., 2015). One underwater data collection methodology that can help combat these limitations is photography.

1.3.2 Photography

Photography, which is derived from Greek and translates to writing with light, was first invented in 1839 even though humans have been attempting to produce photographs since ancient times as evidenced by instances of pinhole cameras cropping up throughout history (Kocak & Caimi, 2005). Much like with scuba diving, technological advances were necessary before humans were able to capture high quality images underwater (Figure 14). Important advancements in underwater photography include artificial light, strobe lights, super-wide-angle lenses, corrected lenses, novel batteries, camera housings and of course the invention of scuba diving itself (Rebikoff, 1967). Thus, even though our fascination with the oceans dates to early times, it wasn't until 1856 when William Thompson photographed seaweed and sand resulting in a flooded camera that we had the first underwater photo (Pateman, 2009).

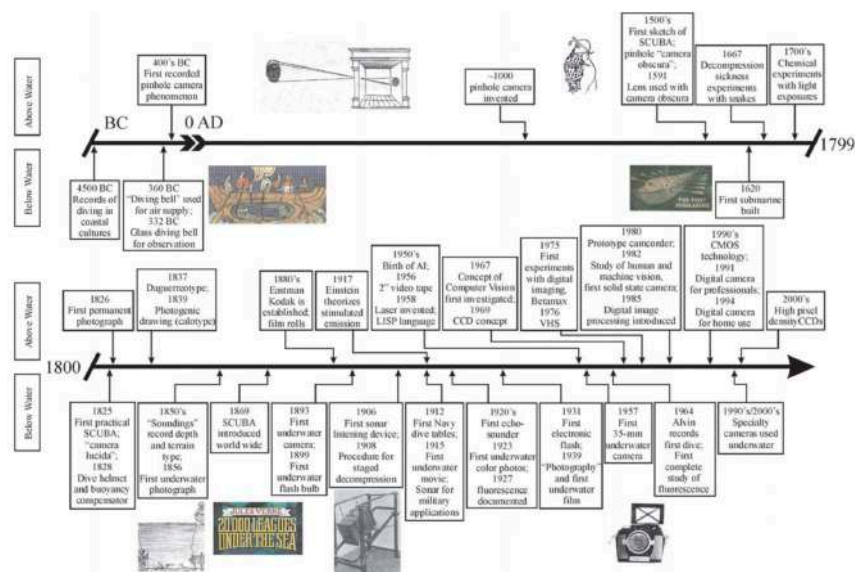


Figure 14: Timeline following the history of underwater imaging (Kocak & Caimi, 2005)

Since William Thompson's first photo, humans have made huge strides in underwater photography; particularly, with the invention of the first underwater water camera approximately 40 years later (Kocak & Caimi, 2005). Nowadays, cameras of higher quality and improved resolution are available at low costs which has increased the accessibility of photography as a methodology for marine research (Bejibom et al., 2015) (Gómez-Rios et al., 2018) (Page et al., 2016). Photographical monitoring provides numerous benefits such as increased sampling speed, non-destructiveness or invasiveness, the ability to repeat sampling of the same colony, the creation of an instant permanent record, and the option to revisit the image at a later date (Todd, Sanderson & Chou, 2001). Thus, photographic monitoring is becoming more commonplace with more and more images being produced (Page et al., 2016) (González-Rivero et al., 2020).

Using photography as a means for underwater data collection saves time during the dive but increases post dive analysis time for image annotation (Hopkinson et al., 2020). With the invention of digital cameras and sizeable memory cards, researchers are now able to capture hundreds of photographs in a single dive. Thus, image acquisition has far outpaced manual annotation and important data is lost as many images remain unannotated (Bejibom et al., 2012). In fact, less than 2% of images are annotated since they require an expert opinion and annotation is labor intensive and time consuming (Mahmood et al., 2016). This is specifically relevant in terms of coral classification, where an expert opinion is essential due to both highly variable morphology and the fact that coral nomenclature is not yet stable with new species being discovered and reclassifications occurring at the family, order, and genus level (Gómez-Rios et al., 2018). Furthermore, errors in terms of consistency and objectivity are two common problems with manual annotations (Bejibom et al., 2012). This is especially apparent when classifying corals based on color in order to assess for bleaching or paling. Not only do individuals see color differently but underwater photographic datasets tend to have variable lighting and color differences associated with changes in depth. All these factors lead to annotator bias and make comparisons difficult (Winters et al., 2009). Thus, underwater photography offers a promising means for speeding up data collection but requires another process such as artificial intelligence tools to speed up photographic annotations.

1.3.3 Artificial Intelligence

Artificial Intelligence or “machines that can think” is another theme sprinkled throughout history and can be found in writings by Homer, Descartes, and Verne and in celebrated works of fiction such as Frankenstein or the wizard of Oz (Buchanan, 2006). The origin of AI can be

traced back to “automata” in Greek mythology which was an artificial intelligence used by the gods for avoiding burdensome tasks whereas the term Artificial Intelligence itself was first coined in 1956 by MIT professor John McCarthy (McCroduct et al., 1977). Some early advances that were important for the development of AI were the Turing Machine famed for cracking the Enigma code in the second world war and ELIZA, a natural language processing tool able to converse with humans (Haenlein & Kaplan, 2019a). Modern examples of AI include Tesla’s self-driving cars, virtual assistants capable of voice recognition such as Siri or Alexa, and automated facial recognition in Facebook images (Kaplan & Haenlein, 2019b). Today, AI can be defined as a “system’s ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation” (Kaplan & Haenlein, 2019a). AI mimics the functioning of the human brain and learns from examples to identify rules and patterns without explicit programming (Kaplan & Haenlein, 2019b). To understand intelligent thoughts and actions, AI uses computers as experimental devices and draws on a multitude of disciplines namely engineering, biology, psychology, mathematics, statistics, and linguistics (Buchanan, 2006). To achieve its potential, AI needed technological advancements such as improvements in memory and processors, and efficient operating systems and languages (Buchanan, 2006). As with photography, this technology is quickly developing resulting in decreased costs and new applications (Kocak & Caimi, 2005).

1.3.3.1 Machine Learning

Machine learning is a subset of AI that focuses on pattern recognition and learning from data and can be divided into supervised and unsupervised learning (Yu et al., 2018). Unsupervised learning aims to find groupings and patterns and does not rely on human produced inputs or labels whereas supervised learning uses labeled training data to train the computer on pre-defined categories (Chirayath & Instrella, 2019). For example, exploiting human assigned labels to identify a never before seen image as a cat or a dog would be an example of supervised learning whereas the speech recognition used by Siri or Alexa would be unsupervised (Haenlein & Kaplan, 2019b). Supervised learning whether deep or not is the most common form of machine learning (Lecun et al., 2015).

1.3.3.2 Deep Learning

Deep learning is a subfield of machine learning where rules and algorithms are used to extract features which are then automatically learned in a hierarchical fashion by stacking multiple

layers, thus enabling the discrimination of more complex and abstract features (Rosebrock, 2017). Deep learning is a massive part of modern-day AI and has led to advances in object detection, motion tracking, action recognition, human pose estimation and semantic segmentation (Voulodimos et al., 2018) (Yu et al., 2018). In traditional methods, rules and algorithms were hand defined to extract features whereas in deep learning features are automatically learned through the training process (Rosebrock, 2017). This is a major advantage of deep learning (Lecun et al., 2015) but means that the mechanics behind outputs from deep networks are not fully understood and are often referred to as a black box (Haenlein & Kaplan, 2019a). Deep learning is leading to rapid progression in the AI field, not only has it beaten other AI techniques in image recognition, but it is also very good at discovering patterns in high dimensional data making it applicable to multiple domains (LeCun et al., 2015).

1.3.3.3 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are an important class of machine learning algorithms made up of multiple layers of processing that involve linear and nonlinear operators (Mahmood et al, 2016). Typically used with images, CNNs are deep neural network-based classifiers that use supervised learning techniques (Raphael et al., 2020) (O’Shea & Nash, 2015). Through the use of convolutional filters, they extract features from image data (Hopkinson et al., 2020). These filters are responsive to different features of the image with lower-level filters picking out simpler properties like edges and colors and higher layers picking out more complex features like shapes or textures (Raphael et al., 2020). These types of models require lots of training samples to achieve good generalization abilities (Gómez-Rios et al, 2018) but perform very well if are trained with large amounts of labeled data (Yu et al., 2018). Due to the success of CNN’s, most large technology companies like Google, Facebook, Microsoft and IBM initiated research and development projects to manipulate this technology (Lecun et al., 2015).

1.3.3.4 Artificial Intelligence in the Medical Field

Medicine has long been a promising field of application for AI with uses including interpretation of electrocardiograms, Disease diagnosis, treatment selection, surgical robots, and administrative tasks such as updating patient records and billing (Davenport & Kalakota, 2019) (Yu et al., 2018). One of the most successful applications of AI in the medical field is automatic medical image diagnosis which offers the benefits of consistent interpretation, high sensitivity and specificity, and instant results similar to those of physicians (Yu et al., 2018) (Gulshan et al., 2016). The use of deep learning in image analysis has allowed for the detection

of clinically relevant features that cannot be observed by the human eye, most commonly in oncology related images (Davenport & Kalakota, 2019). Essentially, the computer algorithm can recognize specific image characteristics or patterns that may not be apparent to a human (Vial et al., 2018). Similarly, deep learning has also been shown to extract new knowledge from raw data by unearthing unknown associations such as using retinal fundus images to accurately predict cardiovascular risk through associating patterns in the images with age, gender, systolic blood pressure, and smoking status (Poplin et al., 2018). Therefore, one can conclude that this technology might provide similar benefits and advances when applied to coral health.

1.3.3.5 Artificial Intelligence in Marine Research

The application of machine learning is constantly progressing but has not yet been fully utilized in marine research due to additional challenges posed by the underwater environment. Two major limiting factors in underwater image acquisition and annotation are variable lighting conditions and changes in water turbidity, both of which affect image colors and thus complicate differentiation and pattern recognition by a machine (Johnson-Roberson et al., 2006) (Bejbom et al., 2012) (Moniruzzaman et al., 2017). Lighting conditions, for example, can change due to environmental factors such as depth, time of day and water clarity which itself fluctuates depending on changing currents, algal blooms, and variable plankton densities (Purser et al., 2019). Further challenges are encountered when focusing specifically on coral reef datasets as they present a complex three-dimensional topography which has ambiguous and organic boundaries among marine species, as well as morphological variations between corals of the same species or similarities between different species in terms of size, color, shape, and texture (Moniruzzaman et al., 2017) (Stokes & Deane, 2009).

Even so, machine learning has already proven effective within the field of marine biology and is used extensively in satellite image analysis and plankton ecology (González-Rivero et al., 2016). Although its implication is relatively new to coral reef systems (González-Rivero et al., 2016) it has still been successful in terms of differentiating coral from non-coral (Tusa et al., 2014)(Mahmood et al.,2016)(Bejbom et al., 2012) (Greer & Benson, 2012), identification studies of fish and coral (Villon et al., 2016)(Stokes & Deane, 2009), benthic composition or percent coral cover (González-Rivero et al., 2014)(Chirayath & Instrella, 2019) (Hopkinson et al., 2020) (Purser et al., 2009)(Bejbom et al., 2012) (Zhang, 2015)(González-Rivero et al., 2020), size measurements (Neal et al., 2015)(Drap et al., 2013), and determining the aesthetic

value of coral reefs (Haas et al., 2015). AI technologies remain relatively unexplored in the examination of individual coral colonies. Due to its overwhelming success in the medical field, AI could prove very promising in applications relevant to coral health.

1.3.4 The Combination of Scuba Diving, Photography, and AI for Coral Reef Conservation

The combination of scuba diving, photography, and AI can be used to create useful tools for coral reef conservation. The use of underwater photography would speed up monitoring by allowing a diver to cover more area and survey more colonies during a dive while the use of machine learning would accelerate image analysis and provide a solution to the existing bottleneck of unannotated images while also reducing costs. González-Rivero et al. (2020) found that machine learning accelerated image analysis 200-fold at a fraction of the cost of manual annotations, with one expert annotated image costing \$5.41 compared to \$0.07 for a machine annotated image coupled with a reduction in ongoing costs. Furthermore, applications of automatic annotation have proven to be highly accurate (94-97%) when compared to expert annotations and would prove beneficial in several different monitoring schemes: bleaching monitoring, coral reef recovery, phase shift events, restoration projects and as a means to monitor the effectiveness of marine protected areas (Nunes et al., 2020). Finally, AI has shown great success in medical imaging and therefore would be interesting to explore in relation to the physiological states of individual corals.

1.4 PhD objectives

There is a negative correlation between technological advances and coral reef health as measured by reef degradation in terms of decreasing percent coral cover. As coral cover continues to drastically decline, technology is fast advancing. In order to capitalize on contemporary innovations, interdisciplinary studies are needed to accelerate scientific research, especially in a conservation context where time is a limiting factor.

My PhD research aims to combine the fields of Coral Biology, Computer Science and Marine Conservation with the **main question of my thesis being: how can artificial intelligence tools be used to assess coral health states from colony photographs?** Since the assessment of individual coral colony health state remains poorly defined, our approach is to use AI tools to assess visual cues such as physically damaging conditions (boring organisms & predation), contact with other organisms (algae, sediment) and color changes (bleaching, pigmentation) that could correlate with health states. This was achieved by utilizing photographic data from the Tara Pacific Expedition to build a first version of AI machines capable of automatically

recognizing these visual cues and then applying this tool to two types of field studies 1. A longitudinal study set up in Moorea, French Polynesia which aimed to investigate coral health as assessed by mortality/partial mortality events. 2. A comparative study between damaged, pristine, and restoration sites in the Raja Ampat, Indonesia. The objective of these studies is to extract the visual cues that distinguish healthy from unhealthy corals. A technological bottleneck of this approach is the ability to rapidly and uniformly annotate in situ photographs to build large datasets with sufficient high-quality examples for machine learning.

Study 1: An important achievement of my thesis was to create a dataset suitable to build first generation AI machines to automatically recognize coral colony visual cues (Chapter 3). To achieve this end, I first manually annotated the photographic data collected during the Tara Pacific Expedition (Chapter 2) which was then used as the foundation for my AI Machines. Utilizing the Tara Pacific photographic dataset in combination with corresponding manual annotations, individual machines were created to automatically annotate nine visual cues : bleaching, predation, pigmentation, algal contact, crustose coralline algae (CCA) contact, presence of boring organisms, tissue appearance, sediment contact, and turf overgrowth. This application helped create successful foundational machines highlighting the promise of AI tools in individual coral colony monitoring (Chapter 3). Furthermore, working with this dataset inspired methodologic changes to resolve some of the issues of the Tara Pacific dataset such as homogeneity and subjectiveness due to post-hoc analysis.

Study 2: I set up a longitudinal study in Moorea, French Polynesia to investigate changes in colony status over time (described briefly in Chapter 3). This study followed the methodology of the Tara Pacific Expedition and included physical as well as photographic sampling while adding a temporal component absent from the original expedition with monitoring and sampling to be conducted yearly. Through the addition of this temporal study, we can follow the survival of colonies over the long term. Then based on survival and mortality, we can investigate patterns from the visual cues measured throughout the years to discern if certain cues play a stronger or lesser role in predicting colony health as measured by survival. The photographs generated during this expedition will not only serve to augment our photographic corpus but can also be used, along with the in-situ annotations, to validate the original AI machines we have created (Chapter 3).

Study 3: In Raja Ampat, Indonesia, I launched a study utilizing a completely non-invasive sampling strategy. Here, four different types of sites (restoration, damaged, pristine, and

polluted) were studied to investigate how ecological changes or differences between sites, globally affect the patterns seen in the visual cues of individual colonies. This study will serve to help define the visual cues relevant to coral health. Additionally, in this study, we added a novel genus, *Acropora*, and methodological improvements such as the incorporation of strobes to combat color loss and the addition of more detailed individual annotations of visual cues to facilitate learning by the machine.

Thus, combining the foundational data from the Tara Pacific expedition with ecological and longitudinal data from two different research studies, I will present my research results and discuss how I achieved my research goal of creating a non-invasive tool that can automatically predict visual cues relevant to the health of an individual coral colony using only photographs.

1.5 Studied reef sites

Our initial research work was generated from the sites visited during the Tara Pacific Expedition which consisted of a set of 32 islands across the Pacific Ocean (Tara-Pacific Expedition). Two additional Research Studies were put in place in Moorea, French Polynesia and Raja Ampat, Indonesia.



Figure 15: Map showing the path of the Tara Pacific Expedition and the 32 islands visited (Planes et al., 2019)

1.5.1.1 The Tara Pacific Expedition

The Tara Pacific Expedition aimed to provide a baseline understanding of the coral holobiont and interaction with its environment across various levels (genes to organisms). This unique expedition was the first pan ecosystemic study of coral reef diversity covering an entire ocean

basin (Planes et al., 2019). The expedition involved over 70 scientists from 18 institutions visiting 30 countries and sailing over 100,000km while collecting nearly 60,000 samples (Planes et al., 2019) (Lombard et al., 2023). 32 islands were selected for sampling to include the full environmental range present in the Pacific Ocean such as the diversity gradient from east to west and the varying temperate regions. Environmental parameters were taken at each of the 32 islands along with the collection of physical samples of coral, fish and water. Additionally, photographs were taken for each organism sampled resulting in almost 12,000 photos. For corals, samples were taken for colonies resembling two scleractinian corals (*Pocillopora meandrina* and *Porites lobata*) and one hydrocoral (*Millepora platyphylla*). These three species were chosen for their wide distribution and abundance.

1.5.1.2 Indonesia

Indonesia has the highest percentage of coral reefs in the world with 18% while France and its territories is fourth on the list with 5% of the world's total at roughly 14,000km² (Sheppard et al., 2017).

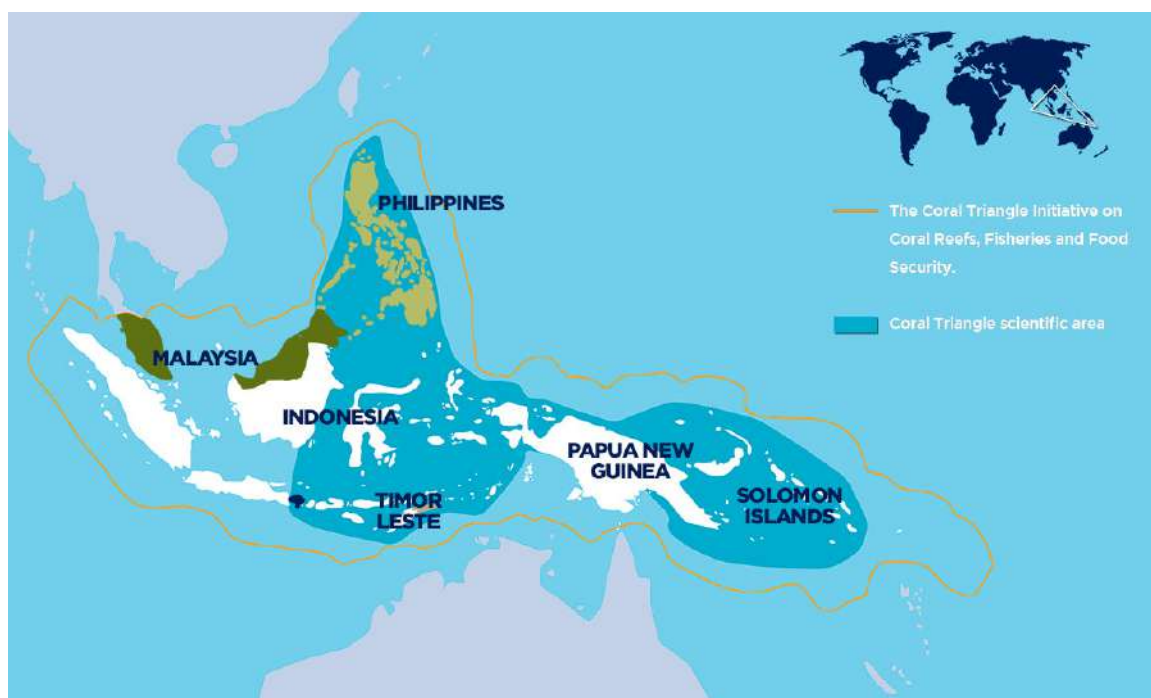


Figure 16: Map showing the coral triangle region (source: <https://www.balinecklaces.com/pages/the-coral-triangle>)

Indonesia is made up of over 17,000 islands and contains 50-100,000km² of coral covered area (Cesar et al., 1997). 25% of Indonesia's gross domestic products come from the marine and coastal resources and activities (Spalding et al., 2001). Coral reef ecosystems play an incredibly important role for the livelihood of the Indonesian population especially as 80% of Indonesians live on the coast with experts predicting this number to increase by more than 200% by 2060

(Neumann et al., 2015) (Habibi & Sartin, 2007). For example, the Bajau people in Wakitobi, Indonesia depend on coral reefs for coastal protection, food, livelihood, and construction materials (Crabbe, 2006). Coral reefs supply income to fishers and act as an important source of food security (Cesar, 1998). Indonesia is the eighth most fish-dependent country in the world with over 50% of its animal protein coming from fish (Abigail et al., 2018). Every year, coral reefs in Indonesia contribute 2.2 billion USD to the fisheries sector, 258 million USD to the tourism sector and 782 million USD to coastal protection highlighting Indonesia's dependence on coral reefs (Burke et al., 2011). Furthermore, shifting reefs to a healthy state by 2030 could add \$2.6 billion USD per year to the Indonesian economy (UN Environment et al., 2018)

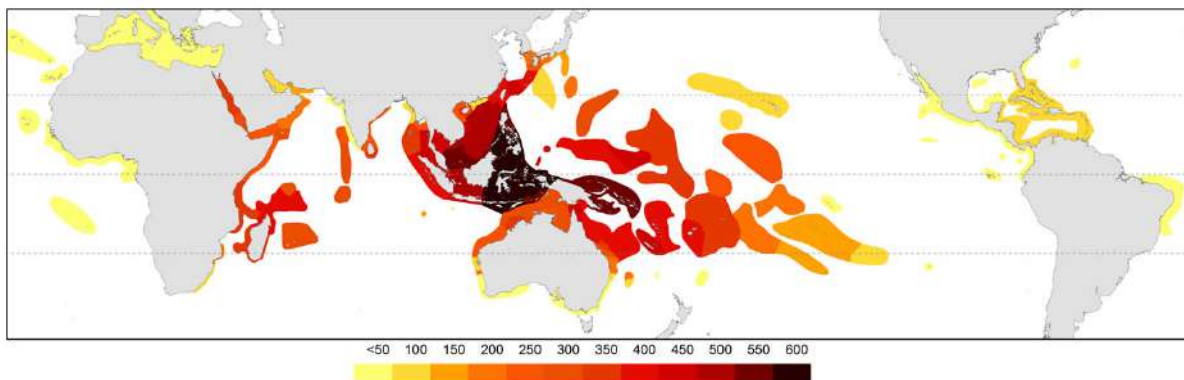


Figure 17: Map depicting global species richness of zooxanthellate corals (Adapted from Veron et al., 2009)

Furthermore, Indonesian coral reefs help make up the Coral Triangle (CT) region which hosts the highest marine biodiversity on the planet rivaling terrestrial rainforests. Within this region, the highest diversity can be found in the Birds Head Peninsula where a single reef can have four times the amount of zooxanthellate scleractinian species as those found in the entire Atlantic Ocean (Veron et al., 2009)**Error! Reference source not found..** An ecoregion of special importance within the Birds Head Peninsula is Raja Ampat. Here, we have the world's highest coral biodiversity with 553 coral species found in this area alone (Turuk and Souhoka, 2003). Moreover, this impressive biodiversity isn't limited to corals and can be seen in fishes, invertebrates and other important habitat forming biota such as seagrasses and mangroves (Sanciango et al., 2013). This rich biodiversity underlines the importance of this area for conservation efforts. To identify trends and develop comprehensive management practices, comprehensive and on-going data collection of coral reefs is necessary (Habibi & Sartin, 2007).

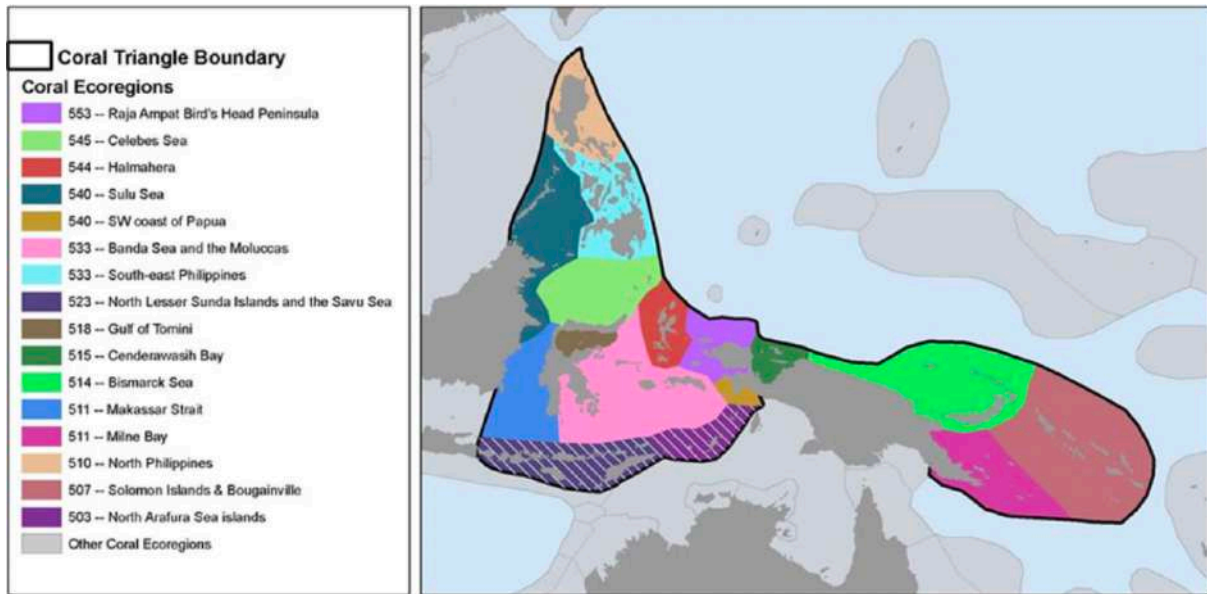


Figure 18: Map of species richness of zooxanthellate corals in different ecoregions of the coral triangle (Veron et al., 2009)

Despite its rich biodiversity and heavy dependence on coral reefs, coral reef monitoring in Indonesia remains limited with few reefs studied regularly and/or over the long-term nonetheless current data indicates that reef condition is declining with an increase in degraded reefs from 10% to 50% over the past fifty years (Burke & Selig, 2002). In 2019, 70% of Indonesian reefs were classified as poor (33.8%) and fair (37.4%) with less than 30% classified as excellent (6.4%) and good (22.4%) (Hadi et al., 2020). The main threats to Indonesia's reefs are overfishing and destructive fishing practice followed by coastal development and sedimentation from inland sources (Burke & Selig, 2002). Destructive fishing methods such as blast and cyanide fishing have been particularly detrimental to Indonesian Reefs (Figure 19). Cesar (1998) predicts that the economic losses from damaged coral reefs far outweigh (up to 50x) the monetary benefits individuals may attain from fishing or mining.



Figure 19: Photograph showing an Indonesian reef affected by dynamite fishing (Left) with a closeup of rubble (broken coral pieces substrate) (Right)

Overall, coral reefs provide an extremely important resource to Indonesia in terms of coastal protection, food, livelihood, and tourism with Indonesia hosting the most biodiverse reefs in the world. Given the current worldwide decline in the state of these ecosystems as well as the present degradation happening in Indonesia, it is of the utmost importance that reefs are monitored thoroughly and over the long-term to best conserve these vital ecosystems with areas of high biodiversity, such as Raja Ampat being of particular concern.

1.5.1.3 French Polynesia

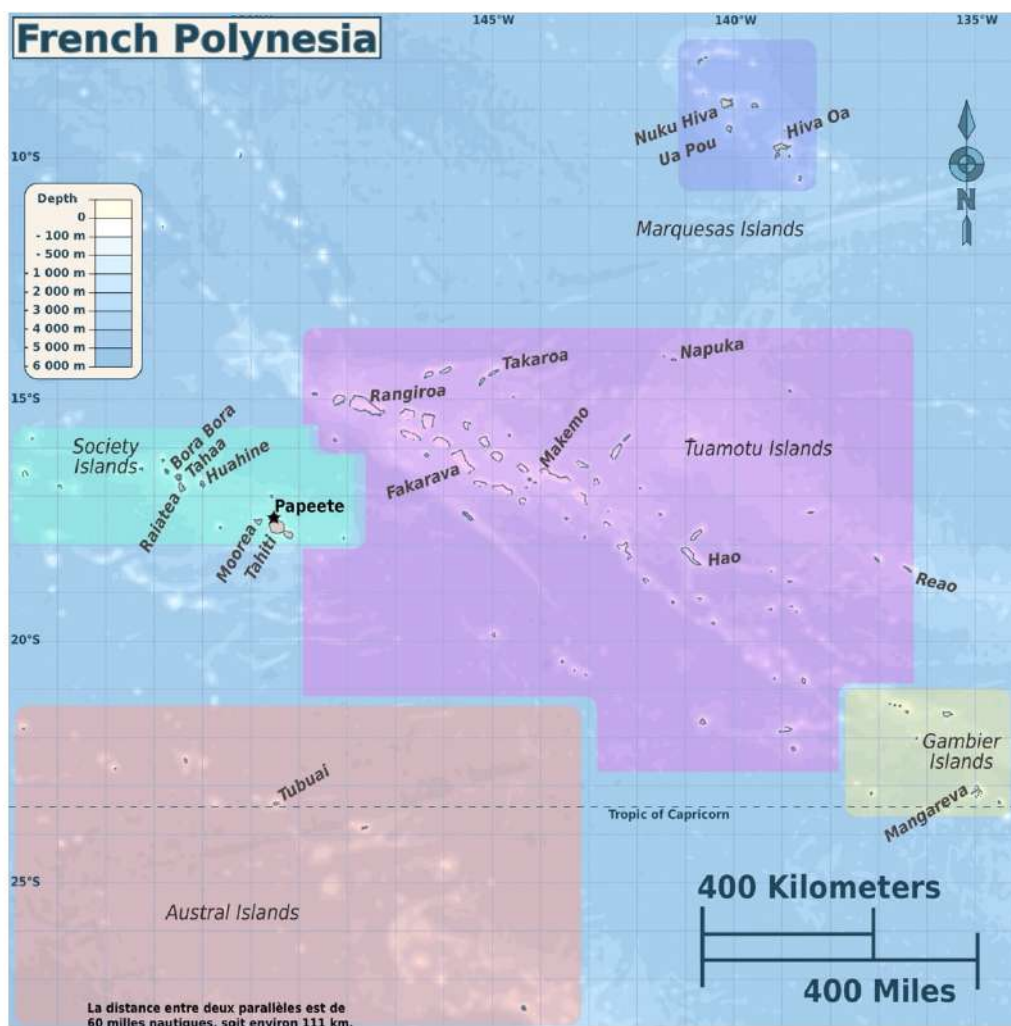


Figure 20: Map showing the French Polynesian islands and the regions they belong to (Source: mapsland.com)

French Polynesia, which became a French colony in 1840 and is now an overseas collectivity, is made up of 118 islands and is one of the largest territories of the Pacific and the largest French overseas area (Figure 20). Since the second world war the population has increased fivefold and continues to increase each year with essentially the entirety of the population centered along the coast and 75% of the population found on the Society Islands of Tahiti and Moorea (Salvat et al., 2008). French Polynesia has historically relied heavily on ocean

resources as can be seen by its two main exports being pearl products and fish accounting for 62% and 13% of exports respectively (Gonsard, 2016). FP has been harvesting pearls since the 1800's and is now one of the leading cultured pearl producing nations in the world supplying 98% of the world market of black pearls earning US\$130 million per year (Spalding et al., 2001) (Johnston et al., 2019). In terms of fisheries, FP exports around 14,000 tons of fish a year resulting in almost \$19,000,000 USD with tuna being the principal export (FAO, 2018). Between 2005 and 2015, exports of fisheries products increased by over 16% indicating the importance of this sector for the local economy (Gonsard, 2016). Furthermore, fish plays a vital role in the nutrition and diet of Polynesians. At the national level, French Polynesians get 65% of their animal derived protein from fish with this percentage reaching 71% in rural areas (Bell et al., 2009). Finally, coral reef related tourism has been a major revenue source for French Polynesia since the nineties and is currently the most important financial sector, bringing in 500 million euros in 2002 and accounting for 8 out of every 10 employees in 2015 (Wilkinson, 1996) (Salvat & Pailhe, 2002) (Gonsard, 2016). Coral reefs are one of the main draws for tourism with 66% of tourists snorkeling and 18% diving (Salvat & Pailhe, 2002).

Along with economic implications, the culture and lifestyle of French Polynesians is centered around the ocean and its resources (Salvat & Pailhe, 2002). Di Castri & Balaji (2002) stated "Polynesian culture is largely reef-oriented in terms of knowledge, traditions, and resources. Nowhere are culture and nature considering coral reefs so intimately associated". Historically, corals were used as construction materials as well as for tools such as pestles, graters, and fishhooks (Molle et al., 2023). Corals are revered within Polynesian cultural which can be traced back to their use in not only the construction of the "marae" or sacred places but also as religious offerings and as one of the materials for which to carve "puna" or magic stones (Molle et al., 2023). The importance of corals and ocean resources is further embedded within the traditional languages such as Tahitian which has the highest number of vernacular names for coral reef habitats and associated flora and fauna of any language in the world (Molle et al., 2023).

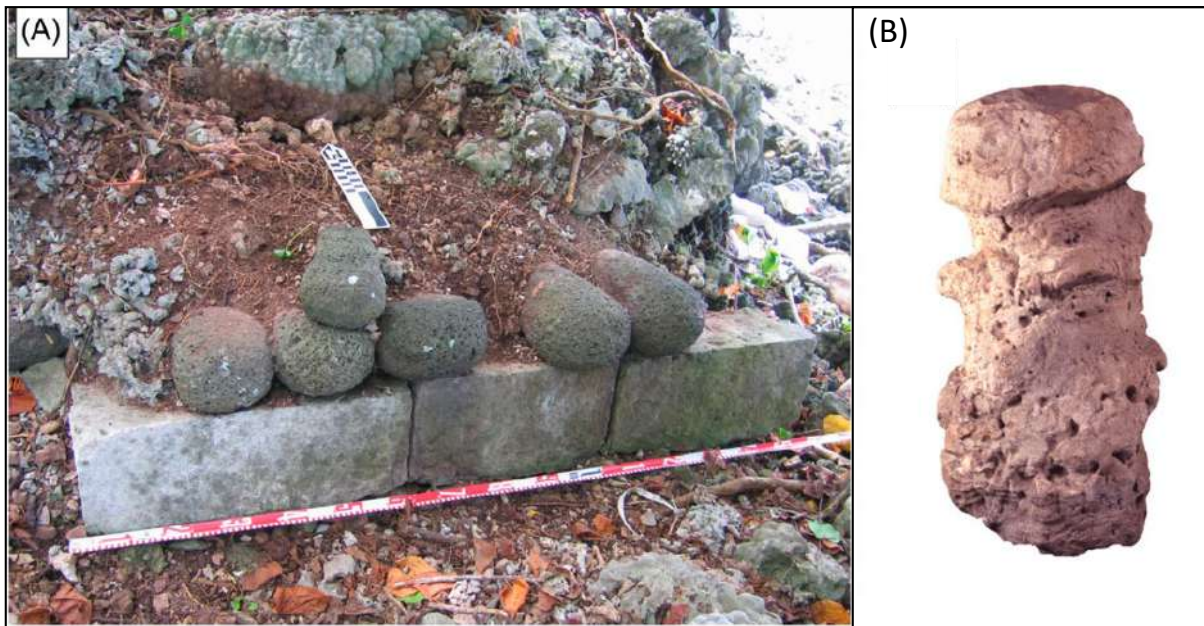


Figure 21: The cultural importance of coral in polynesian culture as depicted by a (A) *Porites* blocks used in a Marae in Moorea (B) A Puna carved out of coral Adapted from Molle et al., 2023

Not only is French Polynesia heavily reliant on ocean resources as indicated by their dependence on pearl production, fisheries, and tourism but they also have a rich cultural history that is centered around the reef thus marking the need to conserve and protect these important resources. Furthermore, the effects of climate change could render low lying islands such as French Polynesia uninhabitable, especially without the coastal protection offered by corals (Wilkinson, 1996). Due to the composition of its flora and fauna reefs in FP tend to be less colorful with less movement than their eastern or Caribbean counterparts as certain taxonomic groups such as gorgonians are absent and others such as soft corals are less abundant (Salvat et al., 2008) (Salvat & Pailhe, 2002). As it is situated on the far eastern region of the Indo-Pacific, FP also shows low diversity per unit area (Spalding et al., 2001). Despite this lowered diversity, FP reefs remain largely important to reef research due to their extended monitoring history and resilience.

One of the most well monitored reefs in the world is in Moorea, FP which has been monitored by the Center for Insular Research and Environmental Observatory (CRIOBE) since 1971 (Salvat et al., 2008). This detailed monitoring follows Moorea's long history of disturbances (bleaching events, outbreaks of the coral predator *Acanthastar sp.* and cyclones) and documents changes in coral cover and composition (Traçon et al., 2010). For example, coral cover was reduced to almost 0% in 2010 but the reefs have since recovered to pre-disturbance levels (Souter et al., 2020). Likewise, in Moorea, coral cover returned to pre

disturbance levels following an *Acanthastar* outbreak in only five years where other reef localities have needed up to 40 years to see similar recovery (Trapone et al., 2010). Disturbances have also affected genus distribution with shifts in the major reef forming corals such as declines in *Acropora* sp. and increases in *Pocillopora* and *Porites* sp (Trapone et al., 2010). It's long history of monitoring as well as the resilience shown by coral species makes Moorea an excellent site choice for studies within French Polynesia.

1.5.1.4 Site Conclusions

As indicated above, both Indonesia and FP rely heavily on coral reefs and the resources they provide. FP is situated further east in the Indo-Pacific and thus shows less species diversity than what we would see in the western pacific in areas such as Indonesia and the coral triangle region. For example, in FP around 176 species of scleractinian corals can be found and 1024 species of fish compared to 605 coral species and 2228 fish species found in the Coral Triangle (Salvat & Pailhe, 2002) (Briggs & Bowen, 2013). Despite this dramatic decrease in diversity, reef structure remains relatively similar between Polynesian reefs and those in the coral triangle (Paulay, 1997). On the other hand, reefs in Indonesia tend to be less monitored whereas FP contains some of the longest temporally monitored sites in the world with reefs demonstrating high resilience. Overall, both sites would benefit from a fast and efficient monitoring program, especially given the distressing rate of reef decline and the ecologic, economic and cultural importance of reefs to both societies.

In order to slow the rapid decline of these ecosystems, we need to improve our understanding of reef structure and spatial and temporal patterns to better understand the consequences of stressors (Gonzalez-Rivero et al., 2014). Currently, our understanding of the effects of negative impacts on coral reefs is limited since we lack not only global baseline data but also region-specific data relevant to reef makeup over long time periods (Chirayath & Instrella, 2019). Moreover, as global bleaching events increase in severity and quantity with shorter spacings between events, crucial information is being lost especially without a fast and effective monitoring scheme in place (Nunes, Cruz & Pinheiro, 2020).

2 Article 1: Open Science Resources from the Tara Pacific Expedition Across Coral Reef and Surface Ocean Ecosystems (2023)

2.1 Chapter 2 Foreword

I present here my first scientific article as a coauthor which details the sampling and data collection procedures used for the entire Tara Pacific Expedition. In this work, I was responsible for the data generated in Section 7 (Coral photographic resources and annotations) dataset 1 (Manual Annotations of *in situ* colony (CO) photos) which describes the photographic annotation process in detail. The manual annotations of the photographs were tedious and time consuming, taking me over 4 months to annotate, working full time. These photos and their corresponding annotations, served as the initial dataset for the rest of my thesis work.

Furthermore, due to the extensive environmental and molecular analyses conducted on the photographed colonies as part of the Tara Pacific expedition multiple other datasets providing complimentary health relevant info such as microbiome, stress biomarkers, and telomere DNA are available for the photographed colonies. Although, not explored during the timeframe of my thesis, analyzing the relationships between these parameters and the photo annotations could help to link to investigate how the molecular mechanisms are associated to the visual annotation of the coral colonies.

2.2 Open Science Resources from the Tara Pacific Expedition Across Coral Reef and Surface Ocean Ecosystems (Lombard et al., 2023)

www.nature.com/scientificdata

scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Open science resources from the Tara Pacific expedition across coral reef and surface ocean ecosystems

Fabien Lombard *et al.**

The Tara Pacific expedition (2016–2018) sampled coral ecosystems around 32 islands in the Pacific Ocean and the ocean surface waters at 249 locations, resulting in the collection of nearly 58 000 samples. The expedition was designed to systematically study warm-water coral reefs and included the collection of corals, fish, plankton, and seawater samples for advanced biogeochemical, molecular, and imaging analysis. Here we provide a complete description of the sampling methodology, and we explain how to explore and access the different datasets generated by the expedition. Environmental context data were obtained from taxonomic registries, gazetteers, almanacs, climatologies, operational biogeochemical models, and satellite observations. The quality of the different environmental measures has been validated not only by various quality control steps, but also through a global analysis allowing the comparison with known environmental large-scale structures. Such publicly released datasets open the perspective to address a wide range of scientific questions.

Background & Summary

Marine ecosystems are facing numerous perturbations either of seasonal, climatic, or biological origin which are now further amplified by perturbations due to anthropogenic activities. The resilience of marine ecosystems to perturbations is a general concern, especially when providing ecosystem services and supporting human activities. Tropical coral reefs maintain important ecological services such as fisheries, tourism, or coastal protection, but are also among the most sensitive ecosystems to environmental changes^{1,2}. The health of stony corals, the foundation species of reef ecosystems, is not only governed by the environment, but also by the composition of the holobiont and its symbiotic interactions encompassing a wide range of eukaryotic organisms (e.g., crustaceans, molluscs, fishes), endosymbiotic microalgae, bacteria, fungi, and viruses³. In the open sea, coral ecosystems are associated with islands and participate in their long-term ecological and geological resilience. Coral ecosystems are hotspots of biological activities and energy flux that have a strong effect on the open sea through nutrient enrichment that could propagate in the open ocean, supporting fisheries or biogeochemical fluxes in other marine ecosystems^{4,5}.

However, a more complete understanding of how coral ecosystems are reacting to environmental stressors is complicated as multiple spatial (from microscale to mesoscale) and temporal (from minutes, day, seasons, or decades) scales are involved, as well as various biological complexity levels (from molecular, genetic, physiological to ecosystem). Monitoring ecosystems features at large biological, spatial, and temporal scales is very challenging. An alternative is to use “space-for-time” substitutions which assumes that processes observed at various static spatial scales could reflect what could happen if the same ecological forcing happens at various temporal scales⁶. Historically, this method was used for centuries, for example when Charles Darwin used it to describe the development of islands from barrier reefs, fringing reefs to atolls⁷. This method is still commonly used in ecology, notably when species distribution⁸ or even diversity⁹ are modelled using niche models.

This type of approach is often limited by the compatibility between datasets, where many observations often originated from separate studies with heterogeneous protocols, methods, or measurements. In this respect, large global expeditions have often paved the way to major scientific breakthroughs from the early expeditions conducted by the Beagle or HSM challenger to the more recent Malaspina or Tara Ocean expeditions^{10–11}.

The Tara Pacific expedition has applied a pan-ecosystemic approach on coral reefs and their surrounding waters at the entire ocean basin scale throughout the Pacific Ocean¹². The aim is to provide a baseline reference

*A full list of authors and their affiliations appears at the end of the paper.

of coral holobiont genomic, transcriptomic, and metabolomic diversity spanning from genes to organisms and their interactions with the environment. *Tara Pacific* focused on widely distributed organisms, two scleractinian corals (*Pocillopora meandrina* and *Porites lobata*), one hydrocoral (*Millepora platyphylla*), and two reef fishes (*Acanthurus triostegus* and *Zanclus cornutus*) together with their contextual biological (plankton) and physicochemical environment¹⁴.

The collaboration of more than 200 scientists and participants during this expedition made it possible to sample coral systems across 32 islands (102 sites), together with 249 oceanic stations, resulting in a collection of 57 859 samples encompassing the integral study of corals, fishes, plankton, and seawater. As with previous *Tara* expeditions¹⁵, organizing and cross-linking the various measurements is a stepping-stone for true open access science resources following FAIR principles (Findable Accessible Interoperable and Reusable¹⁶). In this effort, the strategy adopted by *Tara Pacific* is to provide open access data and early and full releases of the datasets once validated or published. Such an approach ensures a long-lasting preservation, discovery, and exploration of data by the scientific community, which will lead to new hypotheses and emerging concepts.

Here, we present an overview of the sampling strategy used to collect coral holobiont specimens in connection with its local, large scale, or historical environment. We also provide a critical assessment of the environmental context. We provide the full registries describing the geospatial, temporal, and methodological information for every sample and connect it to the various sampling events or stations. Extensive environmental context is also provided at the level of samples or stations. Such registries and environmental context collections are essential for researchers to explore the *Tara Pacific* data and will be updated and complemented when additional datasets will be released to the public. Throughout the entire text, terms stated [within brackets] refer to the terms used within the registry or in environmental context datasets.

Methods

Sampling locations. *Tara Pacific* deployed a standardized sampling and analysis protocol to offer a comparative suite of samples covering the widest environmental envelope while optimizing cruising and sampling time over the 2.5 years of the sampling effort. Protocols and global objectives of the *Tara Pacific* expedition were previously detailed for coral samples¹³ and are detailed here in connection with the sample registry. Similarly, protocols and global objectives for ocean and atmosphere sampling were previously described¹⁴ for the 249 stations sampled during daytime (noted [OA001] to [OA249]); night-time sampling between stations and other non-systematic sampling events were noted [OA000]).

A set of 32 island systems (noted [I01] to [I32] in the registry; Table 1, Fig. 1) were targeted to cover the widest range of conditions possible, from temperate latitudes to the equator, from the low diversified system of the eastern Pacific to the highly diverse western Pacific warm pool¹⁷. The variety of coral reef systems explored includes continental islands, remote volcanic islands up to atolls, with varying island sizes or human populations (Table 2). Generally, 3 sites ([S01] to [S03]) per island were selected to conduct the full sampling strategy within 4 days. Occasionally only 2 or up to 5 sites were selected (Table 1).

Sampling coral reef systems. The sampling event sequence and protocols were performed consistently over the whole expedition. Sampling was conducted following the same procedure, approximate timing, and articulated around the same standardized “sampling events” (Fig. 2) which allowed the same collection of samples with a standardized protocol (Table 3). On rare occasions, the timing and protocols were adapted due to sailing conditions and to fit the schedule. Sampling events are characterized by their mode of sampling, which could be either indirectly from *Tara*’s dinghy [ZODIAC] or directly either using scuba-diving ([SCUBA]) or snorkeling ([SNORKEL]). In addition, the sampling device and strategy are included in the sample identifier.

The first set of sampling events (usually in the morning) was mostly devoted to the sampling event [SCUBA-3X10] to sample coral colony fragments. In the meantime, another team pumped underwater, with the [SCUBA-PUMP] to collect coral surrounding water ([CSW]), while the third team snorkeled to capture a total of 10–15 fish using a speargun ([SNORKLE-SPEAR]). A small CTD probe (Castaway CTD) was also deployed from the dinghy down to the reef (generally ~5 to 10 m) to record temperature and conductivity profiles.

The second set of sampling events (usually in the afternoon) was devoted to a survey of coral diversity ([SCUBA-SURVEY]) concurrently with sampling surface water for biogeochemistry ([ZODIAC-NISKIN]), plankton in the size-fractions smaller than 20 μm ([ZODIAC-PUMP]), and plankton in the size-fractions between 20 to 2000 μm ([SCUBA-NET-20]). Finally, over a last dive a coral core was recovered over a large colony of *Porites sp* or *Diploastrea sp* ([SCUBA-CORER]).

Sampling coral colonies [SCUBA-3X10]. During this typical sampling event, a total of 30 coral colonies [C001] to [C030], including 10 colonies for each of the 3 target species (*Pocillopora meandrina*, *Porites lobata*, and *Millepora platyphylla*) were sampled. Each colony was first photographed ([PHOTO]) using a 20 cm quadrat as a scale, their depth recorded and then sampled to collect about 70 grams of each coral by mechanical fragmentation using hammer and chisel. Fragments were placed in Ziploc bags labelled by colony ID and brought back to the boat.

Sampling coral surrounding water [SCUBA-PUMP] and [ZODIAC-NISKIN]. Two *Pocillopora meandrina* coral colonies [C001] and [C010] were marked with small buoys, and [CSW] samples were collected as close as possible to the coral colony before the actual SCUBA-3X10 sampling to avoid contamination of the water samples with fragments or tissues released during the mechanical fragmentation of coral colony. Then, water was pumped using a manual membrane pump onboard *Tara*’s dinghy that was stationary above the coral colony. A scuba diver was holding a clean water tubing next to the colony while the operator onboard the dinghy was

Island code	isl_name (s)	Archipelagos/synonym names	Country	latitude	longitude	station nb
I01	Chapera/Mogo Mogo/Bartolome	Islas de las Perlas	Panama	8.4061	-79.0605	3
I02	Brincaco/del canal de Afuerta/Jicarita	Coiba	Panama	7.4667	-81.7833	3
I03	Malpelo		Colombia	4	-81.6081	2
I04	Rapa Nui	Easter Island	Chile	-27.1167	-109.367	4
I05	Ducie	Pitcairns	United Kingdom	-24.6833	-124.783	4
I06	Tenoko/Tekava/Kamaka	Gambiers	France	-23.14	-134.94	3
I07	Moorea	French Polynesia	France	-17.5333	-149.833	3
I08	Aitutaki	Cook	New Zealand	-18.8561	-159.785	3
I09	Niue		Niue	-19.05	-169.917	3
I10	Upolu	Samoa	Samoa	-13.5833	-172.333	4
I11	Futuna	Futuna/Horn Islands	France	-14.2833	-178.15	3
I11	Alofi	Futuna/Horn Islands	France			
I12	Tuvalu		Tuvalu	-8.5067	179.0979	4
I13	Abaiang	Kiribati	Kiribati	1.4167	173	3
I14	Chuuk	Micronesia	Micronesia	7.4167	151.7833	3
I15	Guam		USA	13.5	144.8	3
I16	Chichi Jima	Ogasawara	Japan	26.9981	142.2181	3
I17	Sesoko	Okinawa	Japan	26.4794	127.9278	3
I18	Fiji		Fiji	-18	179	3
I19	Heron		Australia	-23.4385	151.9084	4
I20	Chesterfield		France	-19.332	158.4727	3
I21	New Caledonia		France	-22.4973	166.4787	3
I22	Guadacanal/Njurokamo/Njapuna	Solomon	Solomon Islands	-8.5672	158.5733	3
I23	Milney Bay		Papua New Guinea	-9.2684	151.4979	3
I24	Kimbe Bay		Papua New Guinea	-5.2801	150.1162	3
I25	Hellen Reef/Tobi/Merir/Pulo Anna/Soronsol	Palau South islands	Palau	2.890117	131.7944	5
I26	Babeldaob	Palau	Palau	7.344777	134.4888	3
I27	Hong Kong		Hong Kong	22.63486	114.1022	2
I28	Taiwan		Taiwan	22.06	121.33	3
I29	Oahu	Hawaii	USA	21.43421	-157.739	3
I30	Isla Cerralvo/Los Frailes/Bahía Chilenos	Baja California	Mexico	24.23236	-109.888	3
I31	Clipperton		France	10.26905	-109.203	3
I32	Brincaco/Rancheria/Jicarita/Las Uvas	Coiba	Panama	8.004	-82.3431	4

Table 1. Summary of the different islands sampled during the Tara Pacific expedition with the associated island code (I01 to I32), their chosen reference name (in bold) corresponding either to the name of the island or of the archipelagos.

pumping the water up to the skiff. First, the water collected was used to rinse the pumping system, as well as a 20 µm metallic sieve and the 50 L carboys that will be used to transport the sample [C010]. Then, 50 L of water was filtered within and around the coral colony onto a 20 µm metallic sieve and directly stored in the dedicated clean 50 L carboy ([SCUBA-PUMP] for [C010]). When available, two replicates of sediment samples (i.e. sand [SSED]) were also taken using two 10 mL cryovials near the sampled colony. Finally, the coral colony [C010] itself was sampled following the [SCUBA-3X10] protocol.

Once the [C010] was sampled, the dinghy was moved on top of colony [C001], where, before any other sampling, carbonate chemistry and nutrient protocols (using a 5 L Niskin bottle for carbonates [CARB] and nutrients [NUT]) as well as for [PH] protocols (using 5 mL polypropylene vials and a 50 mL Falcon tube) were performed. The [PH] was first sampled using two vials (5 mL polypropylene vials for samples), and a falcon 50 mL tube (for later use to rinse the probe) were first lowered closed, opened next to the colony, rinsed with the [CSW], and closed tightly making sure no bubbles were trapped inside the vials. Next, the Niskin bottle was immersed open by the diver [ZODIAC-NISKIN], well rinsed along the descent and with the coral surrounding water near the targeted colony, and finally closed as close as possible to the colony [C001]. The tubing, the sieve, a 4 L Nalgene (protected with reflective tape to isolate the sample from sunlight), and the 50 L carboy dedicated for [C001] were rinsed with the [C001] [CSW]. The 5 L Nalgene bottle was filled with [C001] [CSW] for high-performance liquid chromatography (HPLC). The 50 L carboy was then filled ([SCUBA-PUMP] for [C001]) and the sediment samples [SSED] were collected following the same procedure as for [C010]. Finally, the coral colony [C001] itself was sampled following the same [SCUBA-3X10] protocol. For safety reasons, carbonate chemistry samples [CARB] could not be preserved with mercury (II) chloride on-board the dinghy due to its acute toxicity. Hence, the Niskin bottle was sampled on the last colony of the sampling sequence to minimize the time between sampling and chemical preservation on-board Tara.

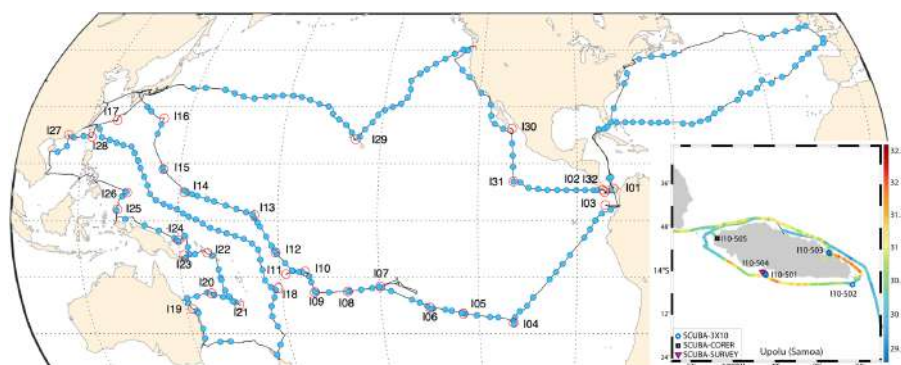


Fig. 1 Tara Pacific expedition (2016–2018) sampling map. Map of sampled coral systems (red circles) and oceanic stations (blue dots). Insert: Example of coral sampling locations around Upolu (Samoa; I10) with overlaid temperature as recorded by the inline thermosalinograph. The absence of sampling during the return trip in the Atlantic Ocean is due to bad weather.

Sampling for fish [SNORKLE-SPEAR]. Fish sampling of two target species (*Acanthurus triostegus* and *Zanclus cornutus*) was operated by spear-fishing and snorkeling for a target number of about 10–15 fishes ([F001] to [Fxxx]) depending on the population present. The targets were speared and immediately stored in labeled individual Ziplock bags to avoid contamination between samples and kept inside a floating container to keep them at water temperature.

Sampling sediments and macroalgae [SCUBA-...]. Sediments and macroalgae samples were sampled when encountered during the different dives. Sediment samples (i.e. sand [SSED]) were taken using two 10 mL cryovials near the sampled colony. Macroalgae, ideally brown macroalgae with thallus morphology type arbutive, ([MA01]-[MAxx]) were photographed ([PHOTO]) and sampled in individual Ziplock bags when encountered.

Coral biodiversity sampling [SCUBA-SURVEY]. Biodiversity sampling transects were conducted in two depths-range environments to sample up to 80 coral colonies ([C041] to [C120]) arbitrarily chosen with ideally up to 40 colonies at a depth of 10–16 m, and up to 40 colonies at a depth of 2–10 m, with an emphasis on sampling across a diverse range of coral hosts at different depths. Two pictures of each colony sampled were taken ([PHOTO]), and small pieces of 1–3 cm² were sampled using a hammer and a chisel or a bone cutter.

Sampling surface seawater [ZODIAC-NISKIN] and [ZODIAC-PUMP]. In addition to the seawater collected next to coral colonies explained above, surface ([SRF]) seawater was sampled at 2 m depth using the manual pump on-board of the dinghy ([ZODIAC-PUMP]). The [SRF] site was chosen to be as close as possible from the coral colonies sampled in the morning but with enough water depth that the plankton net sample could be taken at 2 m depth and at least 5 m above the seafloor. When the sampling site was shallower than 7 m, the site was chosen where these sampling conditions could be met within 100 m around the [CSW] sampling site. The water collected was treated similarly to the [SCUBA-PUMP] samples, with the difference that 100 L [SRF] water was collected into two 50 L carboys. The 4 L Nalgene bottles protected from sunlight were also filled with water at 2 m below the dinghy for HPLC filtrations on-board Tara.

Sampling large size plankton [SCUBA-NET-20]. During this surface water pumping, plankton larger than 20 µm were sampled at 2 m below the sea surface using two small diameter bongo plankton nets with 20 µm mesh size, attached to an underwater scooter ([SCUBA-NET-20]) and towed for about 15 min at maximum speed ($0.69 \pm 0.04 \text{ m} \cdot \text{s}^{-1}$). The average maximum speed of the net tow was estimated in Taiwan (island 28 site 03) measuring the time it took the diver with full gear on and the nets attached, to travel between two buoys separated by a 9-meters line held tight and floating with the current, to avoid any impact of the current. The measurement was repeated three times facing the current, three times in the same direction as the current, and five times with the current sideways. Each net was equipped with flowmeters, but the speed of the underwater scooter was insufficient to trigger their rotation, therefore the time of sampling was precisely timed to estimate theoretically the volume filtered using the following equations:

$$\text{Volume filtered} = \text{Opening area} * \text{Tow speed} * \text{Tow duration} \quad (1)$$

With

Island code	used name	Island type	land area (km ²)	max elevation	population	density (humans/km ²)
I01	Islas de las Perlas	continental isl.	332.9	na	4500	13.52
I02	Coiba	continental isl.	503	416	0	0
I03	Malpelo	island	3.5	320	0	0
I04	Rapa Nui	island	164	507	7750	47.26
I05	Ducie	atoll	3.90	4	0	0
I06	Gambiers	island	31	441	1592	51.35
I07	Moorea	island	134	1207	17718	132.22
I08	Aitutaki	atoll	16.80	124	2194	130.60
I09	Niue	island	260	68	1591	4.60
I10	Upolu	island	2944	1113	193483	62.50
I11	Futuna	island	46.28	524	3225	69.68
I11	Futuna	island	17.78	417	1	0.06
I12	Tuvalu	atoll	26	1.80	11342	436
I13	Abaiang	atoll	16.40	1.80	5568	339.51
I14	Chuuk	island	116.20	238	48651	419
I15	Guam	island	549	406	164229	299
I16	Ogasawara	island	104	916	2821	27.13
I17	Okinawa	island	1201	503	1230000	1024.15
I18	Fiji	island	18270	1324	935974	51
I19	Heron	atoll	0.29	3.60	na	na
I20	Chesterfield	atoll	< 10	6	0	0
I21	New Caledonia	island	18575.50	1629	271407	15
I22	Solomon	island	28400	2335	652857	18.10
I23	Milney Bay	continent	462840	4509	8300000	14
I24	Kimbe Bay	continent	462840	4509	8300000	14
I25	Palau South islands	island	0.85	6	30	35.29
I26	Palau	island	330	242	6000	18.18
I27	Hong Kong	continent	1104	957	7466441	6763
I28	Taiwan	island	35980	3952	23603049	656
I29	Hawaii	island	1545.40	1220	976372	631.79
I30	Baja California	continental isl.	143396	3096	712029	0.89
I31	Clipperton	atoll	1.70	29	0	0
I32	Coiba	continental isl.	503	416	0	0

Table 2. Geological, topological and human population characteristics of the sampled Islands recovered from various sources.

$$\text{Opening area} = \pi * \text{net radius}^2 \quad (2)$$

The volume estimated from the flowmeter reading was about 60 times smaller than the volume calculated theoretically, implying that the flow rate was below the level ensuring proper functioning of the flowmeter. thus, only the theoretical volume will be used in concentration calculations. After 15 minutes of towing, the divers surfaced the two nets and the two cod-ends were sieved through a 2000 µm metallic sieve, into a 2 L Nalgene (r) bottle. The bottle was topped-up with 0.2 µm filtered seawater from the same sampling site and kept at ocean temperature in a bucket during transportation to Tara. Finally, [PH], [NUT] and [CARB] samples were taken at 2 m depth just before leaving the sampling site following the same protocol than for [CSW] sampling and using the same cleaned 5 L Niskin bottle for [CARB] and [NUT], and two 5 mL polypropylene vials as well as a Falcon 50 mL tube for [PH].

Sampling coral cores [SCUBA-CORER]. During the last dive, coral cores were sampled ([SCUBA-CORER]) on *Porites* colonies previously identified and photographed ([PHOTO]). To prevent contamination with coral fragments and tissues released during coring, two [CARB] samples of seawater were taken (one at the surface and one close to the coral colony) before coring and using two 500 mL glass stoppered bottles. Grease was applied to the glass stopper before the dive to allow opening under pressure next to the coral colony. The diver lowered the bottles closed, opened one at 2 m below the surface, and one next to the coral colony. Another seawater sample was taken with a 60 mL HDPE plastic bottle at 2 m depth for subsequent analysis of trace isotopes in relation to the core analysis. Once all seawater was sampled, a 250 mm diameter, 600 round per minute corer from Melun Hydraulique was used to coral cores ([CORE]). Forty coral skeletal cores (40–150 cm long) were collected from

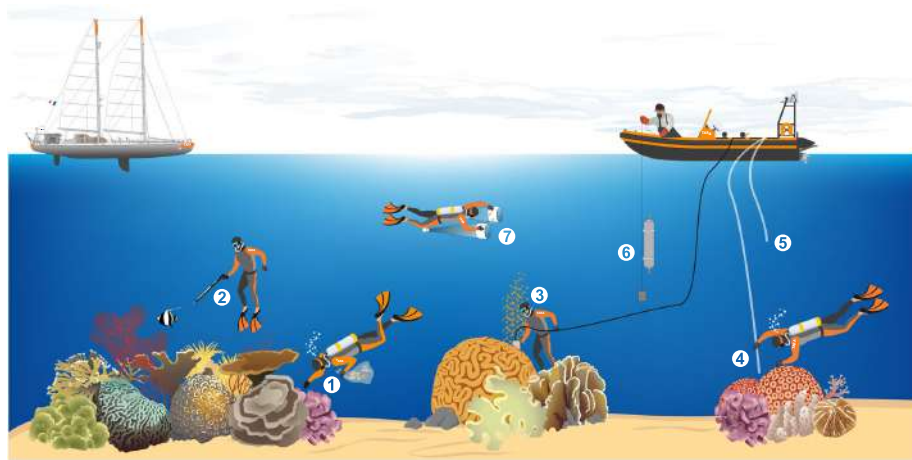


Fig. 2 Schematic overview of the various sampling events conducted during the Tara Pacific expedition while sampling on coral systems. The different events are represented by the different numbers. (1) [SCUBA-3X10] and [SCUBA-SURVEY]; (2) [SNORKLE-SPEAR]; (3) [SCUBA-CORER]; (4) [SCUBA-PUMP]; (5) [ZODIAC-PUMP]; (6) [ZODIAC-NISKIN]; (7) [SCUBA-NET-20].

colonies living between 3 m (Moorea Island-I07) and 20 m (Futuna Islands-I11) depth. From island I19 (Great Barrier Reef) the same protocols were also carried out on large *Diploastrea heliophora* colonies when encountered. **Samples processing.** Benthic samples Once back onboard Tara, the material collected during each sampling event was immediately processed into various samples. Samples were labeled with their target analysis (e.g. sequencing ([SEQ]), imaging, microscopical or morphological inspection ([IMG]) or biogeochemical measurements ([BGC])).

Coral samples obtained from [SCUBA-3X10] events were immediately sorted and separated using bone cutters, in several sub-samples usually labeled with the amount of material used or with the targeted analysis (Table 3). [CS4] and [CS4L] samples containing ~4 g of coral material, were stored at -20°C in 15 ml Falcon tubes and 6 ml of DNA/RNA Shield (Zymo Research, Irvine, CA, USA) for subsequent metabarcoding, metagenomic and metatranscriptomic analyses. [CS4L] only differs from [CS4] by the addition of lysing matrix beads. [CS10] and [CS40] samples, that contain respectively 10 g and 40 g of coral material, were stored in Whirlpak® sample bags, immediately flash frozen in liquid nitrogen, and kept at -20°C . These samples are intended for subsequent metabolomic analysis for [CS10], physiologic/stress biomarkers (symbiont and animal biomasses, antioxidant capacity and protein damages) and telomeric DNA length for [CS40]. Morphological taxonomic identification [CTAX] samples were performed by drying 5 g of material in 50 ml Falcon tubes, and removing organic material with the addition of 3–4% bleach solution during approximately 2 days. After discarding the bleach solution, clean skeletons were preserved dry at room temperature. For histological measurements of reproduction status [CREP], 5 g of each coral colony was preserved in a 50 ml Falcon tube filled with a 3.5% formaldehyde solution and stored at room temperature. Lastly, for transmission electron microscopy examination of coral intracellular details including viruses [CTEM], 0.1 g of coral tissue was preserved with 250 μL 2% glutaraldehyde and conserved at 4°C in a fridge.

Macroalgae samples ([MA]), and the seawater collected with them, were firmly shaken to resuspend attached epiphytic organisms. 20 mL of water was transferred into glass vials and fixed with 2% acidic Lugol and stored at 4°C for future benthic dinoflagellates identification and counts using microscopy ([BD1]), while 100 mL of each replicate were filtered onto a 10 μm pore size polycarbonate filter which was flash frozen and preserved in liquid nitrogen for future metabarcoding analysis ([BDS]).

[SSED] samples were immediately flash frozen when brought back on-board Tara.

About 30 to 40 mL of the seawater that was sampled with the coral fragments of [C001] and [C010] and transported in the coral individual Ziplock bags were transferred immediately after the dive into 50 mL falcon tube and stored at water temperature in non-direct ambient light to recover cultures of plankton species closely associated with coral colonies ([IMG-LIVE]).

When fish were recovered onboard, a [PHOTO] was taken, their sex and length were determined before taking a sample of skin mucus ([MUC]) by collecting 1 cm^2 of skin. The fish were then dissected to recover about 3 cm long of the final section of the digestive tract ([GT]) that was preserved in 2 mL cryotubes with 1 ml of DNA/RNA shield and then stored at -20°C for metagenomic and metabarcoding analyses. One fin sample ([FIN]) was dissected, and preserved into an Eppendorf tube filled with 95% ethanol for population genetic analyses. Lastly, the otolith ([OTO]) was also dissected and stored dry into an Eppendorf tube at room temperature for later aging of each fish.

Analysis category	Sample type	n (samples)	n (rep.)	Material sampled (sample-material_label)	Amount of material	Processing	container	conservative	conservation temperature	Targeted analysis
SEQ	CS4	2703	1	Coral	4g	cut coral parts	Falcon 15 ml	DNA/RNA shield	-20°C	MetaB, metaG, metaT
SEQ	CS4L	2651	1	Coral	4g	cut coral parts	Falcon 15 ml	DNA/RNA shield + lysing matrix beads	-20°C	MetaB, metaG, metaT
SEQ	CS10	2738	1	Coral	10g	cut coral parts	Whirlpak bag	flash frozen	-20°C	Metabolomic
SEQ	CS40	2701	1	Coral	40g	cut coral parts	Whirlpak bag	flash frozen	-20°C	biomarkers and telomere length
IMG	CTAX	2763	1	Coral	5g	cut coral parts, dried, and bleach added for few hours	Falcon 50 ml	Bleach	RT	morphology, taxonomy
IMG	CREP	2649	1	Coral	5g	cut coral parts	Falcon 50 ml	3.7% formaldehyde	RT	histological analysis of reproduction
IMG	CTEM	2385	1	Coral	0.1 g	cut coral parts	2 ml cryotubes	2% glutaraldehyde	4°C	transmission electron microscopy analysis
IMG	PHOTO	10830	2	Coral, Fish	—	—	—	—	—	morphology, taxonomy
SEQ	MUC	1059	1	Fish	—	dissection	coton swab+ 2mL cryotube	DNA/RNA shield	-20°C	MetaB, metaG
SEQ	GT	1059	1	Fish	—	dissection	2 ml cryotubes	DNA/RNA shield	-20°C	MetaB, metaG
SEQ	FIN	1059	1	Fish	—	dissection	eppendorf	ethanol	RT	population genetic analyses
IMG	OTO	1057	1	Fish	—	dissection	eppendorf	—	RT	aging
SEQ	CDIV	2628	1	Coral	<0.5g	cut coral parts	2 ml cryotubes	DNA/RNA shield	-20°C	MetaB, metaG
SEQ	SSED	351	1	Sediment	7.5 ml	seawater replaced with DNA/RNA shield or ethanol and homogenized	15 mL Flacon tubes	DNA/RNA shield or ethanol	-20°C	MetaB, metaG
IMG	CORE	92	1	Coral	26–126 cm	dried 24–48h	plastic bubble wrap	—	RT	morphologic and isotopic analysis
BGC	MTE-LSCE	170	1	Seawater	60 mL	—	60 mL HTPE vial	—	RT	Trace elements (Li, Bo) isotopes measurements
BGC	PH	364	2	Seawater	5 mL	analysed onboard	5 ml plastic vial	—	—	pH measurements
BGC	CARB	364	1	Seawater	500 mL	—	500 mL glass bottle	Hg2Cl2	RT	Carbonate system measurements
IMG	BDI	152	1	benthic dinoflagellates (on brown algae)	20mL	water from shaken macroalgae	20 ml scintillation vials	2% acidic lugol	4°C	microscopic count
SEQ	BDS	124	1	benthic dinoflagellates (on brown algae)	100 mL	water from shaken macroalgae	45 mm 10µm PC filter stored in cryotubes	flash frozen	LN	MetaB, metaG, metaT
BGC	HPLC	944	2	Water, pigments	2L	filtered on a 25mm-diameter, 0.7-µm-pore glass fiber filter	1.5 ml cryotubes	flash frozen	LN	HPLC pigment analysis
BGC	NUT	862	2	Seawater	20 mL	filtered through a 0.45 µm-pore size cellulose acetate membrane with a syringe	20 mL polyethylene vials	—	-20°C	macronutrients dosing
SEQ	S023	1104	2	Plankton (0.2–3µm)	50 L	Filtration	5 mL cryotubes	flash frozen	LN	MetaB, metaG, metaT
SEQ	S320	1086	2	Plankton (3–20µm)	50 L	Filtration	5 mL cryotubes	flash frozen	LN	MetaB, metaG, metaT
SEQ	S<02	874	2	Water, Viruses (<0.2µm)	10L	FeCl3 precipitation and filtration	5 mL cryotubes	—	4°C	Sequencing
SEQ	S<02>	127	1	Water, membranes vesicles (<0.2µm)	80 L	—	50 mL Falcon tube	—	-20°C	Sequencing
IMG-SEQ	SCG	1056	1	Plankton	4 ml	—	5 mL cryotubes	600µl of 48% Glycine Betaine, flash frozen	LN	single cells sequencing

Continued

Analysis category	Sample type	n (samples)	n (rep.)	Material sampled (sample-material_label)	Amount of material	Processing	container	conservative	conservation temperature	Targeted analysis
Sample protocol label										
IMG	FCM	1078	2	Plankton	1.48 5mL	mix and incubate 15min at RT	2 ml cryotubes	15µL Glutaraldehyde 25%/PoloXamer 10%; flash frozen	LN	Flow cytometry
IMG	SEM	566	1	Plankton	500 mL	filtered onto 47mm 0.22µm PC membranes, dried 2h at 50°C	petrislides	—	RT	Scanning electron microscopy
IMG-SEQ	FISH	562	1	Plankton	225 mL	incubate 1–24 h with PFA 10x; filter on 25mm, 0.22µm PC filter, rinse with ethanol, dry for 5–10 minutes	petrislides	—	–20 °C	Fluorescent in situ hybridation
SEQ	S20	714	2	Plankton	1L (250ml per filter)	filtered onto 47mm 10µm PC membranes	5 mL cryotubes (two filters per tube)	flash frozen	LN	MetaB, metaG, metaT
IMG	H20	422	1	Plankton	45 mL	—	50 mL Flacon tubes	5mL of 10% paraformaldehyde and 500µl of glutaraldehyde 25% EM grade	4°C	High throughput confocal microscopy
IMG	LIVE20	358	1	Plankton	50 mL	analysed using Flowcam onboard	—	—	—	Quantitative imaging analysis
IMG-SEQ	E20	444	1	Plankton	100–250 mL	concentrated onto 20µm sieve, stored in ethanol during 24h before serving again to change the ethanol	15 mL Flacon tubes	95% mollecular grade ethanol	–20 °C	single cells sequencing
IMG-SEQ	SCG20	212	1	Plankton	4 mL	—	5 mL cryotubes	600µl of 48% Glycine Betaine, flash frozen	LN	single cells sequencing
BGC	SAL	50	1	Seawater	250 ml	—	—	—	RT	salinity measurements
IMG	L20	243	1	Plankton	250 mL	concentrated onto 20µm sieve, resuspended using filtered sea water	50 mL Flacon tubes	1mL of acidic Lugol solution	4°C	microscopic observations
IMG	F20	240	1	Plankton	45 mL	—	50 mL Flacon tubes	1mL of acidic formalin 37%, and filled up to 50mL with sodium teraborate decahydrate buffer	RT	microscopic observations
IMG	F300	510	1	Plankton	1L	concentrated onto 200µm sieve, resuspended using filtered sea water with sodium teraborate decahydrate buffer	250 mL double closure bottles	30mL of 37% formalin solution	RT	Quantitative imaging analysis
SEQ	S300	603	2	Plankton	1L(250 mL per filter)	prefiltered onto 2mm metallic sieve, filtered onto 47mm 10µm PC membranes	5 mL cryotubes (two filters per tube)	flash frozen	LN	MetaB, metaG, metaT
IMG	AI	1323	1	Aerosols	~21.6 m3	—	petrislide	dried	RT	microscopic observations
SEQ	AS	1300	1	Aerosols	~21.6 m3	—	2 ml cryotubes	flash frozen	LN	MetaB
SEQ-BGC	ABS	1306	2	Aerosols	~21.6 m3	—	2 ml cryotubes	flash frozen	LN	MetaB and biogeochemistry
BGC	MTE-USC	249	1	Seawater	125 mL	—	acid cleaned 125 mL low density PET bottle	—	RT	Trace metal analysis

Table 3. Correspondence between samples types and their associated events and a summary of the protocol used and targeted analysis. RT: Room temperature, LN: Liquid Nitrogen, MetaB: metabarcoding, MetaG: metagenomic, MetaT: metatranscriptomic, PC: Polycarbonate, PET: Polyethylene.

Coral samples obtained from [SCUBA-SURVEY] were collected for symbionts and coral diversity analysis ([CDIV]) using different marker genes (metabarcoding, 18S, 16S and ITS2). About 0.5 g of material was preserved with DNA/RNA shield and stored into 2 mL cryotubes at -20°C .

Finally, samples collected during [SCUBA-CORER] events were also processed and stored onboard Tara. The [CORE] were rinsed with freshwater, air dried for 24–48 h before being wrapped into a plastic bubble wrap for sclerochronological and geochemical analysis, to recover historical water biogeochemical properties. The [PH], [CARB] and [MTE-LSCE] samples associated with the coral core [CORE] were processed following the same protocol than the water samples collected with the [SCUBA-PUMP] and [ZODIAC-PUMP] (explained in section 2.2.2), with the exception that the [CARB] and [MTE-LSCE] samples were already stored in their final container during sampling on the dinghy.

Water samples for biogeochemistry The [PH] was measured from the two replicates 5 mL polypropylene vials onboard Tara using an Agilent Technologies Cary 60 UV-Vis Spectrophotometer equipped with an optical fiber. The detailed protocol was previously described¹⁸, but briefly, the 5 mL vials and the 50 mL falcon tube were kept closed and acclimated to 25°C for 2–3 h. Absorbance at specific wavelengths was then read before and after the addition of 40 μL meta-Cresol Purple dye to each 5 mL vial. The probe was rinsed between each measurement using the 50 mL falcon tube containing the same seawater as the 5 mL vials samples. TRIS buffer solutions¹⁸ were measured regularly along the cruise to validate the method and correct for potential drifts of pH of the dye solution.

The Niskin bottles of the morning ([CSW] for [C001] colony) and afternoon ([SRF]), carefully kept closed since sampling on the dinghy, were each used to rinse and fill one 500 mL glass stoppered bottle on Tara. Some grease was applied to the glass stopper, and bottles were filled with water samples leaving 2 mm of air below the bottom of the bottleneck. Note that the [CARB] samples associated with the [CORE] samples were already stored in their final container and grease was already applied to the glass stopper before the dive. The water level of these samples was simply adjusted to 2 mm below the bottleneck. All [CARB] samples were immediately poisoned with 200 μL of saturated mercury (II) chloride solution (HgCl_2) and stored at room temperature.

The Niskin water was then used to rinse and fill up trace element samples in 60 mL HDPE plastic bottles [MTE-LSCE]. These samples were stored at room temperature and used to confirm the absence of local influence on Li and B isotopic signals. Similar to [CARB] associated with [CORE] samples, the [MTE-LSCE] samples associated with [CORE] samples were already stored in their final containers, therefore, were just stored at room temperature.

The water remaining from the Niskin bottle, sampled in the morning ([CSW] for [C001] colony) and the afternoon ([SRF]), was used to prepare macronutrient samples ([NUT]). A 50 mL syringe was rinsed with the sampled seawater three times. A filter 0.45 μm -pore size cellulose acetate membrane was then connected to the syringe and ~ 20 mL of sample water was run through it to rinse the filter. Once the syringe, filter and vials were properly rinsed twice, two 20 mL polyethylene vials were filled running the sampled water through 0.45 μm -pore updisic syringe filter. Nutrient samples were stored vertically at -20°C .

Two replicates of two liters of seawater sampled in the 4 L Nalgene bottle from the [SCUBA-PUMP] and [Zodiac-PUMP] events, were filtered onto 25 mm-diameter, 0.7 μm -pore glass fiber filters (Whatman GF/F) and immediately stored in liquid nitrogen for later High-Performance Liquid Chromatography (HPLC) analysis to obtain pigments concentration.

Water samples for genomics and imagery The water collected during the [SCUBA-PUMP] and [Zodiac-PUMP] events was treated similarly, with the only difference that while [Zodiac-PUMP] samples were treated in duplicates, the two 50 L samples collected during [SCUBA-PUMP] correspond to [C001] and [C010] colonies. This applies only for sequencing samples ([SEQ-S]), while all other samples were taken in duplicates. Additionally, all genomic samples were processed to be as comparable as possible with previous existing samples from Tara Oceans^{12,15}.

As soon as back on-board Tara, the water collected was used to rinse and fill one (for each [CSW]) or two (for [SRF]) 50 L carboy and two 2 L Nalgene(r) bottles. The content of the 50 L carboys was immediately size-fractionated by sequential filtration onto 3 μm -pore-size polycarbonate membrane filters and 0.22 μm -pore-size polyethersulfone Express Plus membrane filters. Both were placed on top of a woven mesh spacer Dacron 124 mm (Millipore) and stainless-steel filter holder “tripods” (Millipore). Water was directly pumped from the 50 L with a peristaltic pump (Masterflex), and separated into samples that contain particles from 3–20 μm ([S320]) and 0.2–3 μm ([S023]) for latter sequencing. To ensure high-quality RNA, the filtering of the first replicate ([C001] for [CSW] samples and any of the two 50 L carboys for [SRF]) were stopped after 15 minutes of filtration while the second was continued for the full volume (or a maximum of 60 min) to maximize DNA yield. Filters were folded into 5 mL cryovials and preserved in liquid nitrogen immediately after filtration. During this filtration 10 L of 0.2 μm filtered water ([S<02]) was collected from each replicate, 1 mL of FeCl_3 solution was added to flocculate viruses¹⁹ for 1 hour. This solution was then again filtered onto a 1 μm -pore-size polycarbonate membrane filter using the same filtration system as for [S320] [S023]. Filters were then stored in 5 mL cryotubes and stored at 4°C for later sequencing of viruses. The 80 L remaining of 0.22 μm prefiltered water was used to filter membranes vesicles ([S<02>]) using an ultrafiltration Pellicon2 TFF system by keeping the pressure below 10 psi until the concentrate was reduced to a final volume of 200–300 mL. This sample was further concentrated using a Vivaflow200 TFF system at a recirculation rate of 50–100 mL/min and less than one bar of pressure until obtaining a final sample of 20 mL. Flushing back the system usually brings this volume to up to 40 mL which was stored in a 50 mL Falcon tube at -20°C .

Two 4 mL samples were taken from the 2 L Nalgene bottles, and stored into 5 mL cryotubes fixed with 600 μL of 48% Glycine Betaine and directly flash-frozen for later single cells genomic analysis ([SCG]). For flow cytometry cell counting ([FCM]), two replicates of 1.485 mL of sampled water were placed into 2 mL cryotubes pre-aliquoted with 15 μL of fixative composed of Glutaraldehyde (25%) and PoloXamer (10%). Tubes were then

mixed gently by inversion, incubated 15 min at room temperature in the dark before being flash-frozen, and kept in liquid nitrogen. For scanning electron microscopy ([SEM]), 500 mL of water was filtered onto a 47 mm, 0.22 µm pore size, polycarbonate filter, placed in a petri slide, dried for two hours at 50 °C and conserved at room temperature. Fluorescence *In Situ* Hybridization ([FISH]) samples were prepared by adding 225 mL of seawater into a 250 mL plastic vial containing 25 mL of 10xPFA. The samples were incubated at 4 °C before filtration onto two 25 mm 0.22 µm pore size polycarbonate filters, rinsed with ethanol, placed in petri slides, dried for 5–10 minutes before being stored at –20 °C.

Samples collected during the [SCUBA-NET-20] were fractionated for sequencing and imaging needs. One litre of the sample collected was filtered onto four 47 mm, 10 µm pore size, polycarbonate membranes (250 mL each). Filters were then placed into 5 mL cryotubes, flash-frozen, and stored in liquid nitrogen for later sequencing ([S20]). 45 mL was subsampled into a 50 mL Falcon tube, fixed with 5 mL of 10% paraformaldehyde and 500 µL of glutaraldehyde 25% EM grade, and stored at 4 °C for future high-throughput confocal microscopy ([H20]; e.g.²⁰). 4 mL was stored in 5 mL cryotubes, fixed with 600 µL of 48% glycine betaine, immediately flash frozen and kept in liquid nitrogen for single cell genomics ([SCG20]). Another sample for single cell sequencing stored in ethanol ([E20]) was done by filtering 100 to 250 mL of the sample onto a 20 µm sieve and re-suspended in EM grade ethanol for 24 h at 4 °C. After incubation, the sample was sieved a second time to remove any trace of seawater, re-suspended with EM grade ethanol into 15 mL falcon tube, and stored at –20 °C. Finally, a 50 mL sample was directly imaged live onboard ([LIVE20]) using a FlowCam²¹ Benchtop B2 series equipped with a 4x lens and processed using the auto-image mode.

Oceanic sampling. To obtain both a large scale and local (around coral reef island) environmental characterization, a comprehensive set of physical, chemical and biological properties of the sea surface ecosystem were recorded while cruising. This sampling scheme was framed to be compatible with the previous Tara Ocean expedition measurements^{12,15}, but also to provide a continuity with water samples conducted directly on the coral reef. Furthermore, while the biology and ecology of surface ecosystems remain largely unknown, they are an essential component of air-land-sea exchanges and are subjected to numerous hydrological, atmospheric, physical and radiative constraints²² and are therefore at the frontline of climate change and pollution.

The main goals and general overview of this sampling are already described^{14,23} and will be briefly presented here in the context of the different sampling events and samples that were generated. Measurements and samples could be separated into two types: i. local samples originating from a local sampling event, and ii. autonomous high frequency continuous measurements of atmospheric and surface seawater properties (e.g. per minute averages of higher frequency measurements). In the case of the discrete water sampling, the different sampling events were either attributed to a station (noted [OA001] to [OA249]) if they were conducted in a reasonably short time lapse (>75 km away, or >0.25 days away from a group of OA Events), or noted [OA000] otherwise. Similarly, every OA station located within 200 nautical miles (370 km) of island were annotated with that island label, i.e. the sampling-design-label of the corresponding OA Events and OA Samples is [OA##-1##-C000]. The continuous sampling was conducted as follows: a. surface seawater measurements were performed by pumping water continuously through the boat hull ([INLINE-PUMP]) at ~1.5 m depth, b. light and atmosphere properties were measured 5 m above the sea level ([PAR + BATOS]), and c. aerosols were sampled by pumping air on top of the mast ([MAST-PUMP]) at ~27 m (15 m during the first trans-Atlantic transect prior to May 2016).

Sampling events. Sampling was organized following several successive events, generally at daily frequency, in the morning. Water collection while cruising was carried out by a custom-made underway pumping system nicknamed the [DOLPHIN] connected by a 4 cm diameter reinforced tubing to a large volume industrial peristaltic pump (max flow rate = 3 m³ h⁻¹) on the deck. The system was equipped with a metallic pre-filter of 2 mm mesh size, two degassers, and a flowmeter to record the volume of water sampled. Unfiltered water was collected first for a series of protocols, water was prefiltered using a 20 µm sieve to rinse and fill two 50 L. Both unfiltered seawater use and 20 µm filtered seawater were labelled as [CARBOY]. To collect larger plankton, water was pumped from the DOLPHIN into a 20 µm net fixed on the wetlab's wall ([DECKNET-20]) for 1 to 2 hours depending on biomass concentration simultaneously to a net tow using a "high speed net" ([HSN-NET-300]). The HSN was equipped with 300 µm mesh sized net and designed to be efficient up to 9 knots. It was towed from 60 to 90 minutes depending on the plankton density. Near islands and in the Great Pacific Garbage Patch, a Manta net ([MANTA-NET-300]) with a 0.16 × 0.6 m mouth opening with a 4 m long net with 300 µm mesh size was used concurrently at a maximum speed of 3 knots. Finally, trace metal samples ([MTE-USC]) were collected from the bow using a metal-free carbon fiber pole [HANDHELD-BOW-POLE] on which a plastic fixation have been added to insert a 125 mL low density polyethylene bottle (LDPE) which was previously pre-washed on land and stored individually in separate Ziploc bags. To avoid contamination from the boat, samples were hand held collected, wearing polyethylene gloves, while cruising upwind on the bow of the boat (i.e., before the boat got in contact with the collected water; Fig. 3).

Samples processing. Water, plankton and aerosols samples collected in the vicinity of islands and from the open sea were processed as much as possible following similar protocols than on islands. Samples collected both on islands and in open sea are marked with asterisks* here, and only the few differences in protocols will be noted. From Dolphin, unfiltered waterUnfiltered seawater collected from the [DOLPHIN] was used to process several samples for biogeochemical purposes ([BGC]). For every station, samples were collected for nutrients [NUT]*, [PH]* measurements and pigments analysis by [HPLC]*. Salinity [SAL], carbonates ([CARB]*) and trace elements [MTE-LSCE]* were sampled on a weekly basis. [SAL] samples were done by sampling 250 mL of seawater in a 250 mL hermetically sealed glass bottle.

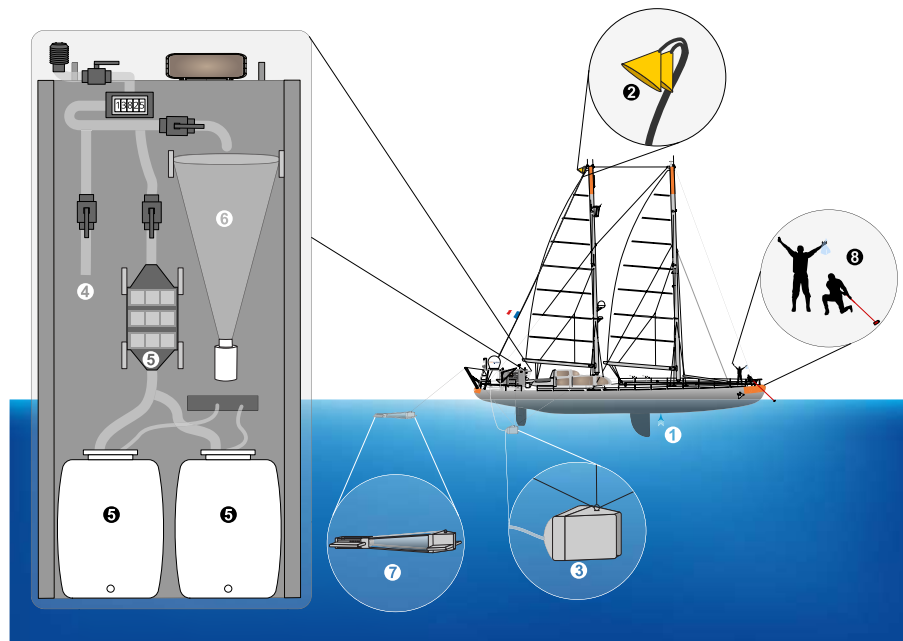


Fig. 3 Schematic overview of the various sampling events conducted during the Tara Pacific expedition while sampling on oceanic stations. The different events are represented by the different numbers. (1) [INLINE-PUMP]; (2) [MAST-PUMP]; (3) [DOLPHIN] pumped water that is either used (4) [RAW], filtered at $20\ \mu\text{m}$ to fill two 50 L (5) [CARBOY], or filtered through (6) [DECKNET-20]; (7) “high speed” [HSN-NET-300] or [MANTA-NET-300] plankton nets; (8) [HANDHELD-BOW-POLE].

From Dolphin, pre-filtered waterThe two 50 L carboys of $20\ \mu\text{m}$ prefiltered seawater were used to produce size fractionated samples for genomic analyses ([S320]* [S023]* [S < 02]*). The same pre-filtered seawater was sampled for flow cytometry cell counting ([FCM]*) and single cell genomic ([SCG]*).

From Dolphin-DecknetOnce the [DECKNET-20] time limit reached (between 1 and 2 hours), the flow was stopped and the net was carefully rinsed with $0.2\ \mu\text{m}$ filtered seawater. The plankton sample was then transferred to a 2 L Nalgene bottle and completed to 2 L with $0.2\ \mu\text{m}$ filtered seawater. The sample was homogenized by repeated smooth bottle flips and split into four 250 mL subsamples for [S20]*, one 250 mL sample for [E20]*, one 250 mL sample for [LIVE20]*, and one 45 mL sample for [H20]*. In addition to these already described protocols, one 250 mL sample was also taken for [L20], for which the seawater was drained using a $20\ \mu\text{m}$ sieve and the plankton was transferred in a 50 mL Falcon tube and fixed with 1 mL of acidic lugol solution for latter microscopic observations. Finally, a 45 mL sample was taken for [F20], transferred in a 50 mL Falcon tube and fixed with 1 mL of 37% formalin solution and completed to 50 mL with sodium tetraborate decahydrate buffer solution for latter microscopic observations.

From HSN/Manta netsOnce recovered, samples collected both by the HSN net and the Manta net followed the same procedure. The net was carefully rinsed from the exterior to drain organisms into the collector. Its content was transferred using $0.2\ \mu\text{m}$ filtered sea water in a 2 L Nalgene Bottle and completed to 2 L. The sample was then homogenized and split in two 1 L samples. The first half was prefiltered onto a 2 mm metallic sieve and filtered onto four 47 mm $10\ \mu\text{m}$ pore size polycarbonate membranes (250 mL each). Filters were then placed into 5 mL cryotubes, flash frozen and conserved in liquid nitrogen for latter sequencing ([S300]). The second fraction was concentrated onto a $200\ \mu\text{m}$ sieve and resuspended in a 250 mL double closure bottle using filtered seawater saturated with sodium tetraborate decahydrate, fixed with 30 mL of 37% formalin solution and stored at room temperature for latter taxonomic and morphological analysis using imaging methods ([F300]).

From Mast-pumpAerosols pumped through one of the ([MAST-PUMP]) inlets were channelled through a conductive tubing of 1.9 cm inner diameter to four parallel 47 mm filter holders installed in the rear hold using a vacuum pump (Diaphragm pump ME16 NT, VACUUBRAND GmbH & Co KG, Wertheim, Germany) at a minimum flow rate of 30 lpm (20 lpm prior to May 2016). Three filter holders were equipped with $0.45\ \mu\text{m}$ pore size PVDF filters for latter aerosol sequencing ([AS]) and biogeochemical analysis together with sequencing ([ABS]), while the fourth one was a $0.8\ \mu\text{m}$ pore size polycarbonate filter for later aerosol imaging ([AI]) analysis

using scanning electron microscope. Twice a day (12 h pumping periods), at approximate dusk and dawn, those filters were changed, [AS] and [ABS] filters were placed into 2 mL cryotubes (2 filters for each [ABS] sample) and immediately flash frozen while [AI] filters were packaged in sterile PetriSlide preloaded with absorbent pads and stored dry at room temperature.

Continuous measurements. As previously described (see^{14,23}), a comprehensive set of sensors were combined to continuously measure several properties of the water but also atmospheric aerosols and meteorological conditions. All sensors were interfaced to be synchronized with the ship's GPS and synchronized in time (UTC time). Surface seawater was pumped continuously through a hull inlet located 1.5 m under the waterline using a membrane pump (10 LPM; Shurflo), circulated through a vortex debubbler, a flow meter, and distributed to a number of flow-through instruments. A thermosalinograph [TSG] (SeaBird Electronics SBE45/SBE38), measured temperature, conductivity, and thus salinity. Salinity measurements were intercalibrated against unfiltered seawater samples [SAL] taken every week from the surface ocean, and corrected for any observed bias. Moreover, temperature and salinity measurements were validated against Argo floats data collocated with Tara. A CDOM fluorometer [WSCD] (WETLabs), measured the fluorescence of coloured dissolved organic matter [fdom]. An [ACS] spectrophotometer (WETLabs) measured hyperspectral (4 nm resolution) attenuation and absorption in the visible and near infrared except between Panama and Tahiti where an AC-9 multispectral spectrophotometer (WETLabs) was used instead. A filter-switch system was installed upstream of the [ACS] to direct the flow through a 0.2 µm filter for 10 minutes every hour before being circulated through the [ACS] allowing the calculation of particulate attenuation [ap] and absorption [cp], by removing the signal due to dissolved matter, drift, and biofouling²⁴. From November 13, 2016 to May 6, 2017, a backscattering sensor [BB3] (WETLabs ECO-BB3) in a flowthrough chamber (BB-box) was added to the underway system, upstream of the switch system, to measure the volume scattering function [VSF] at 124° and 3 wavelengths (470, 532, 650 nm) and estimate the backscattering coefficient [bbp]. From May 7th 2017 to the end of the expedition, the BB-box and the [BB3] were moved downstream of the filter-switch system to run 0.2 µm filtered seawater for 10 minutes every hour in order to remove the biofouling signal and improve [bbp] estimations. Chlorophyll a content [chl] was estimated from [ap]²⁵ and [cp] (when [cp] is hyperspectral²⁶), as well as other pigments (when [ap] is hyperspectral²⁷). The [chl] estimated from [ap] was then calibrated against the [HPLC] [chl]²⁵. The particulate organic carbon concentration [poc] was estimated both using an empirical relation²⁸ between measured [poc] and measured [cp], or applying an empirical relation between measured [poc] and [bbp]²⁹. Phytoplankton organic carbon [cphyto] was estimated by an empirical relationship with [bbp]³⁰. An indicator for size distribution of particles between 0.2 and ~20 µm [gamma] was calculated from [cp]³¹. A brief description of the methods to analyse, calibrate, correct, and estimate bio-optical proxies are detailed in the section Technical Validation and more extensively explained in each processing report attached with the dataset.

An Equilibrator Inlet Mass Spectrometer [EIMS] (Pfeiffer Vacuum Quadrupole 1–100 amu) measured the Oxygen to Argon ratio in percent [o2ar], coupled with an optode (Aanderaa optode 4835) measuring oxygen concentration in the seawater [O2]. Concurrently with samples collected through the [MAST-PUMP], two instruments were installed aboard Tara to measure the size distribution and abundance of atmospheric aerosol particles: a scanning mobility particle sizer ([SMPS], SMPS-C GRIMM Aerosol Technik Ainring GmbH & Co. KG, Ainring, Germany) measuring particles in the size range 0.025–0.70 µm, and an optical particle counter ([EDM]; EDM180 GRIMM Aerosol Technik Ainring GmbH & Co. KG, Ainring, Germany) measuring all particles in the size range 0.25–32 µm. The SMPS was set to perform a full scan of particle distribution every 5 min and the EDM produced a particle size distribution every 60 s. Data provided from [EDM] includes both the total particle concentration (nb cm⁻³) in the size range 0.25–32 µm every 60 seconds, and through a second dataset averaged every 30 minutes, both the particle concentration (nb cm⁻³) together with its normalized size distribution (dN/dlogDp (nb cm⁻³), i.e., the concentration divided by the log of the size width of the bin), while data from [SMPS] are averaged at the hour scale and provided both at the scale of particle concentration (nb cm⁻³) together with its normalized size distribution (dN/dlogDp (nb cm⁻³)).

Together with navigation data such as speed over ground [sog] and course over ground [cog] meteorological station (BATOS-II, Météo France) measured air temperature, relative humidity, and atmospheric pressure at 7 m above sea level. True and apparent wind speed and direction was measured at about 27 m above sea level. In October 2016 a Photosynthetically Active Radiation [par] sensor (Biospherical Instruments Inc. QCR-2150) was mounted at the stern of the boat (~5 m altitude).

Data Records

The full collection of datasets has been deposited either at Pangaea or at Zenodo depending on their nature, but also on the likelihood to be updated.

Provenance metadata. Tara Pacific datasets are articulated around a consistent set of provenance metadata that provide temporal (UTC date and time) and spatial (latitude, longitude, depth or altitude) references as well as annotations about environmental features and place names, using controlled vocabulary from the environmental ontology (<https://www.ebi.ac.uk/ols/ontologies/envo>) and the marine regions gazetteers (<https://www.marineresources.org/>). These metadata are available at three granular levels: sampling stations and sites, sampling events, and samples collected at a specific depth.

A [sampling-design-label] is provided to facilitate the identification and integration of data that originate from the same open ocean station (OA###), island (I##), site (S##) or coral colony (C###), and hence share provenance and environmental context. For example, data originating from coral colony number twelve on the second site of the fourth island visited by Tara will bear the sampling design label OA000-I04-S02-C012. Similarly, data collected at station number 99 in the middle of the Pacific Ocean will bear the sampling design

label OA099-100-S00-C000, and data collected at open ocean station number 41 within 200 nautical miles of island number four will bear the sampling design label OA041-104-S00-C000.

Each sample is also characterized by its sampling event which have several properties such as its date and time (UTC) of sampling ([sampling-event_date_time-utc]), the type of event from which the sample originates ([sampling-event_device_label]), the material sampled ([sample-material_label]; see Table 3), the protocol used ([sampling-protocol_label]; see Table 3) and finally the barcode attributed to the final sample obtained and replicated on the logsheets ([sample-storage_container-label]). Finally, each sample, in addition to its original barcode was characterized by an event label and a sample label composed of that previous information such as:

Sample label: TARA_SAMPLE_[sampling-event_date_time-utc]_[sampling-design_label]_[sampling-environment_feature_label]_[sample-material_label]_[sampling-protocol_label]_[sample-storage_container-label]

Event label: TARA_EVENT_[sampling-event_date_time-utc]_[sampling-design_label]_[sampling-day-night_label]_[sampling-environment_feature_label]_[sample-material_label]_[sampling-protocol_label]_[sample-storage_container-label]

The provenance context of all samples collected during the Tara Pacific Expedition is available as a single UTF-8 encoded tab-separated-values file, in open access at Zenodo and replicated in part at BioSamples (XYZ). In addition to georeferences and place names, the provenance metadata includes sample unique identifiers, taxonomic annotation from NCBI, and links to sampling logsheets and campaign summary reports.

Additionally, the full repository containing the campaign summary reports, sampling authorisations, logsheets and the full record of coral images could be consulted on Pangaea (<https://store.pangaea.de/Projects/TARA-PACIFIC/>). The full list of sampling events is consultable on the following dataset²²: <https://doi.org/10.1594/PANGAEA.944548>.

Environmental context for data analysis. Rich collection of environmental parameters collected from either samples, on-board measurements, satellite imagery, operational models or even calculated from astronomical atlas were compiled and made available for further analysis. These environmental measurements were provided in a multi-layered way in open access to either Pangaea or Zenodo (Tables 4 and 5), depending on the potentiality to require updates, with (1) raw measurements at the measure level for both physical samples or for on-board continuous measurements, accompanied with their quality check flags (2) a combined version regrouping all measurement at the sampling event level and adding satellite imaging and results obtained from operational models. (3) This latter was propagated, together with all measurements done on samples, to provide an environmental context to every collected samples belonging to the same station, but by also providing indices of the spatial ([dxy]), temporal ([dt]) and vertical ([dz]) discrepancies between the various measures and the designed sample and their variability (as assessed by mean, standard deviation, number of measures and 5, 25, 50, 75, 95 percentiles when possible); (4) a simplified version at the site level where all synonym measurements were cross-compared and chosen by level of quality. (5) At the scale of the site level, a series of Lagrangian and Eulerian diagnostics were calculated using satellite-derived and modelled velocity fields, providing multiple information on water mass transport and mixing (6) Finally, and for coral sites only, historical data of temperature were extracted (see (6) Historical data on coral sites) from satellite imagery to provide an historical overview of past heatwave experienced by the sampled coral reefs (since 2002 up to the sampling date).

Raw measurements from samples or sensors. From sensors, the measurements were standardized at the minute scale when possible (including standard deviation and the number of observations within the minute when available) and accompanied with their UTC time and GPS position. These data sets regroup data obtained from the [TSG] the [ACS] the [WSCD] the [BB3] the [EIMS] the [optode], the [EDM], the [SMPS], the [PAR] and the navigation data. These are available as ten distinct data sets, one for each package of sensors. Similarly, measurements made from discrete samples collected on board Tara (see Methods Section 3.3), together with quality assessment flags, are provided as six distinct data sets, one for each type of analysis ([NUT], [MTE-USC], [CARB], [FCM], [HPLC], and [CTD]). For [CARB], additional parameters of the carbonate system were calculated with CO2SYS.m v3.1.1³³ using *in situ* temperature, total alkalinity, total dissolved inorganic carbon, salinity, phosphate and silicate concentrations as inputs together with recommended parameters³⁴⁻³⁷ (K1K2 = 10; KSO4 = 3; KF = 2; BOR = 2). Data sets are available in open access at the Data Publisher for Earth & Environmental Science PANGAEA.

Combined version at the event level. A compilation of all environmental measures obtained during a given sampling event was produced by compiling the boat's sensor data available during the time-lapse of the station and measurements originating from satellite imagery (MODIS-AQUA satellite - Level 3 mapped product, 8-day average, 4 km resolution) recovered using OpenDAB protocols at <https://oceandata.sci.gsfc.nasa.gov>. The zone corresponding to the station position and date was recovered either by taking a two-pixel buffer around the given location (total zone being a 5 by 5 pixels square of 20 km side) and in order to propose an alternative measure in the inevitable case where clouds were present an alternative 12-pixels buffer was taken (total zone being a 25 by 25 pixels square of 100 km side).

The corresponding variables recovered are chlorophyll a^{38} (OCx algorithm³⁹, [Chl_Sat]; mg m^{-3}), the sea surface temperature⁴⁰ (4 μm shortwave algorithm; [T_Sat]; $^{\circ}\text{C}$), daily mean photosynthetically available radiation at the ocean surface⁴¹ ([PAR_Sat]; $\text{Einstein m}^{-2} \text{d}^{-1}$), concentration of particulate inorganic carbon⁴² ([PIC_Sat]; mol m^{-3}), concentration of particulate organic carbon⁴³ ([POC_Sat]; mol m^{-3}), the diffuse attenuation coefficient for downwelling irradiance at 490 nm⁴⁴ ([Kd490_Sat] related to light penetration in water column modified by particulate matter; m^{-1}), and the particulate backscattering coefficient at 443 nm derived from the

Name	Number of measurements	Variables	Link/doi	Reference
Raw continuous measurements				
TSG	>590 000	T, S	https://doi.org/10.1594/PANGAEA.943675	⁶⁷
EDM	>15 000	Aerosols concentration (0.25–32 µm)	https://doi.org/10.1594/PANGAEA.943694	⁸¹
		1 min and 30 min versions	https://doi.org/10.1594/PANGAEA.943691	⁸²
EIMS	>230 000	O/Ar ratio	https://doi.org/10.1594/PANGAEA.943714	⁸³
Optode	>280 000	Oxygen concentration	https://doi.org/10.1594/PANGAEA.943790	⁸²
Navigation	>1 271 000	Navigation and Meteo	https://doi.org/10.1594/PANGAEA.944365	⁸⁴
ACS	>411 000	Chla, phytoplankton size, POC	https://zenodo.org/record/6449893	⁸⁵
BB3	>350 000	Backscattering, phytoplankton carbon	https://doi.org/10.1594/PANGAEA.943793	⁸⁶
WSCD	>553 000	relative DOM fluorescence (sd, n)	https://doi.org/10.1594/PANGAEA.943739	⁸⁷
PAR	>830 000	Photosynthetically active radiations (sd, n)	https://doi.org/10.1594/PANGAEA.943740	⁸⁸
SMPS	>4600	Aerosols concentration, particle size distribution (25–685 nm), sd	https://doi.org/10.1594/PANGAEA.943856	⁸⁹
Raw discrete measurements				
NUT	849	NO ₂ , NO ₃ , PO ₄ , Si(OH) ₄	https://doi.org/10.1594/PANGAEA.944289	⁹⁰
MTE-USC	523	Fe, Zn, Cd, Ni, Cu, Pb, Mn	https://doi.org/10.1594/PANGAEA.944395	⁹¹
CARB	325	Alkalinity, Carbonates, pH, pCO ₂ , fCO ₂ , [HCO ₃ ⁻], [CO ₃ ²⁻], CO ₂ , Ω-Calcite, Ω-Aragonite	https://doi.org/10.1594/PANGAEA.944420	⁹²
FCM	1041	Pico-, Nano-, Picoplankton abundance and scattering	https://doi.org/10.1594/PANGAEA.944490	⁹³
HPLC	551	Pigment concentrations	https://doi.org/10.1594/PANGAEA.944281	⁹⁴
CTD	4246	T,S, conductivity, conductance, density, sound velocity, depth, pressure	https://doi.org/10.1594/PANGAEA.943869	⁹⁵

Table 4. Data sets providing the environmental context for future analysis and provided as raw measurements by sensors and from samples.

Garver-Siegel-Maritorena algorithm⁴⁵ ([GSM_Sat] which gives a good indication of the concentration of suspended organic and inorganic particles such as sediments in the water; m⁻¹).

This compilation of environmental data at the scale of the event was further enriched using data from reanalyzed (ie. forced with observations) operational models obtained from Copernicus Marine Services (GLOBAL_REANALYSIS_PHY_001_030⁴⁶, daily mean for sea surface height, salinity, temperature, current speeds, mixed layer depth; GLOBAL_REANALYSIS_BIO_001_029⁴⁷ daily mean for Chl a, phytoplankton carbon, O₂, NO₃, PO₄, SIOH, Fe concentrations, Primary production, pH and CO₂ partial pressure and GLOBAL_REANALYSIS_WAV_001_032-TDS⁴⁸ for sea surface waves) but also using almanach^{49,50} to calculate essential sun and moon parameters (position, rises and sets, phase, etc).

Environmental context at the granularity of samples. The environmental context of all samples collected during the Tara Pacific Expedition is available together with the provenance file in open access at Zenodo. The environmental context of each sample is provided based on environmental data sets described above for continuous and discrete measurements, as well as those generated from almanacs, satellite imagery and models.

Environmental context is provided in eleven UTF-8 encoded tab-separated-values files, all with the same structure, but each providing a different statistic: number of values (n), mean value (mean), standard deviation (stdev), 05, 25, 50, 75 and 95 percentiles (P05, P25, P50, P75, P95), lag in time (dt), i.e. difference between the collection date/time of the sample and that of the environmental context provided, lag in horizontal space (dxy), i.e. distance between the collection location of the sample and that of the environmental context provided, and lag in vertical space (dz), i.e. difference between the collection depth/altitude of the sample and that of the environmental context provided.

Missing value terms are: “nav” = not-available, i.e. the expected information is not given because it has not been collected or generated; “npr” = not-provided, i.e. the expected information has been collected or generated but it is not given, i.e. a value may be available in a later version or may be obtained by contacting the data providers; “nac” = confidential, i.e. the expected information has been collected or generated but is not available openly because of privacy concerns; “nap” = not-applicable, i.e. no information is expected for this combination of parameter, environment and/or method, e.g. depth below seabed cannot be informed for a sample collected in the water or the atmosphere

Simplified version at site level. In some cases, certain parameters were not available at specific sampling sites due to technical issues or sensor availability, however, various basin scale studies and statistical tests require a complete dataset for all sampled sites. During the Tara Pacific expedition, many parameters were concurrently measured *in-situ*, estimated from remote sensing and/or modelled. For instance, sea surface temperature was measured on the boat using the thermosalinograph included in the underway system, but also with satellite

Name	Number of measurements	Variables	Link/doi	Reference
Environmental context at the granularity of sampling events				
Inline sensors + Almanach + Copernicus + Modis Aqua (2 and 12 pixels around)	4155	all Inline data with n, sd, quartiles, local sun/moon set/rize, local zenith, nutrients, hydrology, plankton quantities, Chla, PAR, PIC, POC, T, GSM, KD490 (with n, sd, and quartiles)	https://zenodo.org/record/6445609	⁹⁶
Provenance metadata and environmental context at the granularity of samples				
Sample provenance	57859	georeference, sample unique identifier, logsheet links, environmental features and place names	https://zenodo.org/record/6299409	⁹⁷
All previous variables extracted at event level	57859	mean and std + dt, dx, dz from sampling timing, position and depth		
Environmental context at the granularity of sampling stations				
all event level variables	655	intercalibrated and combined version	https://zenodo.org/record/6474974	⁹⁸
Lagrangian Descriptors	246	Eulerian and Lagrangian diagnostics of water dynamic	https://zenodo.org/record/6453376	⁹⁹
Environmental context at the granularity of coral sampling sites				
historical heat and cold stress indicators	113	TSA, DHW, recovery time etc...	https://zenodo.org/record/6499374	¹⁰⁰
raw time series	>6000 × 113	SST at 1, 3 and 9 pixels, seasonal average, DHW, DCW		
Reefcheck bleaching occurrence	106	Bleaching (% of colony or % of population)	https://zenodo.org/record/6511406	¹⁰¹
Photo annotations				
Qualitative photographic annotations	5606 photo, 2216 colonies	identification to the genus level, algal contact (genus of algae), presence of boring organisms (type), contact with sediment,	https://zenodo.org/record/6364768	¹⁰²
Taxonomic annotations of coral diversity (CDIV) surveys	2470	18 S based taxonomic annotations, corresponding morphological annotation based on photo	https://zenodo.org/record/6327048 https://ecotaxa.obs-vlfr.fr/prj/4176	¹⁰³

Table 5. Data sets providing the provenance and the environmental context for future analysis and provided aggregated at the sample, event and site levels.

and estimated from a model. Each of these three modes of acquisition have their caveat and accuracy, however, within a certain confidence interval, missing *in-situ* data can be replaced by its remotely sensed or modelled equivalent. We provide here a simplified version at the sampling site level by replacing missing *in-situ* data by their closest and most accurate satellite or modelled equivalent. In each case, *in-situ* data was considered as the most accurate source of data, with a preference to HPLC pigments analysis followed by measurements done by the ACS, while satellite and modelled data were used only if *in-situ* data was not available. We evaluated the accuracy of ACS and of each satellite and modelled datasets by linear regressions with their *in-situ* counterparts. A bias of the modelled or satellite data was identified when the slope of the regression was different to 1 and/or an intercept was different to 0. The satellite and modelled data were forced to match the *in-situ* data by dividing by the slope and subtracting the intercept. This is the case for SST. When large bias persisted between matchups with observations, the corrected data was not used to replace missing *in-situ* data. This is the case for chl. The same approach was then applied to fill missing data with modelled values (MERCATOR-Copernicus).

A correction for the bias in the following variable was applied for SST, SSS, PO₄, and SiOH. As previously done, if large bias persisted between observations and corrected data, they were not used to replace missing *in-situ* data. This is the case for chl, NO₃, and Fe.

The [MTE] samples were sometimes sampled in the afternoon instead of the morning alongside all the other water samples, thus were located in between two sampling stations. These [MTE] samples could not be assigned to a sampling station following the criterion presented in the section 3, therefore, the missing values of the corresponding morning stations were interpolated linearly.

The same approach was used for pH measurements, with a preference from measurements provided by total carbonate system quantifications, followed by direct pH measurements and then modeled values (MERCATOR-Copernicus).

Lagrangian and Eulerian diagnostics. In order to provide a description of the dynamical properties of the water masses sampled, different Eulerian and Lagrangian diagnostics were calculated. Here, we report a general description of the information each of them provides. In the next subsection, we provide the details of how they were calculated for each station.

The following Eulerian diagnostics were calculated: Absolute velocity ([Uabs], m s⁻¹): $\sqrt{u^2 + v^2}$, where u and v are the zonal and meridional components of the horizontal velocity field used (described below); Kinetic energy ([Ekin], m²s⁻²): $0.5*(u^2 + v^2)$; Divergence ([EulerDiverg], d⁻¹): $du/dx + dv/dy$; Vorticity ([Vorticity], d⁻¹): $dv/dx - du/dy$; Okubo-Weiss ([OW], d⁻²): $s^2 - \text{vorticity}^2$, where s^2 is $(du/dx - dv/dy)^2 + (dv/dx + du/dy)^2$. If negative, it indicates that the station sampled was inside an eddy.

The following Lagrangian diagnostics were calculated: Finite-Time Lyapunov Exponents ([Ftle], d⁻¹, Shadden *et al.*, 2005): it indicates the rate of horizontal stirring, and it is a means to quantify the intensity of turbulence in a given region. FTLE are commonly used to identify Lagrangian Coherent Structures, i.e. barriers to transport. In this case, a strong FTLE value indicates a region separating water masses which were far away backward in time. Lagrangian betweenness⁵¹ ([betw], adimensional): this diagnostic draws inspiration from Lagrangian Flow Network Theory⁵². It can identify regions which act as bottlenecks for the circulation,

in that they receive waters coming from different origins, and that are then spread over several different destinations. These can represent possible hotspots driving biodiversity⁵¹. Lagrangian Divergence⁵³ ([LagrDiverg], d^{-1}). This diagnostic was calculated by integrating the Eulerian divergence along the backward trajectories. If positive, it indicates a water mass that, during the previous days, was subjected to a strong divergence, thus to a possible upwelling. If negative, it indicates a strong convergence, thus possible downwelling. Retention Time⁵⁴ ([RetentionTime], d). This diagnostic indicates how many days a water mass has spent inside an eddy in the previous period. If the water mass is outside an eddy, then its retention time is set to zero. Extraction of the Eulerian and Lagrangian diagnostics For each of the 246 stations sampled, we proceeded as follows.

We identified the water mass sampled at the given station. This was considered as a stadium shape with the two semi-circles centered on the starting and ending points of the transect, respectively. The radius of the stadium semi-circles was considered 0.1° , which is in accordance with previous studies^{51,55,56}. The stadium was filled with virtual particles separated by 0.01° .

For each virtual particle inside the stadium shape, we calculated a Eulerian or Lagrangian diagnostic (described above). The Eulerian diagnostics were extracted directly from the velocity field of the day of sampling. Concerning the Lagrangian diagnostics, these were obtained by advecting the virtual particle backward in time for an amount of time τ from the day of sampling day_S. For the Lagrangian betweenness, the advection was performed between day_S + $\tau/2$ and day_S - $\tau/2$, so that the advective time window was centered on the sampling day (details in⁵¹).

For the Lagrangian diagnostics, we used the following advective times τ : 5, 10, 15, 20, 30, and 60 days. The only exception is the retention time, which, by construction, was calculated only with the largest advective time, namely $\tau = 60$ days.

Once that, a given diagnostic (Eulerian or Lagrangian) was calculated for all the virtual particles filling the stadium shape, we calculated the mean value, and the 25, 50, and 75 percentiles. The percentiles were calculated in order to quantify the spatial variation of the diagnostic inside the stadium shape. Therefore, we associated each station with four values (mean, 25, 50, and 75 percentiles) of a given diagnostic.

Furthermore, two different velocity fields were used, which are described as follows.

Velocity fields and trajectory calculation Both the velocity fields were downloaded from E.U. Copernicus Marine Environment Monitoring Service (CMEMS, <http://marine.copernicus.eu/>). The first velocity field used was MULTIOBS_GLO_PHY_REP_015_004⁵⁷ [GlobEkmanDt]. This was produced by combining the altimetry derived geostrophic velocities and modelled Ekman surface currents. It had a spatial resolution of 0.25° and a temporal resolution of one day. The second velocity field was GLOBAL_REANALYSIS_PHY_001_030⁴⁶ [GloryS12]. It was obtained by a NEMO model assimilating altimetry and other observations. It had a spatial resolution of $1/12^\circ$ and a temporal resolution of 1 day.

Historical climate data and indices for climate variability for coral collection sites. It's becoming increasingly clear that stress resilience, in particular thermal tolerance, is shaped not only by maximum monthly mean temperatures (MMMs), but also by long-term and short-term climate variability, even at the scale of reefs^{58–60}. In order to provide an overview of past climate variability and marine heatwaves experienced by corals sampled at each site, we built a high-resolution historical dataset that spans from 2002 to each sites' sampling date. Ocean skin temperature (11 and 12 μm spectral bands longwave algorithm) was extracted from 1 km resolution level-2 MODIS-Aqua and MODIS-Terra from 2002 to the sampling date and from level-2 VIIRS-SNPP from 2012 to the sampling date. Day and night overpasses were used to maximize data recovery. Following recommendations from NASA Ocean Color (OB.DAAC), only SST products of quality 0 and 1 were used. The 9 closest pixels to the sampling sites of each scene were extracted. All the extracted pixels from the 3 platforms were then averaged daily to obtain daily SST averages and standard deviations time series for each sampling site, from 2002 to the sampling date.

Each time series was first averaged on a Julian day basis to provide a seasonal average. This yearly seasonal average was triplicated and concatenated into a 3-year seasonal cycle to apply a digital low pass filter on the middle year without generating artefacts. A digital low pass filter (filter order 3, pass band ripple 0.1; "filfilt" function in matlab) with 36 Julian days windows was applied to the concatenated time series to remove high frequency noise. The middle year was then extracted from the concatenated time series to recover the seasonal cycle. The sea surface temperature anomaly was calculated as the SST minus the seasonal cycle over the full time series. Considering the short periods of missing data (mean of the 95th percentile of the duration of consecutive days with missing data: 9.8 ± 4.1 days), the missing values in the SST and SST anomaly time series were linearly interpolated in order to calculate thermal stress indices. The SST anomaly frequency was calculated as the number of days over the past 52 weeks when the SST anomaly is greater than or equal to 1°C . Thermal stress indices relevant to coral reef health were then calculated using methodology developed for the Coral Reef Temperature Anomaly Database (CoRTAD)⁶⁰ (Table 6). Events of cold temperature accumulation were also reported to cause bleaching and mortality^{61,62}, therefore, the same set of indices were calculated for cold stress adapting the CoRTAD method, but using the minimum weekly climatologies (Table 6). Further to that, we checked for previous occurrences of bleaching events at sampled reef sites by matching island coordinates to the Reef Check dataset (reefcheck.org) obtained from Sully *et al.*^{58,63}. For each Tara Pacific island, coordinate we determined that Reef Check site that was closest (in terms of distance in km) and considered only Reef Check data that was within a 10 km circumference.

A condensed table containing single values associated with each sampling site was created extracting the minimum, maximum, sum, averages, standard deviations, and value recorded at the sampling day of each of these indices (detailed in the readme file provided with the dataset). Additional metrics of the last heating and cooling events as well as the time of recovery is also provided to represent the state of thermal stress at the day of sampling.

Name	Acronym	Description	Reference
SST daily average 9 pixels	[sst_mean_9pixel]	Daily average of the 9 closest pixels around the sampling site	
Seasonal average 9 pixels	[seasonal_average_9pixel]	Seasonal average SST calculated from 2002 to the sampling date	
SST anomaly 9 pixels	[SST_anomaly_9pixel]	SST anomaly calculated as: sst_mean_9pixel minus seasonal_average_9pixel	
SST daily average interpolated	[SST_mean_interp]	SST daily average with missing values interpolated linearly	
SST anomaly interpolated	[SST_anomaly_interp]	SST anomaly with missing values interpolated linearly	
SST anomaly frequency	[SST_anomaly_freq]	number of days in the past 52 weeks when $SST_anomaly_interp > = 1^{\circ}C$	CoRTAD
Heat Thermal Stress Anomaly	[TSA_heat]	Daily SST average interpolated minus the maximum weekly climatology	CoRTAD
TSA heat frequency	[TSA_heat_freq]	number of days in the past 52 weeks when $TSA_heat > = 1^{\circ}C$	CoRTAD
TSA degree heating week	[TSA_DHW]	sum of the past 12 weeks when TSA_heat is greater than or equal to $1^{\circ}C$	CoRTAD
TSA degree heating week frequency	[TSA_DHW_freq]	number of days in the past 52 weeks when TSA_DHW is greater than or equal to $1^{\circ}C$	CoRTAD
Cold Thermal Stress Anomaly	[TSA_cold]	Daily SST average interpolated minus the minimum weekly climatology	Custom
TSA cold frequency	[TSA_cold_freq]	number of days in the past 52 weeks when $TSA_cold < = -1^{\circ}C$	Custom
TSA degree cooling week	[TSA_DCW]	sum of the past 12 weeks when TSA_cold is lower than or equal to $-1^{\circ}C$	Custom
TSA degree cooling week frequency	[TSA_DCW_freq]	number of days in the past 52 weeks when TSA_DCW is lower than or equal to $1^{\circ}C$	Custom

Table 6. Description of historical SST values and thermal stress indices calculated following CoRTAD⁶⁰ method and modified to also represent cooling events.

Coral photographic resources and annotations. The [PHOTO] resource consists of two datasets. The first, obtained from the [SCUBA-3X10] protocol, was annotated for genus validation, gross morphological characteristics of the colony, algal contact, presence of boring organisms, sediment contact, predation, and health factors (such as presence of disease and coloration). The acquisition protocol of these annotations is described below. This dataset is also used for the description of morphotypes within each genus for taxonomic annotation in combination with genetic data. The second dataset, obtained following [SCUBA-SURVEY] protocol was used for the taxonomic annotation (as close to genus level as possible) of the coral host of the [CDIV] samples. Of a total of 2,470 CDIV samples, 1711 samples had one or more pictures associated (3,085 total pictures), 759 samples had no photos. Overall, 11,460 coral photographs were generated and annotated allowing for a permanent record of all colonies sampled. All [PHOTO] were transferred to EcoTaxa⁶⁴.

(1) Manual Annotations of *in situ* colony (CO) photos:

Photo analysis for the genus validation and environmental context was conducted using Matlab with code developed and written specifically for the Tara Pacific Expedition⁶⁵. Photos were annotated individually, and annotations were conducted from January to April 2020. To prevent observer bias, photos were randomized, and the annotator was blind to any information regarding the location or the sampling site. The analysis included 1) identification to the genus level, 2) algal contact with types of algal genus if identifiable (*Halimeda*, *Turbinaria*, *Dictyota*, *Lobophora*, Crustose Coralline Algae (CCA), *Sargassum*, *Galaxaura*, other), 3) presence of boring organisms with types if identifiable (Bivalve, *Spirobranchus*, *Tridacna*, Urchin, Other Polychaete, Sponge, and Other), 5) contact with sediment (sand), 6) presence of predation marks. Most annotations were boolean operators (yes/no) with identifications added if possible. Several indicators of coral health were also annotated such as if the coral looked unhealthy or showed tissue loss (Yes/No), coloration (light, normal, dark, or bleached), and presence of a pigmentation response (Yes/No). If a pigmentation response was present, the annotator was prompted to determine if it was trematodiasis (Yes/No). Finally, additional notes included but were not limited to the quality of the photo (blurry, poor visibility, coloration), contact with neighbouring hard or soft coral colonies, fish presence in the photograph, snail(s), or hermit crab(s) on the coral, an object in the photograph, etc.

(2) Taxonomic annotations of coral diversity (CDIV) surveys:

All images imported in EcoTaxa have been identified at the genus level by taxonomic experts, and crosslinked with genomic identification from metabarcoding based on the V9 region of the 18S rDNA. Analysis of the 18S marker aimed to generate coral host taxonomic annotations to the level of genus for every sample. The annotation was generated based on each sample's most abundant 18S sequence by aligning to the NCBI 'nt' database with taxonomic labels. A 'lowest common ancestor' approach was used when there were multiple best hits. These alignment-based annotations were verified phylogenetically (i.e. taxonomic similarity agreed with sequence similarity). More than half of the samples were not annotated at genus or better level using this approach, due to the lack of resolution of the 18S V9 marker. Where available, host taxonomic assignments were based on photo annotations. Otherwise, 18S-based annotations were used.

Technical Validation

Numerous steps of quality control were operated at different levels of acquisition to ensure good quality of the different datasets and may vary depending on the type of measurement operated and if it originates from sensors on-board or from samples.

Inline measurements, models, and satellite data validity. [PAR] measurement validity was checked by first removing physically wrong data (ie. values greater than $0.45 \mu E \text{ cm}^{-2} \text{ sec}^{-1}$ or lower than $0 \mu E \text{ cm}^{-2} \text{ sec}^{-1}$) and compared with clear sky matchup measurements from MODIS-Aqua & Terra. Comparison confirmed the good agreement between datasets but also the absence of sensor drift. Temperature and salinity were acquired by the [TSG]. The quality of the whole time series was manually checked, and the temperature validity was assessed

by comparing the temperature reading of the two sensors placed at two different places along the inline system. Potential drifts of the temperature sensor was investigated by comparing the temperature time series with satellites' sea surface temperature. Salinity measurements were intercalibrated against unfiltered seawater samples [SAL] taken every week from the surface ocean, and corrected for any observed bias. Moreover, temperature and salinity measurements were validated against Argo floats data collocated with Tara. The [ACS] absorption and attenuation signal due to dissolved matter, drift, and biofouling were estimated between two filter events by interpolating filtered water absorption and attenuation following the shape of the [fdom] from the [WSCD], when available. This method improves data quality in case of strong variation of dissolved matter absorption that the frequency of filter event would not capture properly (e.g. approaching coastal waters or entering a lagoon). When [fdom] data was not available, the filtered absorption and attenuation were linearly interpolated between filter events before being removed from the total absorption and attenuation. From November 13, 2016 to May 6, 2017, the [BB3] was located upstream of the switch system, thus measured total (non-filtered) water all the time. During this period, the volume scattering coefficient of seawater was removed from the raw data counts to obtain the particulate backscattering coefficient [bbp]. The biofouling and instrument drift were estimated comparing values before and after each cleaning events. The biofouling was estimated between two cleaning events by fitting an exponential or linear model to the raw data before removing it from the signal. We advocate to use this period with caution as the data was corrected with theoretical assumptions (i.e. pure seawater scattering and linear or exponential biofouling) that may differ from reality. From May 7th 2017 to the end of the expedition, the [BB3] was located downstream of the filter-switch system so that, like for the [ACS] processing, the biofouling signal could be estimated and removed between two filter events and [bbp] quality improved. The correspondence between total particulate scattering [bp] estimated from the [ACS] and [bbp] was investigated for the whole expedition. [bbp] values were discarded when [bbp]/[bp] was unusually low (< 0.002 ; see range of [bbp]/[bp] in natural waters⁶⁶). A similar modelling and correction for biofouling than the one performed for the [BB3] was applied to the [WSCD] data. The [PAR], [TSG], [BB3], [ACS], and [WSCD] data were processed following the last recommendations for processing inline²⁴, using custom software available at <https://github.com/OceanOptics/InLineAnalysis>. The entire time series of measurement were automatically QC to remove artifacts and manually checked and QC for obviously inaccurate measurements due to saturated sensor, low flow rate, bubbles, or poor filtered seawater measurements. The full processing and QC procedure and reports could be accessed together with each dataset.

Sample measurements technical validation. For nutrients [NUT] samples a quality check was done in several steps. First a visual inspection was done to determine if samples were overfilled or not frozen vertically which may induce sample leakage during the frosting procedure. Secondly any readings too close to detection limits or when duplicate measurements differed by more than 10% were flagged. In this last case, when the difference between two values of the same sample is greater than 10%, it is considered that the high value is not acceptable and is not reported. Finally, the overall quality of the dataset was established by comparing measurements values with Copernicus Marine Services modelling outputs.

For trace metals ([MTE-USC]), any samples in which concentrations were close to detection limits were flagged. A standard produced by the GEOTRACES program (coastal surface seawater standard) was included in each sample run. If the metal concentrations of the standard were outside the GEOTRACES community consensus values, the sample run was rejected. Trace metal concentrations had an average error of 5%.

[HPLC] samples were analysed as described in Ras *et al.* 2008. All pigments peaks were inspected and quality controlled as good, acceptable or qualitative. Any measurements below detection limits were disregarded.

[FCM] samples were analysed with a FACS Canto II Flow Cytometer equipped with a 488 nm laser⁶⁷ and every measurement where cell populations were either complicated, needed manual curation or were impossible were flagged.

Nets collection validity. To estimate the technical validity of the different nets collection we analysed the raw abundance of living organisms collected conjointly by the [HSN-NET-300] and [MANTA-NET-300] at the same stations, but sequentially in time. Indeed [MANTA-NET-300] is operated at different speeds (3 knots maximum) compared to [HSN-NET-300] (9 knots maximum) and therefore were not deployed simultaneously. Manta nets are commonly used and recognized as a reference type of net while investigating surface plankton⁶⁸⁻⁷⁰ and we therefore used a set of 24 stations where both were deployed concurrently to estimate the efficiency of the [HSN-NET-300]. For this [F300] samples collected by both nets were imaged using the ZooScan⁷¹ to obtain images of each object collected. Images were then transferred to EcoTaxa⁶⁴ and sorted taxonomically to the deepest taxonomic level possible. All results were used to calculate the normalized biovolume size spectra⁷² (NBSS) of living organisms for both nets, which is an analogue to abundance per size categories. This NBSS spectra allows investigating the potential under- or over-sampling while investigating it over various sizes of organisms. The NBSS of both nets were giving about the same order of magnitudes of abundances (Fig. 4A) and when inspected along the size spectra between pairs of observations (Fig. 4B) they did not differ largely from 1:1 in 13 cases over the different deployments. A large variability between nets could however be observed at a few stations which could possibly be caused by local plankton patchiness⁷³ resulting in more variability for [HSN-NET-300] and less for [MANTA-NET-300] due to larger sampling volume. Overall, we can conclude that [HSN-NET-300] and [MANTA-NET-300] are collecting plankton with a relatively similar efficiency even if the larger sampling volume of [MANTA-NET-300] allows a better collection of larger, rare, organisms, as seen from spectra extending to larger sizes (Fig. 4A). Nevertheless, these results show that the use of [HSN-NET-300] may be really useful for underway zooplankton sampling in the situations when it is not possible to stop the ship for regular sampling or on ships of opportunity.

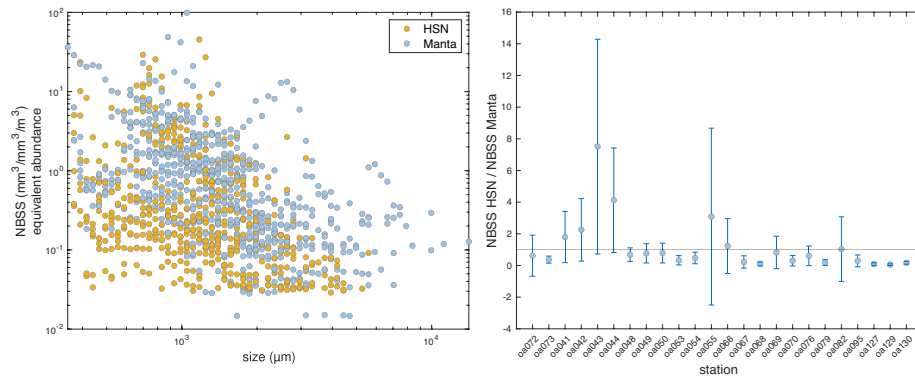


Fig. 4 Technical validation of net sampling. Comparison of normalized biovolume size spectra (NBSS) of living organisms sampled with [HSN-NET-300] and [MANTA-NET-300] over a set of 24 stations where both were deployed together. From both NBSS, a sampling efficiency of the HSN net compared to the MANTA net was calculated as a mean and standard deviation over all the size classes considered.

Overall biogeochemical data validity. To assess the overall quality and homogeneity of the collected environmental parameters, we conducted a quick multivariate exploration of the dataset to compare it with known biogeography of biogeochemical provinces^{74,75} and their associated biogeochemical signatures. For this, we first used data simplified at the site version (see section 4 of Data records), selected only datasets providing a full overview over the geographical range of the expedition, used a box-cox transformation and centred-reduced each variable to equally consider those. This dataset was then analysed through a PCA analysis (Fig. 5). The 3 first components of the PCA analysis were recovered to code for a RGB (red, green, blue) color-coding of each station and better visualize the biogeochemical signature of the station on a map. Finally, those were compared with known biogeochemical provinces extracted from⁷⁵. Despite the different temporal resolution between instantaneous sampling and biogeochemical provinces representing a consensus over several years and seasons, we can see that the main biogeochemical provinces (and associated macroscale oceanic features) as well as their progressive boundaries are well captured by our sampling scheme. Among the notable features, the western Pacific coast of Americas are marked by a strong upwelling signature (with high amount of nutrients and trace metals), the southern Pacific gyre with a high salinity but a low iron and silicate concentration, the central Pacific zone is characterized by high temperature, light and sea surface height, small phytoplankton size (high gamma), with low chlorophyll *a* and low NO_3 and trace metals (Ni, Cu, Zn, Pb or Cu) concentrations, with the exception of the few stations centred on the equator which clearly display some indicators of local upwelling such as those potentially created by the equatorial upwelling. This first overview clearly shows correspondence with known features related to nutrients and nutrient limitation of plankton, trace metals or even global biogeochemistry^{76–78} and further shows that the sampling scheme used allowed to sample corals and plankton across a large variety of environmental constraints either on oceanographic, climatic or chemical aspects. The same analysis repeated only using sites realized around islands further confirms this large variety of environmental constraints (Fig. 6). To evaluate the variety of the past temperature history, and notably the impact of past seasonality and heat/cold waves, we further reproduced this analysis using historical temperature and heat/cold waves experienced on coral sites. However, since temperature anomalies and their accumulated degree cooling weeks (DCW) could be negative, only a basic normalization of data was made since box-cox normalization is not suited for negative values. The first axis of the PCA separate islands that suffered intense and recurrent heat-waves (positive values) from those that rather experienced cold-waves (negative values) while the second axis separate cold and highly seasonal islands (positive values) from islands with warm environments with low seasonality (negative values). This analysis further confirms that the selected location also displays a full variety of past history of temperature and heat-waves but also reflects known geographical patterns of bleaching events^{58,79}.

Usage Notes

We recommend paying close attention to the various quality flags provided with the raw datasets to avoid using lower quality data if needed. Similarly, to provide the more complete set of observations for each sample, we provided the lag in time (dt), as well as distance in horizontal (dxy) and vertical (dz) space, between the collection timing, latitude/longitude and depth/altitude of the sample and that of the environmental context provided. Depending on the scientific question, future users are encouraged to carefully define reasonable time lag and distances to consider in their study, to avoid including unrealistic associations between samples. Moreover, we extracted contextual data at the event level to simplify the data extraction task. We also provide simplified version at the site level by combining and cross-calibrating all similar variables (e.g. using different sources of SST data to fill gaps of missing data and obtain one merged SST variable). We prioritised observations

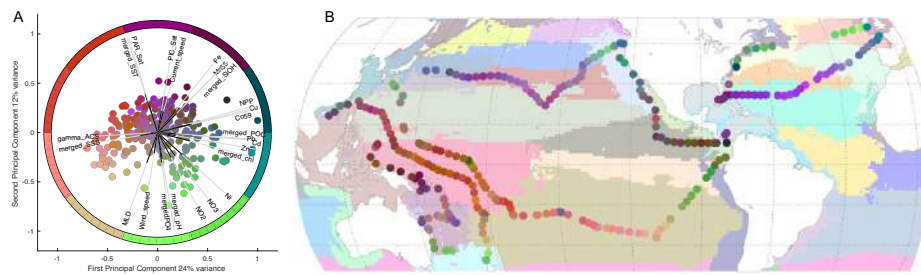


Fig. 5 Technical validation of the main hydrological and biogeochemical environmental variables compared with biogeochemical provinces as extracted from⁷⁵. (A) Environmental data were normalized through a box-cox normalization and analysed through a PCA analysis to better display their typical environmental signature. The position of each station in the 3 first axes of the PCA were further used to provide a red blue green color-coding, allowing to (B) project their environmental signature on a map and compare it with known biogeochemical provinces.

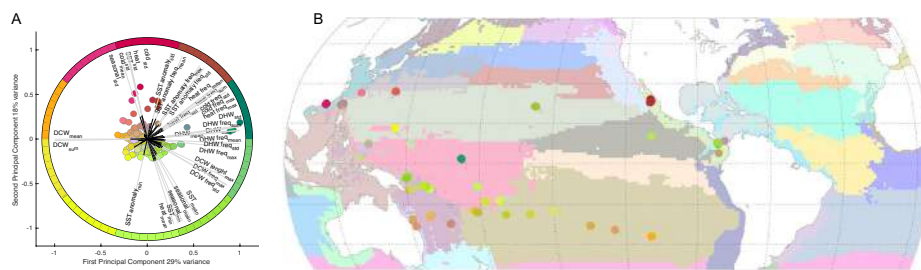


Fig. 6 Technical validation of the historical SST heatwaves and cold waves parameters compared with biogeochemical provinces as extracted from⁷⁵. (A) Environmental data were normalized and analysed through a PCA analysis to better display their typical environmental signature. Only 50% of the variables having the more influence on the analysis are displayed here. The position of each station in the 3 first axes of the PCA were further used to provide a red blue green color-coding, allowing to (B) project their environmental signature on a map and compare it with known biogeochemical provinces.

originating from *in-situ* samples over satellite data, and over modelled data (MERCATOR), and evaluated their correspondence by linear regressions. Potential biases of satellite and modelled data in comparison to *in-situ* data were corrected applying the slope and intercept of their linear regression to force satellite and modelled data to best match *in-situ* data. Similarly, we also chose to interpolate some environmental variables that were sampled only few hours before or after the site itself to maximize data recovery for each sampling station. We acknowledge merging different sources of data can introduce differences in variance depending on the source of data used, therefore, we encourage the user to cautiously evaluate the relevance of this merged dataset for their study. Considering the intrinsic heterogeneity of variance between the different datasets, and their potential non-normal distribution, we recommend using appropriate normalisation methods before any multivariate statistical analysis. Here we chose to use box-cox transformation and centred-reduced each variable.

In this version of the dataset the satellite data used is 8-days averages while the *in-situ* measurements are instantaneous measurements of optical properties averaged over the station sampling period. The 8-days averaging tends to attenuate extreme values and reduces the potential differences between stations. While suited for macro-ecological processes which depend on large temporal and spatial variations of their environment, the use of 8-day average satellite products could be inaccurate to study shorter life cycles of the pico-, nano and micro-plankton.

Moreover, phytoplankton can adjust their light harvesting pigment concentrations according to light exposure, nutrient availability and temperature. These variations are negligible over periods shorter than a day but can become significant over 3–5 days⁸⁰, referencetherein. Therefore, we advise the users to cautiously use the merged bio-optical variables of this dataset and to verify its compatibility with the research question and potentially replace this 8-day average with shorter time observations if available. As presented in section “3.3. Continuous measurements”, the [poc] was estimated from the underway system, both using the measured [cp]²⁸,

and [bbp]²⁹. The [BB3] sensor have a low signal-to-noise ratio due to its high sensitivity to bubbles in the water line and to accumulation of particles in the sensor, therefore, the [poc] estimated from the [BB3] was used to fill the missing [poc] estimated from the [ACS]. When the [bbp] from the [BB3] was used to estimate [POC], the [bbp] values from the 470 nm wavelength were prioritized over the 532 nm wavelength and 650 nm wavelength and the same merging method was applied to correct for bias between [poc] estimated from the [ACS] and the [BB3], and between wavelength of the [BB3].

Code availability

The different codes used to process the different datasets are indicated within the text and are repeated here and includes:

- Inline optical processing (<https://github.com/OceanOptics/InLineAnalysis>)
- Satellite products used^{38,40–45}
- Mercator products^{46–48,57} used.
- Astronomical almanac to calculate sun/moon position and day-nights parameters from sites positions and time^{49,50}.
- Additional parameters of the carbonate system were calculated with CO2SYS.m v3.1.1³³ using *in situ* temperature, total alkalinity, total dissolved inorganic carbon, salinity, phosphate and silicate concentrations as inputs together with recommended parameters^{34–37} (K1K2 = 10; KSO4 = 3; KF = 2; BOR = 2).
- EcoTaxa⁶⁴ server github (<https://github.com/ecotaxa/ecotaxa>).
- EcoTaxa data processing (<https://github.com/ecotaxa/ecotaxatoolbox>)
- Morphological qualitative annotations⁵⁵.

Received: 24 May 2022; Accepted: 10 October 2022;

Published online: 01 June 2023

References

1. GCRMN. *GCRMN Status of Coral Reefs of the World: 2020*. Edited by: David Souter, Serge Planes, Jérémy Wicquart, Murray Logan, David Obura and Francis Staub. (2021).
2. Hughes, T. P. *et al.* Coral reefs in the Anthropocene. *Nature* **546**, 82–90 (2017).
3. Pogoreutz, C. *et al.* The coral holobiont highlights the dependence of cnidarian animal hosts on their associated microbes. In T. C. G. Bosch & M. G. Hadfield (Eds.), *Cellular Dialogues in the Holobiont* (pp. 91–118). CRC Press. <https://doi.org/10.1201/9780429277375-7> (2020)
4. Gove, J. M. *et al.* Near-island biological hotspots in barren ocean basins. *Nature communications* **7** (2016).
5. Messié, M. *et al.* The delayed island mass effect: How islands can remotely trigger blooms in the oligotrophic ocean. *Geophysical Research Letters* **47**, e2019GL085282 (2020).
6. Pickett, S. T. Space-for-time substitution as an alternative to long-term studies. in *Long-term studies in ecology* 110–135 (Springer, 1989).
7. Darwin, C. *The Structure and Distribution of Coral Reefs*, London (1874). *First published* (1842).
8. Reygondeau, G. Current and future biogeography of exploited marine groups under climate change. in *Predicting future oceans* 87–101 (Elsevier, 2019).
9. Ibarbalz, F. *et al.* Global trends of marine plankton diversity across kingdoms of life. *Cell*, <https://doi.org/10.1016/j.cell.2019.10.008> (2019).
10. Laursen, L. Spain's ship comes. *Nature* **475**, 16–17 (2011).
11. Gross, L. Untapped bounty: sampling the seas to survey microbial biodiversity. *PLoS Biology* **5**, e85 (2007).
12. Karsenti, E. *et al.* A Holistic Approach to Marine Eco-Systems Biology. *PLoS Biology* **9**, 1–5 (2011).
13. Planes, S. *et al.* The Tara Pacific expedition—A pan-ecosystemic approach of the “-omics” complexity of coral reef holobionts across the Pacific Ocean. *PLoS Biology* **17**, e3000483 (2019).
14. Gorsky, G. *et al.* Expanding Tara Ocean protocols for underway, eco-systemic sampling strategy of surface ocean/atmosphere plankton during Tara Pacific expedition (2016–18). *Frontiers in Marine Sciences* (under press).
15. Pesant, S. *et al.* Open science resources for the discovery and analysis of Tara Oceans data. *Scientific data* **2**, 150023 (2015).
16. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018–160018 (2016).
17. Huang, D. & Roy, K. The future of evolutionary diversity in reef corals. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**, 20140010 (2015).
18. Nemzer, B. & Dickson, A. The stability and reproducibility of Tris buffers in synthetic seawater. *MARINE CHEMISTRY* **96**, 237–242 (2005).
19. John, S. G. *et al.* A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environmental microbiology reports* **3**, 195–202 (2011).
20. Colin, S. *et al.* Quantitative 3D-imaging for cell biology and ecology of environmental microbial eukaryotes. *Elife* **6**, e26066 (2017).
21. Sieracki, C. K., Sieracki, M. E. & Yentsch, C. S. An imaging-in-flow system for automated analysis of marine microplankton. *Marine Ecology Progress Series* **168**, 285–296 (1998).
22. Helm, R. R. The mysterious ecosystem at the ocean's surface. *PLoS Biology* **19**, e3001046 (2021).
23. Flores, J. M. *et al.* Tara Pacific Expedition's Atmospheric Measurements of Marine Aerosols across the Atlantic and Pacific Oceans: Overview and Preliminary Results. *Bull. Amer. Meteor. Soc.* **101**, E536–E554 (2019).
24. Slade, W. H. *et al.* Underway and moored methods for improving accuracy in measurement of spectral particulate absorption and attenuation. *Journal of Atmospheric and Oceanic Technology* **27**, 1733–1746 (2010).
25. Boss, E. *et al.* The characteristics of particulate absorption, scattering and attenuation coefficients in the surface ocean; Contribution of the Tara Oceans expedition. *Methods in Oceanography* **7**, 52–62 (2013).
26. Houskeeper, H. F., Draper, D., Kudela, R. M. & Boss, E. Chlorophyll absorption and phytoplankton size information inferred from hyperspectral particulate beam attenuation. *Applied optics* **59**, 6765–6773 (2020).
27. Chase, A. *et al.* Decomposition of *in situ* particulate absorption spectra. *Methods in Oceanography* **7**, 110–124 (2013).
28. Gardner, W. D., Mishonov, A. V. & Richardson, M. J. Global POC concentrations from *in-situ* and satellite data. *Deep Sea Research Part II: Topical Studies in Oceanography* **53**, 718–740 (2006).
29. Cetinić, I. *et al.* Particulate organic carbon and inherent optical properties during 2008 North Atlantic Bloom Experiment. *Journal of Geophysical Research: Oceans* **117** (2012).

30. Graff, J. R. *et al.* Analytical phytoplankton carbon measurements spanning diverse ecosystems. *Deep Sea Research Part I: Oceanographic Research Papers* **102**, 16–25 (2015).
31. Boss, E., Twardowski, M. & Herring, S. Shape of the particulate beam attenuation spectrum and its inversion to obtain the shape of the particulate size distribution. *APPLIED OPTICS* **40**, 4885–4893 (2001).
32. Bourdin, G. *et al.* Sampling events from the Tara Pacific Expedition 2016–2018. *PANGAEA* <https://doi.org/10.1594/PANGAEA.944548> (2022).
33. Sharp, J. D. *et al.* CO2SYSv3 for MATLAB. *Zenodo* <https://doi.org/10.5281/zenodo.4546015> (2021).
34. Lueker, T. J., Dickson, A. G. & Keeling, C. D. Ocean pCO₂ calculated from dissolved inorganic carbon, alkalinity, and equations for K₁ and K₂: validation based on laboratory measurements of CO₂ in gas and seawater at equilibrium. *Marine Chemistry* **70**, 105–119 (2000).
35. Waters, J. F. & Millero, F. J. The free proton concentration scale for seawater pH. *Marine Chemistry* **149**, 8–22 (2013).
36. Perez, F. F. & Fraga, F. Association constant of fluoride and hydrogen ions in seawater. *Marine Chemistry* **21**, 161–168 (1987).
37. Lee, K. *et al.* The universal ratio of boron to chlorinity for the North Pacific and North Atlantic oceans. *Geochimica et Cosmochimica Acta* **74**, 1801–1811 (2010).
38. NASA Ocean Biology Processing Group. *MODIS-Aqua Level 3 Mapped Chlorophyll Data Version R2018.0* <https://doi.org/10.5067/AQUA/MODIS/L3M/CHL/2018> (2017).
39. Hu, C., Lee, Z. & Franz, B. Chlorophyll algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. *Journal of Geophysical Research: Oceans* **117** (2012).
40. NASA/JPL. *MODIS Aqua Level 3 SST Thermal IR 8 Day 4km Daytime V2019.0*, <https://doi.org/10.5067/MODSA-8D4D9> (2020).
41. NASA Ocean Biology Processing Group. *MODIS-Aqua Level 3 Mapped Photosynthetically Available Radiation Data Version R2018.0* <https://doi.org/10.5067/AQUA/MODIS/L3M/PAR/2018> (2017).
42. NASA Ocean Biology Processing Group. *MODIS-Aqua Level 3 Mapped Particulate Inorganic Carbon Data Version R2018.0* <https://doi.org/10.5067/AQUA/MODIS/L3M/PIC/2018> (2017).
43. NASA Ocean Biology Processing Group. *MODIS-Aqua Level 3 Mapped Particulate Organic Carbon Data Version R2018.0* <https://doi.org/10.5067/AQUA/MODIS/L3M/POC/2018> (2017).
44. NASA Ocean Biology Processing Group. *MODIS-Aqua Level 3 Mapped Downwelling Diffuse Attenuation Coefficient Data Version R2018.0* <https://doi.org/10.5067/AQUA/MODIS/L3M/KD/2018> (2017).
45. NASA Ocean Biology Processing Group. *MODIS-Aqua Level-3 Mapped Garver, Siegel, Maritorena Model (GSM) Data Version R2018.0* <https://doi.org/10.5067/AQUA/MODIS/L3M/GSM/2018> (2017).
46. Mercator Ocean International. *Global Ocean Physics Reanalysis. E.U. Copernicus Marine Service Information* <https://doi.org/10.48670/moi-00021> (2018).
47. Mercator Ocean International. *Global ocean biogeochemistry hindcast. E.U. Copernicus Marine Service Information* <https://doi.org/10.48670/moi-00019> (2018).
48. Mercator Ocean International. *Global Ocean Waves Reanalysis WAVEYRS. E.U. Copernicus Marine Service Information* <https://doi.org/10.48670/MOI-00022> (2019).
49. Mahooti, M. Sun/Moon Rise/Set, MATLAB Central File Exchange. (2018).
50. Ofek, E. MAAT: MATLAB Astronomy and Astrophysics Toolbox. Astrophysics Source Code Library, record ascl:1407.005. (2014).
51. Ser-Giacomi, E. *et al.* Lagrangian betweenness as a measure of bottlenecks in dynamical systems with oceanographic examples. *Nature Communications* **12**, 4935 (2021).
52. Ser-Giacomi, E., Rossi, V., López, C. & Hernandez-Garcia, E. Flow networks: A characterization of geophysical fluid transport. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **25**, 036404 (2015).
53. Hernández-Carrasco, I., Orfila, A. & Rossi, V. & Garçon, V. Effect of small scale transport processes on phytoplankton distribution in coastal seas. *Scientific Reports* **8**, 8613 (2018).
54. d'Ovidio, F., De Monte, S., Penna, A. D., Cotté, C. & Guinet, C. Ecological implications of eddy retention in the open ocean: a Lagrangian approach. *Journal of Physics A: Mathematical and Theoretical* **46**, 254023 (2013).
55. Baudena, A. *et al.* Fine-scale structures as spots of increased fish concentration in the open ocean. *Scientific Reports* **11**, 15805 (2021).
56. Chambault, P. *et al.* Swirling in the ocean: Immature loggerhead turtles seasonally target old anticyclonic eddies at the fringe of the North Atlantic gyre. *Progress in Oceanography* **175**, 345–358 (2019).
57. Mercator Ocean International. *Global Total Surface and 15m Current (COPERNICUS-GLOBCURRENT) from Altimetric Geostrophic Current and Modeled Ekman Current Reprocessing. E.U. Copernicus Marine Service Information* <https://doi.org/10.48670/moi-00050> (2018).
58. Sully, S., Burkepile, D. E., Donovan, M. K., Hodgson, G. & van Woesik, R. A global analysis of coral bleaching over the past two decades. *Nature Communications* **10**, 1264 (2019).
59. Voolstra, C. R. *et al.* Standardized short-term acute heat stress assays resolve historical differences in coral thermotolerance across microhabitat reef sites. *Global Change Biology* **26**, 4328–4343 (2020).
60. Saha, K. *et al.* The Coral Reef Temperature Anomaly Database (CoRTAD) Version 6 - Global, 4 km Sea Surface Temperature and Related Thermal Stress Metrics for 1982 to 2019. *NOAA National Centers for Environmental Information* <https://doi.org/10.25921/ffw7-cs39> (2019).
61. Lirman, D. *et al.* Severe 2010 Cold-Water Event Caused Unprecedented Mortality to Corals of the Florida Reef Tract and Reversed Previous Survivorship Patterns. *PLOS ONE* **6**, e23047 (2011).
62. González-Espinosa, P. C. & Donner, S. D. Predicting cold-water bleaching in corals: role of temperature, and potential integration of light exposure. *Mar Ecol Prog Ser* **642**, 133–146 (2020).
63. InstituteForGlobalEcology. <https://github.com/InstituteForGlobalEcology/Coral-bleaching-a-global-analysis-of-the-past-two-decades/blob/7925b8fd550065a7a7c92ed45f0d4133ab7756f/Reef%20Check%20Data%20Raw.csv>.
64. Picheral, M., Colin, S. & Irissou, J.-O. EcoTaxa, a tool for the taxonomic classification of images. <http://ecotaxa.obs-vlfr.fr> (2017).
65. McMind, R. rmcinds/tara_pacific_in_situ_photos: 2020_as_used, *Zenodo*, <https://doi.org/10.5281/zenodo.6286316> (2022).
66. Twardowski, M. S. *et al.* A model for estimating bulk refractive index from the optical backscattering ratio and the implications for understanding particle composition in case I and case II waters. *Journal of Geophysical Research: Oceans* **106**, 14129–14142 (2001).
67. Reverdin, G. *et al.* Temperature and salinity data collected using thermosalinograph [TSG] during the Tara Pacific Expedition 2016–2018, *PANGAEA*, <https://doi.org/10.1594/PANGAEA.943675> (2022).
68. Oozeki, Y. Comparison of catch efficiencies between the Manta net and surface ring net for sampling larvae and juveniles of Pacific saury, *Cololabis saira*. *Bulletin of the Japanese Society of Fisheries Oceanography* **64**, 18–24 (2000).
69. Karlsson, T. M., Kärrman, A., Rotander, A. & Hassellöv, M. Comparison between manta trawl and *in situ* pump filtration methods, and guidance for visual identification of microplastics in surface waters. *Environmental science and pollution research* **27**, 5559–5571 (2020).
70. Eriksen, M. *et al.* Microplastic sampling with the AVANI trawl compared to two neuston trawls in the Bay of Bengal and South Pacific. *Environmental Pollution* **232**, 430–439 (2018).
71. Gorsky, G. *et al.* Digital zooplankton image analysis using the ZooScan integrated system. *Journal of Plankton Research* **32**, 285–303 (2010).

72. Platt, T. & Denman, K. Organisation in the pelagic ecosystem. *Helgoländer Wissenschaftliche Meeresuntersuchungen* **30**, 575–581 (1977).
73. Robinson, K. L., Sponaugle, S., Luo, J. Y., Gleiber, M. R. & Cowen R. K. Big or small, patchy all: Resolution of marine plankton patch structure at micro- to submesoscales for 36 taxa. *Science Advances* **7**, eabk2904.
74. Longhurst, A. R. *Ecological geography of the sea*. (Academic Press, 1998).
75. Reygondeau, G. *et al.* Dynamic biogeochemical provinces in the global ocean. *Global Biogeochemical Cycles* **27**, 1046–1058 (2013).
76. Longhurst, A. R. Chapter 5 - NUTRIENT LIMITATION: THE EXAMPLE OF IRON. in *Ecological Geography of the Sea (Second Edition)* (ed. Longhurst, A. R.) 71–87, <https://doi.org/10.1016/B978-012455521-1/50006-1> (Academic Press, 2007).
77. Moore, C. *et al.* Processes and patterns of oceanic nutrient limitation. *Nature geoscience* **6**, 701 (2013).
78. Ustic Lucas, J. *et al.* Metagenomic analysis reveals global-scale patterns of ocean nutrient limitation. *Science* **372**, 287–291 (2021).
79. Sully, S., Hodgson, G. & van Woessik, R. Present and future bright and dark spots for coral reefs through climate change. *Global Change Biology* n/a (2022).
80. Tomkins, M., Martin, A. P., Nurser, A. J. G. & Anderson, T. R. Phytoplankton acclimation to changing light intensity in a turbulent mixed layer: A Lagrangian modelling study. *Ecological Modelling* **417**, 108917 (2020).
81. Flores, J. M. *et al.* Aerosol concentration and size distribution (0.25–32 µm) measured with an optical particle counter during the Tara Pacific Expedition 2016–2018, PANGAEA, <https://doi.org/10.1594/PANGAEA.943694> (2022).
82. Flores, J. M. *et al.* Aerosol concentration and size distribution (0.25–32 µm) measured with an optical particle counter at a 30min granularity during the Tara Pacific Expedition 2016–2018, PANGAEA, <https://doi.org/10.1594/PANGAEA.943691> (2022).
83. Lin, Y. *et al.* Oxygen to Argon ratio measured with Equilibrator Inlet Mass Spectrometer [EIMS] during the Tara Pacific Expedition 2016–2018, PANGAEA, <https://doi.org/10.1594/PANGAEA.943714> (2022).
84. Bourdin, G. *et al.* Navigation and meteorological data collected during the Tara Pacific Expedition 2016–2018, PANGAEA, <https://doi.org/10.1594/PANGAEA.944365> (2022).
85. Bourdin, G. *et al.* Chlorophyll a concentration, particulate organic carbon, and particle mean size index [γ ; 0.2–20 µm] measured using an hyperspectral spectrophotometer [ACS, Wetlabs] during the Tara Pacific Expedition 2016–2018, PANGAEA, <https://doi.org/10.5281/zenodo.6449893> (2022).
86. Bourdin, G. *et al.* Particles Volume scattering, backscattering, POC, phytoplankton carbon, using backscattering sensor [BB3] (WETLabs ECO-BB3) during the Tara Pacific Expedition 2016–2018, PANGAEA, <https://doi.org/10.1594/PANGAEA.943793> (2022).
87. Bourdin, G. *et al.* Relative dissolved organic matter DOM fluorescence, using CDOM fluorometer [WSCD] (WETLabs) during the Tara Pacific Expedition 2016–2018, PANGAEA, <https://doi.org/10.1594/PANGAEA.943739> (2022).
88. Bourdin, G. *et al.* Photosynthetically Active Radiation using PAR sensor (Biospherical Instruments Inc. QCR-2150) during the Tara Pacific Expedition 2016–2018, PANGAEA, <https://doi.org/10.1594/PANGAEA.943740> (2022).
89. Flores, J. M. *et al.* Aerosol concentration and size distribution (25–685 nm) measured with a scanning mobility particle sizer [SMPS] during the Tara Pacific Expedition 2016–2018, PANGAEA, <https://doi.org/10.1594/PANGAEA.943856> (2022).
90. Pujó-Pay, M., Conan, P. & Ghiglione, J.F. Nutrients collected during the Tara Pacific Expedition 2016–2018, PANGAEA, <https://doi.org/10.1594/PANGAEA.944289> (2022).
91. Kelly, R. L. *et al.* Metal data collected during the Tara Pacific Expedition 2016–2018, PANGAEA, <https://doi.org/10.1594/PANGAEA.944395> (2022).
92. Douville, E. *et al.* Seawater carbonate chemistry dataset collected during the Tara Pacific Expedition 2016–2018, PANGAEA, <https://doi.org/10.1594/PANGAEA.944420> (2022).
93. Marie, D. & Romac, S. Flow cytometry data of phytoplankton, bacteria and viruses obtained for samples collected during the Tara Pacific Expedition 2016–2018, PANGAEA, <https://doi.org/10.1594/PANGAEA.944490> (2022).
94. Dimier, C., Ras, J. & Uitz, J. Pigment concentration database of the sea surface layer collected during the Tara Pacific Expedition 2016–2018, PANGAEA, <https://doi.org/10.1594/PANGAEA.944281> (2022).
95. Bourdin, G. *et al.* CTD dataset collected during the Tara Pacific Expedition 2016–2018 using a castaway CTD, PANGAEA, <https://doi.org/10.1594/PANGAEA.943869> (2022).
96. Lombard, F. *et al.* Environmental data at the sampling event level collected with Inline instruments, almanach, models and satellites during the Tara Pacific Expedition 2016–2018, Zenodo, <https://doi.org/10.5281/zenodo.6445609> (2022).
97. Pesant, S. *et al.* Tara Pacific samples provenance and environmental context - version 2, Zenodo, <https://doi.org/10.5281/zenodo.6299409> (2020).
98. Bourdin, G. *et al.* Environmental context observed during the Tara Pacific Expedition 2016–2018, simplified version at site level, Zenodo, <https://doi.org/10.5281/zenodo.6474974> (2022).
99. Baudena, A., Da Silva, O. & Lombard, F. Eulerian and Lagrangian diagnostics of the dynamical properties of the water masses sampled during the Tara Pacific Expedition 2016–2018, Zenodo, <https://doi.org/10.5281/zenodo.6453376> (2022).
100. Bourdin, G. *et al.* Historical Sea Surface Temperature (SST) data and thermal stress indices of the Tara Pacific Expedition's coral reef sampling sites, from May 1st 2002 to August 31st 2018, Zenodo, <https://doi.org/10.5281/zenodo.6499374> (2022).
101. Woolstra, C. R. *et al.* TARA Pacific Bleaching Prevalence of Sampling Sites (Islands), Zenodo, <https://doi.org/10.5281/zenodo.6511406> (2022).
102. Clampitt, M. *et al.* Tara Pacific Qualitative Photo Annotations, Zenodo, <https://doi.org/10.5281/zenodo.6364768> (2022).
103. Hume, B. C. C. *et al.* TARA Pacific CDIV cnidarian host taxonomic annotation release version 1_1, Zenodo, <https://doi.org/10.5281/zenodo.6327048> (2022).

Acknowledgements

We are keen to thank the commitment of the people and the following institutions for their financial and scientific support that made this singular expedition possible: CNRS, PSL, CSM, EPHE, Genoscope/CEA, Inserm, Université Côte d'Azur, ANR, agnès b., UNESCO-IOC, the Veolia Environment Foundation, Région Bretagne, Billerudkorsnas, Amerisource Bergen Company, Lorient Agglomération, Smilewave, Oceans by Disney, the Prince Albert II de Monaco Foundation, L'Oréal, Biotherm, France Collectivités, Fonds Français pour l'Environnement Mondial (FFEM), the Ministère des Affaires Européennes et Etrangères, the Museum National d'Histoire Naturelle, Etienne BOURGOIS, the Tara Ocean Foundation's teams and crew. Tara Pacific would not exist without the continuous support of the participating institutes. The authors also particularly thank Serge Planes, Denis Allemand and the Tara Pacific consortium. This study has been conducted using E.U. Copernicus Marine Service Information and Mercator Ocean products. We acknowledge funding from the Investissement d'avenir project France Génomique (ANR-10-INBS-09). FL is supported by Sorbonne Université, Institut Universitaire de France and the Fondation CA-PCA. The in-line and atmospheric optics dataset was collected and analysed with support from NASA Ocean Biology and Biogeochemistry program under grants NNX13AE58G and NNX15AC08G and the HPLC processing under NSF award 2025402 to University of Maine. JMF is supported by a research grant from Scott Eric Jordan and the Bernard & Norton Wolf Family Foundation. NCo was supported by a grant from

the Simons Foundation/SFARI (544236). Work from MC, RMM, ER, DF, PF, EG is supported by the French Government (National Research Agency, ANR) through the grant “Coralgene” ANR-17-CE02-0020 as well as the “Investments for the Future” programs LABEX SIGNALIFE ANR-11-LABX-0028 and IDEX UCAJedi ANR-15-IDEX-01. NC and YL were supported by the “Laboratoire d’Excellence” LabexMER (ANR-10-LABX-19) and co-funded by a grant from the French government under the program “Investissements d’Avenir”. FL, SP, CdV and ZM are funded by the European Union’s Horizon 2020 research and innovation programme “Atlantic Ecosystems Assessment, Forecasting and Sustainability” (AtlantECO) under grant agreement No 862923. ED and FL have been supported by the French National Research Agency under the grant number ANR-22-CE02-0025 (project COR-Resilience). FL and ZM are supported by the Belmont Forum project “World Wide Web of Plankton Image Curation” (WWWPIC, grant No ANR- 18-BELM-003). The support of Pr. Alan Fuchs, President of CNRS, was crucial for the success of the surface sampling undertaken during the *Tara* Pacific expedition. We thank A. Gavilli from TECA Inc. France, and E. Tanguy and D. Delhommeau from the Institut de la Mer, Villefranche-sur-Mer for the helpful collaboration in the conception of the High Speed Net and the Dolphin systems. The AT and CT data were analysed at the SNAPO-CO2 service facility at LOCEAN laboratory and supported by CNRS-INSU and OSU Ecce-Terra. We thank the EMBRC platform PIQv for image analysis and CCPv for samples storage and supported by EMBRC-France, whose French state funds through the ANR within the Investments of the Future program under reference ANR-10-INBS-02. The *Tara* Pacific expedition would not have been possible without the participation and commitment of over 200 scientists, sailors, artists and citizens (see <https://zenodo.org/record/3777760#.YfEsfXMLjB>). This publication is number 20 of the *Tara* Pacific Consortium.

Author contributions

Conceptualization and methodology: F.L., G.B., S.P., E.B., N.C., E.D., J.M.F., S.G.J., I.K., M.L.P., J.P.M.P. G.R., S.R., E.R., A.V., C.R.V., B.B., C.B., D.F., P.F., P.E.G., E.G., S.R., S.S., O.T., R.T., R.V.T., P.W., D.Z., D.A., S.P., M.B.S., C.d.V., E.B., G.G. Sample collection: F.L., G.B., S.P., S.A., E.B., P.C., O.D.S., E.D., A.E., J.M.F., J.F.G., B.C.H., Y.L., R.M., D.A.P., M.L.P., J.P., G.R., S.R., E.R., C.R.V., G.I., D.F., P.F., P.E.G., E.G., S.R., S.S., O.T. R.V.T., P.W., D.Z., D.A., S.P., C.d.V., E.B., G.G. Samples analysis and data analysis: F.L., G.B., S.A., E.B., N.C., M.C., P.C. C.D., E.D., A.E., J.F., J.M.F., J.F.G., B.C.H., L.J., S.G.J., R.L.K., Y.L., D.M., R.M., Z.M., N.M., D.A.P., M.L.P., M.P., J.R., G.R., S.R., E.R., C.R.V., B.B. Data production (models/satellites): F.L., G.B., A.B., O.D.S., C.R.V. Data Curation and validation: F.L., G.B., S.P., A.B., M.C., O.D.S., C.D., E.D., A.E., J.F., J.M.F., L.J., S.G.J., R.L.K., I.K., Y.L., D.M., R.M., Z.M., N.M., M.P., J.R., G.R., E.R., A.V., C.R.V. Funding Acquisition: F.L., N.C., P.C., E.D., J.F., J.M.F., J.F.G., S.G.J., I.K., D.M., N.M., M.L.P., M.P., G.R., E.R., A.V., C.R.V., B.B., C.B., D.F., P.F., P.E.G., E.G., S.R., S.S., O.T. R.T., R.V.T., P.W., D.Z., D.A., S.P., C.d.V., E.B., G.G.. Project Administration and supervision: F.L., S.P., S.A., E.B., J.M.F., E.R., C.R.V., C.M., B.B., C.B., D.F., P.F., P.E.G., E.G., S.R., S.S., O.T., R.T., R.V.T., P.W., D.Z., D.A., S.P., M.B.S., C.d.V., E.B., G.G. Visualization: F.L., G.B., Z.M. Writing – Original Draft Preparation: F.L., G.B., S.P., S.A., A.B., E.B., N.C., M.C., O.D.S., E.D., J.F., J.M.F., S.G.J., R.L.K., R.M., Z.M., C.R.V. All authors have read and reviewed the manuscript.

Competing interests


The authors declare no competing interests

Additional information

















Correspondence and requests for materials should be addressed to F.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

Fabien Lombard ^{1,2,3}✉, Guillaume Bourdin ^{1,4}, Stéphane Pesant⁵, Sylvain Agostini⁶, Alberto Baudena ¹, Emilie Boissin⁷, Nicolas Cassar ^{8,9}, Megan Clampitt^{10,11,12}, Pascal Conan^{13,14}, Ophélie Da Silva¹, Céline Dimier¹, Eric Douville¹⁵, Amanda Elineau¹, Jonathan Fin¹⁶, J. Michel Flores¹⁷, Jean-François Ghiglione ¹³, Benjamin C. C. Hume ¹⁸, Laetitia Jalabert¹, Seth G. John ¹⁹, Rachel L. Kelly¹⁹, Ilan Koren ¹⁷, Yajuan Lin^{8,9,20}, Dominique Marie²¹, Ryan McMinds ^{10,22,23}, Zoé Mériguet¹, Nicolas Metzler ¹⁶, David A. Paz-García ²⁴, Maria Luiza Pedrotti¹, Julie Poulain ²⁵, Mireille Pujo-Pay ¹³, Joséphine Ras¹, Gilles Reverdin¹⁶, Sarah Romac ^{2,21}, Alice Rouan ^{10,11,12}, Eric Röttinger ^{10,11,12},

Assaf Vardi²⁶, Christian R. Voolstra¹⁸, Clémentine Moulin²⁷, Guillaume Iwankow⁷, Bernard Banaigs⁷, Chris Bowler^{2,28}, Colombar de Vargas^{2,21}, Didier Forcioli^{10,11,12}, Paola Furla^{10,11,12}, Pierre E. Galand^{2,29}, Eric Gilson^{10,11,12,30}, Stéphanie Reynaud^{12,31}, Shinichi Sunagawa³², Matthew B. Sullivan³³, Olivier P. Thomas³⁴, Romain Troublé²⁷, Rebecca Vega Thurber²³, Patrick Wincker²⁵, Didier Zoccola^{12,31}, Denis Allemand^{12,31}, Serge Planes⁷, Emmanuel Boss⁴ & Gaby Gorsky^{1,2}

¹Sorbonne Université, Laboratoire d'Océanographie de Villefranche, UMR 7093, CNRS, Institut de la Mer de Villefranche, 06230, Villefranche sur mer, France. ²Research Federation for the study of Global Ocean Systems Ecology and Evolution, FR2022/Tara GOSEE, 75000, Paris, France. ³Institut Universitaire de France, 75231, Paris, France. ⁴School of Marine Sciences, University of Maine, Orono, Maine, 04469, USA. ⁵European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK. ⁶Shimoda Marine Research Center, University of Tsukuba, 5-10-1, Shimoda, Shizuoka, Japan. ⁷PSL Research University: EPHE-UPVD-CNRS, USR 3278 CRILOBE, Laboratoire d'Excellence CORAIL, Université de Perpignan, 52 Avenue Paul Alduy, 66860, Perpignan, Cedex, France. ⁸Nicholas School of the Environment, Duke University, Durham, NC, USA. ⁹Laboratoire des Sciences de l'Environnement Marin, UMR 6539 UBO/CNRS/IRD/IFREMER, Institut Universitaire Européen de la Mer, Brest, France. ¹⁰Université Côte d'Azur, CNRS, INSERM, Institute for Research on Cancer and Aging, Nice (IRCAN), Nice, France. ¹¹Université Côte d'Azur, Institut Fédératif de Recherche - Ressources Marines (IFR MARRES), Nice, France. ¹²LIA ROPSE, Laboratoire International Associé Université Côte d'Azur - Centre Scientifique de Monaco, Nice, Monaco. ¹³Sorbonne Université, CNRS, Laboratoire d'Océanographie Microbienne, LOMIC, 66650, Banyuls Sur Mer, France. ¹⁴Sorbonne Université, CNRS, OSU STAMAR - UAR2017, 75252 Paris, France. ¹⁵Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France. ¹⁶Laboratoire LOCEAN/IPSL, Sorbonne Université-CNRS-IRD-MNHN, Paris, 75005, France. ¹⁷Weizmann Institute of Science, Department of Earth and Planetary Sciences, Rehovot, Israel. ¹⁸Department of Biology, University of Konstanz, Konstanz, Germany. ¹⁹Department of Earth Science, University of Southern California, Los Angeles, CA, USA. ²⁰Environmental Research Center, Duke Kunshan University, Kunshan, China. ²¹Sorbonne Université, CNRS, Station Biologique de Roscoff, UMR 7144, AD2M, Roscoff, France. ²²Université Côte d'Azur, Maison de la Modélisation, de la Simulation et des Interactions (MSI), Nice, France. ²³Department of Microbiology, Oregon State University, Corvallis, OR, USA. ²⁴Centro de Investigaciones Biológicas del Noroeste (CIBNOR), La Paz, Baja California Sur, 23096, México. ²⁵Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, France. ²⁶Weizmann Institute of Science, Department of Plant and Environmental Science, Rehovot, Israel. ²⁷Tara Ocean Foundation, Paris, France. ²⁸Institut de Biologie de l'Ecole Normale Supérieure, Ecole Normale Supérieure, CNRS, INSERM, Université PSL, Paris, France. ²⁹Sorbonne Université, CNRS, Laboratoire d'Ecogéochimie des Environnements Benthiques, UMR 8222, LECOB, Banyuls-sur-Mer, France. ³⁰Department of Medical Genetics, CHU, Nice, France. ³¹Centre Scientifique de Monaco, 8 Quai Antoine 1er, MC-98000, Antoine, Monaco. ³²Department of Biology, Institute of Microbiology and Swiss Institute of Bioinformatics, ETH Zürich, Zurich, Switzerland. ³³Department of Microbiology and Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, OH, USA. ³⁴School of Biological and Chemical Sciences, Ryan Institute, University of Galway, University Road, Galway, Ireland. ³⁵e-mail: fabien.lombard@imev-mer.fr

3 Article 2: Utilizing Artificial Intelligence Tools to Automatically Annotate Visual Cues Related to Reef-building Coral Health (in prep)

3.1 Chapter 3 Foreword

Here, I present my first author paper (in prep) which describes the novel machines created to automatically annotate visual health related cues of individual coral colony photographs based on the photographic training data set from the Tara Pacific Expedition (detailed in Chapter 2).

Presently, the assessment of coral health remains elusive with ample focus on diseased and bleached states rather than a multifactorial approach that includes the plethora of internal and external factors and interactions that affect the survival of the coral animal. Here, we present AI monitoring algorithms that could help to decipher how visually assessed physiological signs play a role in overall coral health.

The goals of this research study were threefold. (1.) Validation - the creation of these models validated the use of AI tools to analyze coral colony photographs. Although, the Tara Pacific photographs were taken without an AI purpose in mind, functioning machines were still able to be created which allowed for the photographic analysis of individual coral colonies and helped us gain insight into important methodological improvements and considerations. (2.) Fieldwork Based Studies – the machines presented here served as the foundation for all the works presented in this Thesis. The creation of these machines was the first step in this research project with their functioning and assessment essential for the fieldwork-based research studies discussed in more detail here and in Chapters 4 and 5. There are also numerous future research possibilities using the AI machines created here which could prove to be valuable tools for monitoring coral reefs and helping conservation strategies. (3.) Tara Pacific Expedition - Finally, although not valorized within the timing of this PhD research, one of the values of the Tara Pacific dataset is the plethora of additional datasets and information linked to sampled colonies and their environment. Thus, there are many future research possibilities to explore in combining and investigating the photographic dataset with the other Tara Pacific datasets (explained in chapter 2). Overall, this chapter presents the foundational research study of this PhD work.

3.2 Utilizing artificial intelligence tools to automatically annotate visual cues related to reef-building coral health.

Authors: Megan Clampitt^{1-3,25*}, Marco Milanese^{25*}, Florentin Remot¹⁻³, Guillaume Bourdin⁴, Quentin Carradec^{5,6}, Benjamin C Hume⁷, Julie Poulain^{5,6}, Julie Lê-Hoang^{5,6}, Eric Armstrong^{5,6}, Sylvain Agostini⁸, Guillem Salazar⁹, Hans-Joachim Ruscheweyh⁹, Jean-Marc Aury⁵, David A. Paz-García¹⁰, Ryan McMinds^{1,11}, Didier Zoccola^{12,3}, Alexis Pey¹⁻³, Clémentine Moulin¹³, Emilie Boissin¹⁴, Guillaume Iwankow¹⁴, Sarah Romac¹⁴, Colomban de Vargas^{15,6}, Bernard Banaigs¹⁴, Emmanuel Boss⁴, Chris Bowler¹⁶, Eric Douville¹⁷, Michel Flores¹⁸, Stéphanie Reynaud^{12,3}, Olivier P. Thomas¹⁹, Romain Troublé¹³, Rebecca Vega Thurber²⁰, Serge Planes¹⁴, Denis Allemand^{12,3}, Stephane Pesant²¹, Pierre E. Galand²², Patrick Wincker^{5,6}, Shinichi Sunagawa⁹, Paola Furla¹⁻³, Christian R Voolstra⁷, Didier Forcioli¹⁻³, Fabien Lombard²³, Eric Röttinger^{1-3@}, Eric Gilson^{1-3,24@}

¹ Université Côte d'Azur-CNRS- Inserm - Institute for Research on Cancer and Aging, Nice (IRCAN), France.

² Université Côte d'Azur, Institut Fédératif de Recherche - Ressources Marines (IFR MARRES)

³ Laboratoire International Associé Université Côte d'Azur - Centre Scientifique de Monaco (LIA ROPSE), France, Monaco.

⁴ School of Marine Sciences, University of Maine, Orono, Maine, United States of America

⁵ Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057, Evry, France.

⁶ Research Federation for the Study of Global Ocean Systems Ecology and Evolution, R2022/Tara Oceans GO-SEE, 75016, Paris, France.

⁷ Department of Biology, University of Konstanz, Konstanz, Germany

⁸ Shimoda Marine Research Center, University of Tsukuba, Shimoda, Japan.

⁹ Department of Biology, Institute of Microbiology and Swiss Institute of Bioinformatics, ETH Zurich, 8092 Zurich, Switzerland

¹⁰ CONAyT-Centro de Investigaciones Biológicas del Noroeste (CIBNOR), Laboratorio de Necton y Ecología de Arrecifes, La Paz, Baja California Sur 23096, México.

¹¹ University of South Florida Center for Global Health and Infectious Diseases Research, Tampa, FL, United States

¹² Centre Scientifique de Monaco, Principality of Monaco

¹³ Tara Ocean Foundation, 8 rue de Prague, 75012 Paris, France.

¹⁴ Laboratoire d'Excellence "CORAIL," PSL Research University: EPHE-UPVD-CNRS, USR 3278 CRIOBE, Université de Perpignan, Perpignan Cedex, France.

¹⁵ Sorbonne Université, CNRS, Station Biologique de Roscoff, AD2M, UMR 7144, ECOMAP, Roscoff, France.

¹⁶ Institut de Biologie de l'Ecole Normale Supérieure (IBENS), Ecole normale supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France

¹⁷ Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France

¹⁸ Weizmann Institute of Science, Department of Earth, and Planetary Sciences, 76100 Rehovot, Israel

¹⁹ School of Biological and Chemical Sciences, Ryan Institute, University of Galway, University Road H91TK33 Galway, Ireland

²⁰ Oregon State University, Department of Microbiology, 220 Nash Hall, 97331 Corvallis OR USA

²¹ PANGEA, Data Publisher for Earth and Environment Science, Bremen, Germany,

²² Sorbonne Université, CNRS, Laboratoire d'Ecogéochimie des Environnements Benthiques, Observatoire Océanologique de Banyuls, Banyuls-sur-Mer, France

²³ Sorbonne Université, Institut de la Mer de Villefranche sur mer, Laboratoire d'Océanographie de Villefranche, Villefranche-sur-Mer, France.

²⁴ Department of Medical Genetics, CHU, Nice, France.

²⁵ Université Côte d'Azur, Center of Modeling, Simulation, & Interaction, France

*These authors contributed equally to this work

@Corresponding authors

Statements and Declarations:

The authors have no competing interests to declare that are relevant to the content of this article.

Acknowledgements:

This work was supported by ANR-15-IDEX-01 via the IDEX UCA^{JEDI} (MSI), the Ecole Doctorale Sciences de la Vie et de la Santé de Université Côte d'Azur, the LIA-ROPSE, The Université Côte d'Azur, la Région SUD, CNRS INSB. The authors are grateful to Erin Bowman, Clara Fricano and Florentin Remot who annotated photos to calculate the interrater agreement. Research was conducted under permit

Ce travail a bénéficié d'une aide du gouvernement française, gérée par l'Agence Nationale de la Recherche au titre du projet investissement d'Avenir UCA^{JEDI} portant la référence n° ANR-15-IDEX-01

Abstract:

Coral reefs are deteriorating at a startling rate and the development of fast and efficient monitoring schemas that attempt to evaluate coral health is essential. Here, we developed artificial intelligence (AI) tools to automatically annotate visual cues (Boring organisms, sediment contact, macroalgal contact, contact with CCA or Turf algae, pigmentation response, predation, tissue appearance, and bleaching) potentially related to the coral state health detected from coral colony photographs generated during the Tara Pacific Expedition. We investigated the use of these tools for photographic monitoring of reef corals in a follow-up field study. Convolutional neural networks were used to train machines on the manual annotations of coral colony photographs to classify photos automatically. Using transfer learning and data augmentation, we were able to achieve F1 Scores averaging 83% percent across a set of attributes. When testing our machines on novel photographs, decent results were achieved for bleaching and predation with results subject to pretreatment, highlighting the need for a strict photographic methodology and increased training samples. Our study shows the tremendous potential of this technology and is the foundational work for the development of future AI machines investigating coral colony health (as defined by susceptibility to mortality).

Key Words: (4-6)

Coral, Artificial Intelligence (AI), coral health state, Convolutional Neural Network (CNN)

Introduction:

Corals are essential to marine life as they provide shelter, and act as nursery and feeding grounds for thousands of marine species (Manderson et al., 2017). Coral reefs also play a critical role in human survival through the provision of numerous ecological goods and services such as coastal protection, medicinal compounds, food provisions, livelihood from tourism and recreation as well as providing aesthetic, spiritual and cultural value (Moberg & Folke, 1999). Despite their importance, approximately 14% of the world's coral reefs have been lost since 2005 (GCRM, 2020) with some predictions stating that all reefs could be lost by the end of the century (Manderson et al., 2017). Anthropogenic climate change paired with overfishing and pollution are some of the main threats destroying these vital ecosystems (Hughes et al., 2017) with predictions stating that even if global warming remains below 2°C, all coral reefs will continue to regress from their current state resulting in destructive consequences for the services they provide (Bindhoff et al., 2019). Coral reef monitoring is essential not only to assess the state of these ecosystems but also to detect the impacts of stressors, assess recovery, and better inform management decisions (Flower et al., 2017).

Coral health, much like human health, is complex and multifactorial making it difficult if not impossible to describe based on a single variable. Currently, coral health measures remain incomplete with a focus on the negative definition of health i.e., the absence of pathology (Breslow, 1972) (López-Otin & Kroemer, 2020). Practically, this means diseased or bleached corals are often defined as unhealthy whereas corals showing no signs of bleaching or disease are viewed as healthy. Disease and/or bleaching are commonly used as the singular determinant of coral colony health aligning with this definition of health. However, health has been more globally defined by the World Health Organization (WHO) as “physical, mental and social well-being, not merely the absence of disease” (Constitution of the World Health Organization, 1948). Thus, a similar global approach that considers other factors aside from the presence/absence of pathology is needed to define coral health.

Currently, as technology is fast advancing, high quality and high-resolution photo equipment is more commonplace making digital imagery an inexpensive, rapid, and widely utilized resource for monitoring marine benthos (Page et al., 2016). However, the manual annotation of collected images is labor-intensive and requires an expert resulting in over 98% of images remaining unannotated (Mahmoud et al., 2016). Currently, the literature on the use of Artificial Intelligence in coral reef monitoring focuses on identification and benthic composition (Mahmoud et al., 2016) (Bejibom et al., 2012) (Stokes & Deane, 2009). We hypothesize that a combination of health-related coral colony attributes, which can easily be extracted from colony photographs, can serve to assess the health status of a given colony, echoing the frailty concept used in geoscience (Fried et al., 2021). Therefore, we aim to develop simple to use tools to assess coral states of decreased reserves which would result in increased vulnerability to diseases and stressors even in the absence of visible bleaching or disease at the time of colony inspection. As a first step toward this goal, we present the design of artificial intelligence (AI) deep learning models to automatically assess the nine following health-related attributes.

Bleaching: Coral bleaching is a stress response that can result from a multitude of factors but is most often attributed to increased water temperature resulting from anthropogenic induced climate change (Hoegh-Guldberg, 1999). Bleaching can be defined as the loss of color in response to the elimination (total or partial) of the population of photosymbiotic dinoflagellates (zooxanthellae) or degradation of the algae's photosynthetic pigment (Douglas, 2003). The zooxanthellae enhance calcification, mediate elemental nutrient flux, and provide carbon to the corals (Glynn, 1996). Bleached corals can lose between 60-90% of their zooxanthellae which

greatly affects the fitness of the coral (Glynn, 1996). Bleaching also has been shown to lead to reduced fecundity and growth, inhibited regeneration from injury, reduced vitality, tissue loss and death (Baird & Marshall, 2002; Ward, Harrison, Hoegh-Guldberg, 2000) (Glynn, 1996). Chow and colleagues have shown that digital photography can be successfully used to monitor bleaching (Chow et al., 2016).

Algal Contact: Competition between corals and algae is a fundamental characteristic that helps determine the status of coral reefs with both sessile organisms competing for the limited resources of space and light (McCook, Jompa and Diaz-Pulido, 2001). The result of this competition can have substantial consequences on long term survival and growth of coral colonies (Lirman, 2000) and varies dependent on the functional group (Barrot et al., 2009, Cetz-Navarro et al., 2013). Interactions between corals and macroalgae and/or turf algae are typically seen as having negative effects whereas the effects of crustose coralline algae (CCA) are typically seen as positive (Cetz-Navarro et al., 2013). For example, Barrot et al., (2009) found that contact with macroalgae and turf algae acted as a source of stress for the corals, causing bleaching, paling or tissue damage whereas this was not the case with CCA. Additionally, interactions between corals and turf or macroalgae are hypoxic whereas CCA is not (Barrott et Rohwer, 2012). Therefore, we distinguished three separate types of algal contact.

Macroalgae: Increases in macroalgae are often associated with decreased reef health with coral dominated reefs shifting to macroalgal dominant reefs as a result of anthropogenic influences such as overfishing and eutrophication (McCook, 1999). Macroalgal contact has been shown to have negative influences on coral colonies; contact with macroalgae not only increases tissue mortality but has also been shown to reduce coral growth and fecundity (Lirman, 2000). Additionally, contact with certain types of macroalgae could lead to disease transmission and macroalgal increases over the last 30 years could be an explanation for the recent surge in coral disease (Nugues et al., 2004).

Turf Algae: Similarly, contact with turf algae has been shown to lead to tissue disruption, pigmentation decreases, bleaching, decreased fecundity, photosynthetic performance, and growth rates, as well as tissue mortality (Wild, Jantzen, and Kremb, 2014; Cetz-Navarro et al. 2013). Coral-turf contact has been shown to alter the coral microbiome with increases in bacteria associated with disease and bleaching at the point of contact (Pratte et al., 2017). The removal of turf-algae in contact with corals, can reduce stress (Cetz-Navarro et al., 2013).

Crustose coralline algae (CCA): In contrast, CCA are known to be beneficial to reef health via substrate stabilization and reef framework creation (Littler & Littler, 2013). They benefit corals throughout their life stages from the promotion of coral recruitment to the inhibition of macroalgal growth and formation on adult colonies (Vermeij, Dailer & Smith, 2011) (Barrott and Rohwer, 2012). CCA helps to maintain coral dominance on reefs and the loss of CCA would lead to alternative stable states such as macroalgal dominant reefs (Weiss and Marrtindale, 2017; Vermeij, Dailer, and Smith, 2011). Corals surrounded by CCA showed less stress (increased tissue thickness, chlorophyll and zooxanthellae density and a lower mitotic index) than colonies surrounded by turf algae (Cetz-Navarro et al., 2013) and contact with CCA does not seem to negatively affect or disrupt the colony (Barrot et al., 2009).

Predation: Corallivory, or direct consumption of live coral, affects corals directly via tissue damage, tissue loss, reduction in growth rates and decreased reproductive potential as well as indirectly through facilitating colonization by algae and other competitors, decreasing immune responses, nutrient sequestering, symbiont dispersal, and disease transmission (Rotjan & Lewis 2008) (Grupstra et al., 2021). Predation has been shown to directly affect the coral microbiome making colonies more vulnerable to stressors (Ezzat et al., 2020). Furthermore, it has also been shown that corallivory affects the abundance and distribution of coral species along the reef (Cox, 1986). Thus, coral predation is a key factor in coral assessment whose beneficial or detrimental effects are still elusive.

Sediment Contact: Sedimentation, either from natural processes or resulting from human activities such as dredging, can be detrimental to coral colonies; particles can become suspended in the water column leading to reduced light levels or accumulate on the colony surface and injure and/or smother the polyps (Rogers, 1990; Rushmore, Ross & Fogarty, 2020; Anthony and Larcombe, 2000). Thus, leading to reduced recruitment, growth and photosynthetic efficiency as well as inducing morphological changes and increasing respiration and colony mortality (Rogers, 1990; Rushmore, Ross & Fogarty, 2021). It has already been suggested a few decades ago that observation and photography over time could be a useful method for monitoring sediment smothering (Sammarco & Risk, 1990).

Pigmentation: Pigmentation response is associated with dead, dying, or regenerating tissue (Kubomura, Wee and Reimer, 2020) and has been shown to be an immune response activated by the coral in response to stress (Palmer, Mydlarz and Willis, 2008). Fluorescent proteins, responsible for the vivid colors seen in corals, are responsible for this phenomenon and have been found to act as antioxidants further solidifying pigmentation response as playing a role in coral stress and immune response (Palmer, Chintan and Mydlarz, 2009).

Boring organisms aka coral associates: Boring organisms here defined as invertebrates that live on or within the coral skeleton and are visible on the colony surface, can impact colony physiology. For example, bivalves of the genus *Lithophaga* have been shown to weaken the skeletal strength of corals, specifically adversely affecting the genus *Porites* which has a very porous, low-density skeleton (Risk & Scott, 1988). It has also been found that polychaete worms, another boring organism, can smother corals and lead to mortality (Smith & Harriot, 1998). It has even been suggested that coral bivalves could be used as an indicator of coral reef status with pristine sites showing fewer infested colonies and less infestations overall compared to highly impacted sites (Scaps & Denis, 2008). Thus, boring organisms might be an overlooked indicator that could be important to not only coral physiology but also as an indicator of the overall status of the reef.

Tissue Appearance: In many monitoring schemas, coral tissue appearance is visually assessed to record other conditions affecting colony health such as disease, tissue loss, physical damage, growth anomalies, and/or bacterial or fungal infections (Swanson et al., 2018; Raymundo et al., 2008). Diseases are typically identified macroscopically by tissue color, exposure of skeleton and extent of tissue loss (Ainsworth et al., 2007). Outbreaks of disease can be accredited as one of the major contributing factors to declining coral cover in the Caribbean, having serious consequences on the entire coral reef ecosystem (Rogers & Miller, 2013).

Here, we examine the combination of photographic monitoring with artificial intelligence to optimize temporal and monetary expenditures during both data collection and analysis. Our results confirm the use of AI machines to assess nine coral health-related attributes in a timely

and cost-effective manner. We believe this approach will be useful for future coral ecophysiology studies and reef monitoring, as well as in identifying visually apparent health-related hallmarks.

Materials & Methods:

Photo Collection and Annotation:

The photos utilized for this study were obtained from two sources:

1. **Tara Pacific Expedition (TPE):** The Tara Pacific Expedition (2016-2018) was designed to study the complexity of the coral holobiont across the Pacific Ocean targeting three coral genera: *Porites sp*, *Pocillopora sp* and *Millepora sp* (Planes et al., 2019). 11,460 photos were obtained during this expedition and used as the basis for our AI model. The methods describing the collection and annotation of these photos are detailed in Lombard et al. (2023). A MATLAB script was developed specifically for this project and can be found as a supplementary file (Code S1). Prior to commencing the annotation focused on the visual attributes of each coral colony, the annotator was prompted to determine characteristics about the photograph including: if the photo is useful (Yes/No), if there is a target colony in the photo (Yes/No), identifying the type of photo (standard, wide-angle or close-up), and identifying if a quadrat (or some object for measurement) was present. Major annotation prompts (Tissue Appearance, CCA, Pigmentation, Predation, Turf Overgrowth, Sedimentation, Boring Organisms, and Macroalgal Contact) were then annotated as binary variables with absence being 0 and presence being 1. However, for the attributes Boring Organisms and Macroalgal Contact, the annotator was first prompted to determine if the attribute was present and then identify, if possible, the boring organisms present or the macroalgal in contact with the colony. Thus, although first annotated as a binary category, more detailed information is available for these attributes (Figure 1A). Boring organisms include *Tridacna*, *Spirobranchus*, Sponge, Other Polychaetes, and Bivalves. Macroalgal Contact includes *Turbinaria*, *Halimeda*, *Dictyota*, *Lobophora*, and *Galaxaura*. For both attributes, range values were calculated as well *i.e.*, the number of distinct boring organism annotations per picture or number of distinct algal genera annotations per picture resulting in a value from 0-5 for each picture. Bleaching was categorical with the annotator prompted to choose from four categories: bleached, pale, normal, or dark.
2. **Moorea Longitudinal Study:** A longitudinal study was set up in Moorea, French Polynesia in November 2021 replicating the methodology used during the Tara Pacific expedition. Sampling will be conducted on a yearly basis with the second time point occurring in November 2022 and the third timepoint scheduled for November 2023...*etc.* At the time of this article, data from 2021 and 2022 was utilized, focusing on the same three coral genera (*Porites*, *Pocillopora*, and *Millepora*) and visiting the same 3 sites (Haapiti, Papetoai, and Temae) as the Tara Pacific expedition (Figure 6). Samples were collected from Papetoai and Temae in 2021, with Haapiti added in 2022. Ten corals from each of the three genera were tagged at each site, totaling 30 colonies possible per site and 90 colonies overall with each colony photographed three times (Figure S1C). Additionally, an *in-situ* survey was conducted at the time of photographing which gathered information about living tissue percentage, bleaching, predation, and disease. Due to weather constraints, *Millepora sp.* was not sampled or photographed at Temae in 2022 resulting in fewer colonies and certain tags were not found, meaning not all colonies were sampled. The photographs collected for this study

were used for the validation tests of the AI machines built from the Tara Pacific Expedition photographs.

Inter Annotator Agreement:

An inter-annotator agreement score was calculated by having a coral specialist separately annotate a sub-set of photos from the TPE following the annotation guidelines used by the original annotator. This included a two-step process with a discussion between specialists after the first round of annotations. The second final round of annotations were used to calculate the inter annotator agreement score using Cohens Kappa Coefficient (Cohen, 1960).

Photo dataset cleaning:

The full Tara Pacific photo dataset (N=11,470) was not created with a machine learning approach in mind. Thus, photographs were carefully selected to achieve our objectives by removing duplicates, photos of non-coral samples, non-target coral colonies, photos where the target colony was unclear, and photos of low quality or containing too much noise (divers, tags, tools...*etc.*) (Figure S1A). The manual annotations were used to discern photos of low quality or non-target colonies. A database was also created for this project which was used to group photos by colony and manually verify the photo quality to remove any further photos in the categories listed above. The resulting “cleaned” dataset contained 5555 photos.

Symbiont Biomass & Coloration Levels: Symbiont biomass (mg^2) data was taken from Porro & Zamoum et al., in press following the sampling procedures described in Lombard et al., 2023. Analyses were performed in the program R version 4.1.2 (R Core Team, 2021). Symbiont biomass was log-transformed to improve normality of the residuals. We ran linear models to assess whether algal levels were related to coloration levels (bleached, light, normal and dark) obtained through photographic annotation. We also calculated the median and 95% CIs for the difference in RTL for individuals in different categories by calculating pairwise contrasts using emmeans version 1.8.6 (Lenth *et al.*, 2023). Data visualization was performed using ggplot2 version 3.4.2 (Wickham *et al.*, 2023). Symbiont biomass and photographic annotations were only available for 68 *Porites sp.* and 118 *Pocillopora sp.* and analyses were conducted individually by genus.

Photo Treatment Selection:

To determine image pretreatment, the entire photo dataset was validated using a simple architecture containing four convolutional layers. Four different image treatments were tested: (1) resize, (2) center crop, (3) resize & center crop (4) random crop. For resize, images were resized to 224 x 224, for center crop a 224 x 224 tile was cropped from the center of the image, for resize and center crop the image was resized to 1024 x 1024 and then a center crop of 224 x 224 was taken. For random crop, the image was randomly cropped to 224 x 224. Resize was selected as the preprocessing method.

Network Architecture & Initial Machines:

We trained 9 deep neural networks, one for each attribute as determined by the separate prompts used in the code (Predation, Boring Organisms, Macroalgae Contact, Sediment, Turf Overgrowth, Bleached, Pigmentation and Tissue Appearance) with the addition of CCA which was originally one of the subcategories for Macroalgae Contact (SI Code 1). CCA was used as its own category with its own machine due to the different functions and effects of CCA vs. Macroalgae in relation to hard corals. Managing each attribute independently allowed for the construction of independent pipelines which in turn allowed us to match different features extracted from the picture to the single qualitative attribute. All networks share a common first

block constituted by a pre-trained ResNet-50 Network, in order to cope with the low number of pictures available. The ResNet-50 architecture was not trained, and weights were only updated for the fully connected layer where the Tara photo dataset (N=5555) was incorporated. This layer consisted of fully connected layers, followed by Relu (Activation function) and a SoftMax for classification (Figure 3A). An 80/10/10 split was used for the training, validation, and testing sets, respectively. Different tests were conducted to evaluate classification performances on both binary labels (i.e., presence or absence of a given attribute) and multi-class labels (bleached, number of boring organisms/ algae). For the latter case, a binary version of the network was tested as follows: A) Bleached: we grouped the 4 possible values into two classes: normal/dark (0) and bleached/pale (1) B) Boring Organisms/Macroalgal contact: we signaled the presence (1) or absence (0) of any number of organisms in the picture. The source code is available at <https://github.com/UCA-MSI/tara-classification>.

Photo Augmentation:

The size of the dataset was increased by randomly selecting and transforming class 1 (presence) images of the desired variable (Figure 4B). The number of photos to be augmented was determined individually for each variable to achieve a ratio of 60% between class 0 (absence) and class 1 (presence). This ratio corresponds to the ratio of the least skewed attribute (Turf Overgrowth) from the initial machines. Image augmentation occurred through random manipulation via rotation, cropping, and changing the contrast of original images to create more images. Since the attribute turf overgrowth already had a ratio of 60%, the number of class 1 photos for the category Turf Overgrowth was augmented to equal the average number of total photos for all other augmented variables. Script is available at <https://github.com/UCA-MSI/tara-augment>.

Additional Analyses of AI Machines:

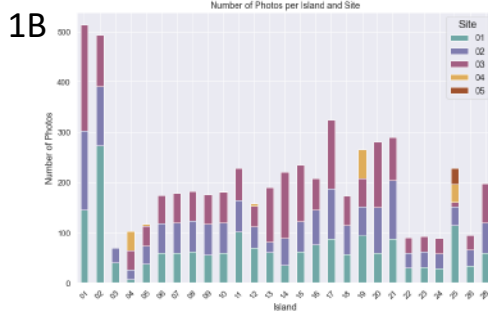
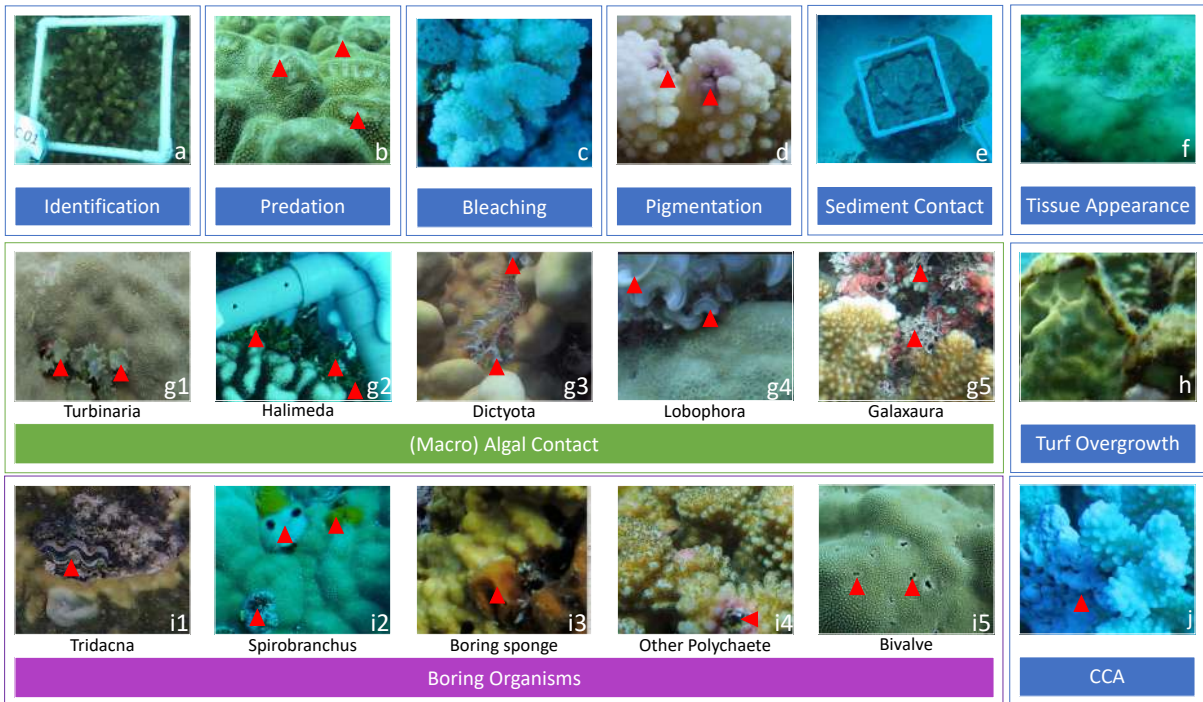
1. Cosine similarity values (CSV) - CSV of the average weights of the final three fully connected layers were calculated for each attribute against all the other attributes. Script is available at <https://github.com/UCA-MSI/tara-classification>.
2. Moorea Longitudinal Testing Set – the photographs and in water annotations collected from Moorea were run through our finalized machines trained on the Tara Pacific data and were used as a validation set with machine predictions compared to the annotations taken *in situ* for (bleaching, predation, and disease). The photos were treated in 3 ways: 1. Resized (224,224), 2. Center Cropped (224,224), 3. Resized (1024) and center cropped (224,224).

Results:

Manual annotation of visually cues from coral colony images is robust and reflects biological data

During the Tara Pacific expedition, all sampled colonies were meant to be photographed twice, resulting in a theoretical number of 6,120 photographs (102 sites x 30 colonies x 2 photographs per colony). Five islands (27, 29, 30, 31, 32) had no photographic data due to technical issues (camera flooding, memory card issues, battery problems...*etc.*) updating the theoretical number to 5,220 total photographs. To perform the manual annotation, a Matlab script was created for this purpose with a pipeline for identifying the physiological hallmarks present in the photographs (Fig1A). The entire image dataset that went through the manual annotation process contained 11,470 images. The first few steps of the manual annotation were used to confirm if a photograph was good quality and contained a target colony. Thus, using the manual

1A



1C

Island	Bivalve	Spirobranchus	Tridacna	Other Polychaete	Boring Sponge	Halimeda	Turbinaria	Dictyota	Lobophora	Galaxaura	CCA	Turf Overgrowth	Tissue Appearance	Sediment Contact	Pigmentation	Predation	Bleached	Light	Normal	Dark
01	237	5	0	11	0	0	0	3	0	3	15	265	28	121	61	255	3	0	215	295
02	147	2	1	14	0	1	0	2	7	34	68	210	25	14	39	242	5	3	272	214
03	56	2	0	13	0	1	0	0	0	12	8	0	5	7	61	0	1	47	22	
04	5	1	0	2	0	4	0	1	0	0	1	20	4	29	5	5	1	17	79	6
05	0	1	0	1	0	0	0	0	0	0	57	13	9	0	0	5	24	40	53	0
06	14	3	0	21	0	12	0	0	0	4	64	12	25	4	16	14	58	96	6	
07	11	15	0	29	0	7	0	0	1	10	59	5	13	16	17	5	54	103	17	
08	12	2	0	36	0	9	0	0	0	61	67	5	5	29	12	0	27	145	11	
09	22	8	1	27	0	47	0	0	5	33	45	3	3	23	10	1	37	125	13	
10	14	11	0	22	1	2	0	0	0	18	63	55	1	7	17	20	3	149	31	
11	17	13	1	35	0	9	0	0	0	42	81	1	14	10	11	4	78	145	1	
12	5	0	0	8	0	23	0	0	2	101	32	2	12	12	10	28	20	102	8	
13	27	2	0	29	0	19	0	0	1	81	62	4	12	14	18	2	5	135	48	
14	15	1	0	18	0	42	2	1	1	40	90	15	20	27	17	29	26	162	4	
15	19	3	0	69	0	29	1	0	39	15	36	113	3	8	55	17	2	14	207	12
16	44	10	1	45	0	0	0	11	6	4	26	89	2	21	45	31	2	7	148	50
17	48	9	0	60	0	0	23	0	0	80	178	4	30	73	60	4	4	259	57	
18	39	3	0	31	0	3	2	0	1	7	25	81	1	12	16	29	0	6	130	37
19	44	21	0	46	0	2	0	2	1	53	119	10	52	5	23	2	27	217	20	
20	21	16	0	30	0	104	1	2	1	129	82	10	44	14	54	2	44	223	11	
21	27	25	0	39	0	1	0	3	1	22	125	0	55	22	43	2	40	232	15	
22	14	6	1	27	0	5	0	0	0	2	30	32	0	8	25	4	1	2	77	10
23	12	0	0	26	0	12	0	1	1	7	27	35	0	15	11	4	0	12	78	2
24	13	1	0	32	2	9	0	0	0	4	29	28	0	9	14	8	0	1	78	10
25	24	4	19	19	0	15	0	0	0	17	91	2	83	10	23	10	9	149	60	
26	10	4	4	23	0	2	0	0	1	15	40	0	11	3	11	1	4	85	5	
28	33	7	1	40	0	1	1	3	1	3	30	101	4	35	32	12	1	11	152	34
Total	930	175	29	753	3	359	30	29	61	107	1107	2185	150	663	589	1018	143	550	3863	999

Figure 22: (A) Annotation categories for the manual annotations of the Tara Pacific Photographs: Identification (Genus), Predation, Bleaching, Pigmentation, Sediment Contact, Tissue Appearance, MacroAlgal Contact (Turbinaria, Halimeda, Dictyota, Lobophora, Galaxaura), Turf Overgrowth, Boring Organisms (Tridacna, Spirobranchus, Boring Sponge, Other Polychaeta, Bivalve), and CCA. Red arrowhead indicates instance(s) of the indicated variable although multiple may be present in the photo (B) Graph depicting the number of coral colony photographs from the TP expedition distributed by island and site, after data cleanup (C) Figure showing the distribution of photographs that were annotated as “present” (1) for each of the annotation categories organized by island with the total number of photos per attribute shown. Islands for which no photographs are available: 27, 29, 30, 31, 32 were not included in Figure B or C.

annotations, we were able to identify and eliminate photos of fish samples, blurry photos, non-target colonies, photographs where the target colony was unclear, photographs containing too much noise, *etc.* (Figure S1A).

After photo cleaning, the actual number of photographs varied per island with islands 1 and 2 containing an elevated proportion of photographs (N~500) and islands 3, 4, 5, 22, 23, and 24 averaging less than 100 photographs per island (Figure 1B, Table S1). The theoretical value of photographs for an island with three sites was 180 photographs. Numbers above or below this value indicate that the average number of photos per colony (theoretical value = 2) varied depending on the island. For example, islands 1 and 2 showed a higher average number of photos per colony (N=7) compared to the rest of the islands (N~2) (Table S1). Furthermore, sometimes islands had over 30 colonies due to coral biodiversity sampling described in more detail in Lombard et al., 2023. Despite this variation, excess photographs per island were not eliminated from the analysis to maximize the size of the dataset. Photos were only eliminated based on photo quality and if no target colony was present. After cleanup, the resulting dataset contained 5555 images which were reduced from the original number and much closer to the theoretical number (Fig1B).

Thus, a final photo dataset containing 5555 images after removing all non-usable photographs was created with annotations identifying genus and absence or presence of 9 attributes of potential physiological hallmarks (predation, pigmentation, macroalgal contact, CCA, turf algae, bleaching, sediment contact, tissue appearance, and boring organisms) in individual colony photographs. The genus analysis resulted in the identification of 2207 *Porites sp.*, 2099 *Pocillopora sp.*, and 1249 *Millepora sp.* colonies (data not shown). Globally, 1631 colonies indicate the presence of boring organisms, 150 colonies appear “unhealthy or show recent tissue loss”, 552 colonies are in contact with macroalgae, 1107 in contact with CCA, 2185 show turf overgrowth, 589 show a pigmentation response, 1018 have predation marks, 663 are in contact with sediment, and 693 annotated as pale or bleached (Fig 1C, Fig 4A). Figure 1C shows the distribution of the number of photographs annotated as present (1) for each attribute by island. In summary, all attributes have higher instances of absence than presence with certain attributes (Tissue Appearance, Pigmentation, Macroalgal Contact, Sediment contact) showing presence in only approximately 10% of annotated photographs. Thus, the manual annotation of the 5555 colony photographs yielded an additional, albeit skewed, dataset that can be used in correlation with all other Tara Pacific datasets (Lombard et al, 2023).

A. Major Categories

Attribute	Value	Agreement Level
Genus	0.950	Very Good
Bleached	0.513	Moderate
Predation	0.451	Moderate
Tissue Appearance	0.332	Fair
Sediment	0.395	Fair
Pigmentation	0.546	Moderate
Boring Organisms	0.664	Good
Algal Contact	0.749	Good
CCA	0.590	Moderate
Turf Overgrowth	0.374	Fair

B. Macroalgal Contact

Attribute	Value	Agreement Level
Halimeda	0.852	Very Good
Turbinaria	0.790	Good
Dictyota	0.556	Moderate
Lobophora	0.652	Good
Galaxaura	0.462	Moderate

C. Boring Organisms

Attribute	Value	Agreement Level
Bivalve	0.522	Moderate
Spirobranchus	0.864	Very Good
Tridacna	0.656	Good
Other Polychaete	0.492	Moderate
Boring Sponge	0.000	None or Missing Data

Table 1: Inter-annotator score for each of the different manual annotation categories calculated using Kohens Cappa Coefficient. The major annotation categories are shown in red (A) and the subcategories for macroalgal contact (B) and boring organisms (C) are also shown. The resulting values can be interpreted as follows: values less than or equal to zero indicate no agreement (or missing data), 0.01–0.20 is considered poor, 0.21–0.40 is interpreted as fair, 0.41–0.60 as moderate, 0.61–0.80 as Good/substantial, and 0.81–.99 as very good with 1 indicating perfect agreement.

To validate the manual annotations and the annotation protocol, an inter-annotator score was calculated for each of the annotation categories. Certain attributes showed a higher level of agreement than others, suggesting that certain attributes are easier to annotate than others (Table 1), the average score for binary annotation categories was .56. This outcome corresponds to a moderate agreement on the Kohen Cappa scale (Cohen, 1960) between human annotators and validates the replicable annotation protocol and thus the results of the manual annotations.

2A

Colour	Estimate	Lower Limit	Upper Limit	Sample Size
Bleached	12.7	11.7	13.6	8
Light	12.7	12.2	13.2	31
Normal	12.3	12.0	12.7	53
Dark	13.9	13.4	14.5	26

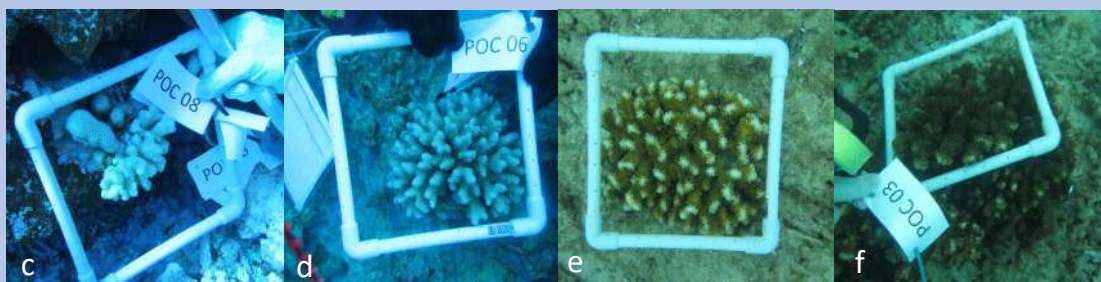
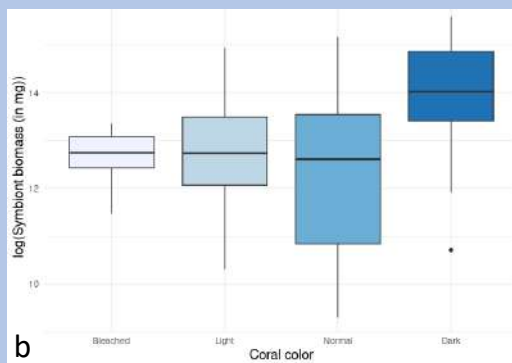
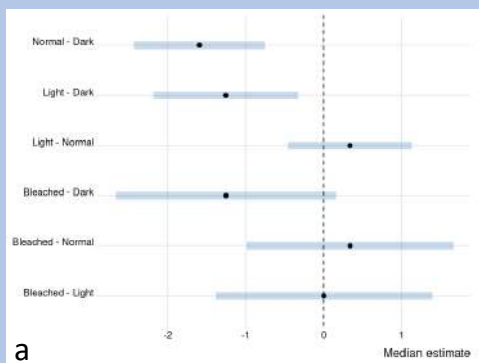
2B

Colour	Estimate	Lower Limit	Upper Limit	Sample Size
Bleached	N/A	N/A	N/A	N/A
Light	13.4	12.7	14.2	5
Normal	13.6	13.3	13.8	57
Dark	13.7	13.0	14.3	6

Table 2: Tables (A, B) representing the estimates of the linear model fitting along with their 95% confidence interval and the sample size for each coloration level (bleached, light, normal, dark). Table 2A is for Pocillopora species and Table 2B is for Porites species.

2A

Pocillopora sp.



2B

Porites sp.

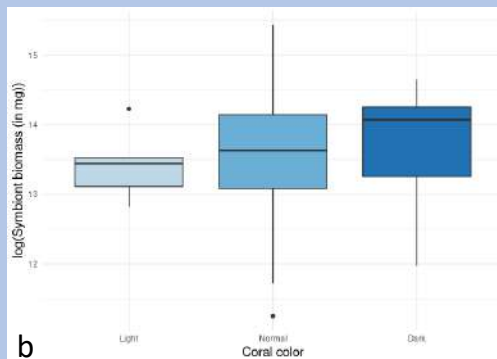
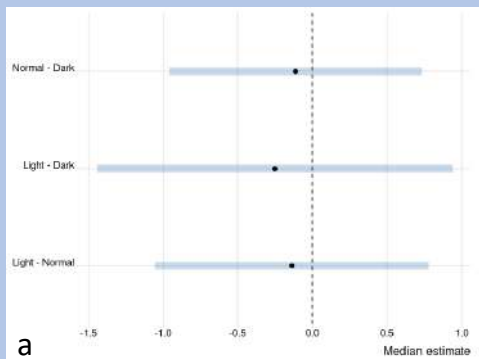


Figure 2: Data comparing the manual annotations from the Tara Pacific Photographs with the algal biomass (mg²) extracted from photographed samples. Panel A shows the data for *Pocillopora sp.* and panel B shows the data for *Porites sp.* Pairwise contrast plots (Aa, Ba) showing differences between symbiont biomass for each color category, Boxplots (Ab, Bb) summarizing symbiont biomass values for each color category and examples of annotated photos from each color category: bleached (Ac, Bc), light (Ad, Bd), normal (Ae, Be), and dark (Af, Bf).

To further validate the manual annotations and determine if biological significance could be given to attributes from the manual annotations, bleaching categories (bleached, light, normal, and dark) of individual colonies were compared to previously experimentally determined values of symbiont algal biomass (mg^2) of the same colonies (Porro & Zamoum et al, *in press*). As the latter data was only available for *Pocillopora sp* and *Porites sp*, *Millepora sp*, were excluded from this analysis. Coloration results for *Pocillopora sp* were linked to symbiont biomass, specifically with dark colonies having more symbionts as compared to bleached, light, and normal colonies (Table 2A). Figure 2Ab shows boxplots for each coloration category with dark showing higher algal biomass than the other categories as confirmed by the pairwise comparisons (Table2Aa). For *Porites sp*, none of the sampled colonies had associated photographs that were annotated as bleached leaving only light, normal, and dark as coloration categories with no differences found between coloration and symbiont biomass (Figure 2B) (Table 2B). It is important to note that sample sizes were very small with *Porites sp*. having fewer overall samples than *Pocillopora* and completely missing samples showing bleaching (Table 2A, Table 2B). Larger sample sizes would be needed to conduct a thorough analysis. Even so, these results, specifically those for *Pocillopora sp*. suggest biological meaning behind the annotation categories. Overall, we have manually annotated 5555 coral colony images using a robust and annotator independent annotation protocol yielding a dataset describing attributes that may be linked to the physiological status of the colony which in the case of bleaching shows a slight correlation for *Pocillopora* photos to experimentally obtained biological values (symbiont biomass).

Artificial Intelligence tools can be used to extract physiological hallmarks from coral colony photographs

Although we showed that manually annotating the TPE image dataset post expedition is feasible, robust, and biologically sound, this approach took close to five months to be completed (not including photo cleaning). Thus, to accelerate and automate annotations of individual coral colonies, separate AI machines were developed for each of the different attributes isolated from the manual annotations. The architecture of our machines is shown in Figure 3. Transfer learning was used to cope with the the relatively small size of our photographic dataset (He et al., 2015). Training was not conducted on the ResNet Convolutional Neural Network, instead weights were taken from this CNN which is trained on the ImageNet Dataset containing millions of pictures (Deng et al, 2009). The pre-trained ResNet network was fine-tuned via a classifier built on top of 3 linear layers. ReLU was used as the activation function and a 40% dropout rate was implemented to avoid overfitting. SoftMax was eventually used to obtain the final probability distribution of the classification outcomes.

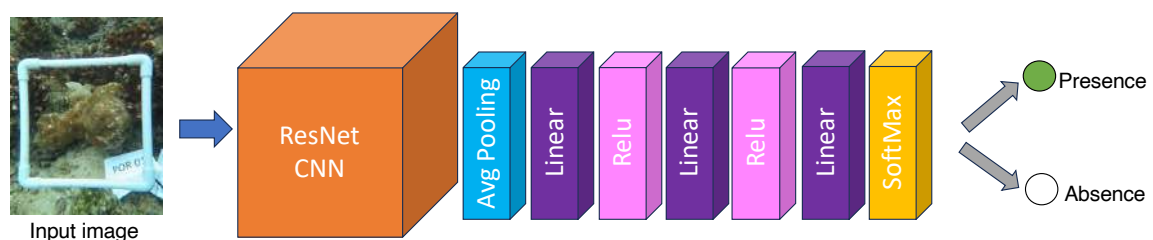


Figure 24 : Network architecture for picture classification showing untrained ResNet Convolutional Neural Network (orange cube) and three fully connected layers (purple linear layers) finetuned to our dataset.

To test the success of AI machines, various metrics can be utilized with common ones being “Accuracy” or “F1 score”. Accuracy compares the model’s prediction with known values to

assess how well the model can predict a target variable (Ismail & Wediawati, 2023). However, a training dataset that is mostly comprised of one class would result in high accuracy scores because the model has learned how to predict the majority class. F1 score is the harmonic mean of precision (true positive/(true positive + false positive)) and recall (true positive/(true positive + false negative)), combining them into a single metric (Ismail & Wediawati, 2023). In the above example of a skewed dataset, the F1 score would result in low precision and recall on the minority class. Therefore, accuracy was not the best metric to measure the success of our model as the training dataset was skewed towards samples showing absence. Thus, we decided to use F1 score for evaluation as it is better adapted for unbalanced datasets (Brownlee, 2020).

Table 3 shows the F1 scores for each of the nine attributes utilizing the cleaned Tara Pacific dataset (N= 5555) for training and validation. F1 scores range from 0 to 1 with a higher score indicating better precision and recall (Ismail & Wediawati, 2023). F1 scores for class 0 were higher than those for class 1 with the average F1 score for class 0 (absence) being 0.80 with values ranging from 0.65- 0.98 and the average F1 score for class 1 (presence) being 0.27 with values ranging from 0.06-0.59. These results can be attributed to the unbalanced nature of the dataset with class 0 having higher class coverage than class 1 (Table 3)(Figure 4A). On average across attributes, class 1 was represented in only 17% of photographs. The categories which had more balanced classes *i.e.*, higher representation of class 1 such as turf algae, and bleached had the highest F1 scores for class 1 (Table 3). These results showed that utilizing AI to automatically annotate coral colony photographs from the Tara Pacific dataset is possible but that balanced datasets or in our case more example photographs showing presence are necessary to create higher performing machines.

Feature	F1-Score (Class 0)	F1-Score (Class 1)	Size	Class 0 Coverage	Class 1 Coverage	Precision	Recall
Boring Organism	0.88	0.29	5555	3924	1631	0.28	0.29
Tissue Appearance	0.98	0.06	5555	5405	150	0.03	0.66
Algal Contact	0.75	0.16	5555	5003	552	0.10	0.37
CCA	0.81	0.16	5555	4448	1107	0.15	0.16
Pigmentation	0.82	0.19	5555	4966	589	0.10	0.97
Predation	0.77	0.31	5555	4537	1018	0.20	0.65
Turf Overgrowth	0.65	0.59	5555	3370	2185	0.39	0.52
Sediment Contact	0.73	0.16	5555	4892	663	0.10	0.35
Bleached	0.84	0.51	5555	3863	1692	0.28	0.30

Table 3: Results comprising F1 scores for Class 0 (absence) and Class 1 (presence) from initial models created using the cleaned dataset (N=5555). Coverage per class is shown for binary variables (0= absence, 1= presence).

To increase performance, we created a dataset with more balanced classes by randomly selecting and transforming (rotating, cropping, zooming) photographs to augment the dataset (Mumuni & Mumuni, 2022) (Figure 4B). Photos were augmented on a per class basis by increasing the ratio of photographs showing absence and presence to 60/40 which was the ratio of the least skewed attribute (turf overgrowth) from our original results (Table 3) (Figure 4C). Figure 4A and 4C show the differences in class coverage between the cleaned and the augmented dataset with the augmented dataset showing much more balanced classes with class 1 now being represented in 40% of photographs. As expected, the augmented dataset (N=6278-8663), lead to increases in F1 scores for class 1. F1 scores for class 0 averaged 0.86 ranging from .0.74 to 0.96 and F1 scores for class 1 averaged 0.79 with a range from 0.62 to 0.93 (Table 4). By augmenting the dataset to create more balanced classes we were able to increase the performance of our machines across attributes.

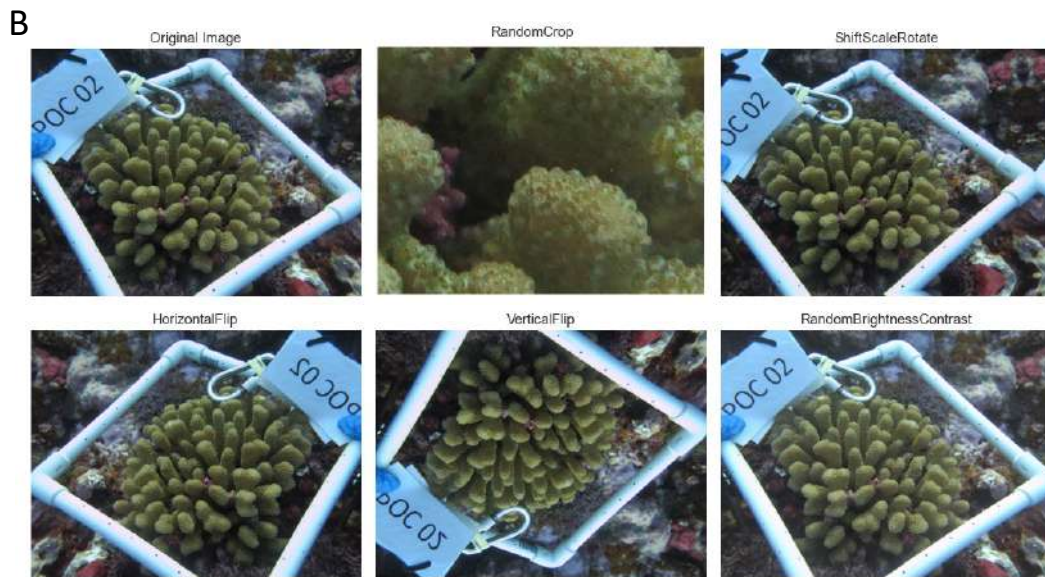
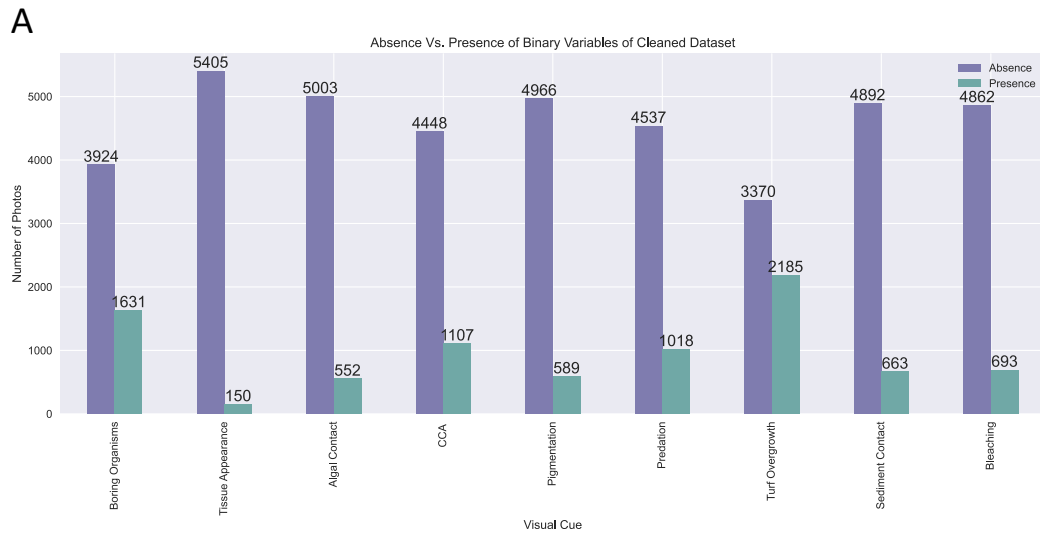


Figure 25: (A) Values distribution for binary variables (Boring Organisms, Tissue Appearance, Algal Contact, CCA, Pigmentation, Predation, Turf Overgrowth, Sediment Contact and Bleaching) of the original dataset ($N=5555$). (B) Examples of different transformations possible (random crop, shift scale rotate, horizontal flip, vertical flip, and random brightness contrast) as compared to the original image. (C) Values distribution for binary variables of augmented datasets.

Feature	F1-Score (Class 0)	F1-Score (Class 1)	Size	Class 0 Coverage	Class 1 Coverage	Precision	Recall
Boring Organism	0.83	0.72	6278	3924	2354	0.72	0.66
Tissue Appearance	0.96	0.93	8663	5405	3258	0.92	0.94
Algal Contact	0.90	0.83	8005	5003	3002	0.80	0.87
CCA	0.84	0.74	7117	4448	2669	0.74	0.71
Pigmentation	0.90	0.83	7945	4966	2979	0.81	0.86
Predation	0.86	0.78	7259	4537	2722	0.83	0.74
Turf Overgrowth	0.74	0.62	7583	4600	2983	0.68	0.59
Sediment Contact	0.89	0.84	7827	4892	2935	0.92	0.77
Bleached	0.86	0.80	7779	4862	2917	0.88	0.78

Table 4: Results comprising F1 scores for class 0 (absence) and class 1 (presence) from machines created using augmented datasets (N=6278-8663). Results are shown for binary variables, for “Bleached” 0 = Normal/Dark, 1= Bleached/Light.

Through the analysis of false negatives and false positives, we can gain more insight into why certain predictions were made. Figure 5 shows examples of predictions generated by our machine built for identifying boring organisms with examples of correct and incorrect predictions for each class. A thorough analysis needs to be conducted to determine trends in photographs classified incorrectly (false positives and false negatives). Such an analysis on the results of the classifiers can provide insights and guidelines for further analysis/annotations/photoshoot procedures.

Finally, to confirm that the individual machines were distinguishing distinct features from the images and thus that we had created unique machines specialized to each visual cue, cosine similarity was calculated on the average of the final three layers. The final three layers were the layers trained on the Tara Pacific data. The cosine similarity measure is a well-known method to evaluate the distance between vectors in geometric space, by calculating the cosine of the angle formed by two vectors with values approaching 1 implying the two vectors are colinear and values close to 0 being perpendicular. The cosine similarity analysis between individual machines showed values close to 0 indicating that the two vectors of the different machines are not the same and therefore supporting the idea that each machine is learning unique features and confirming the need for different machines for each attribute (Table S2).

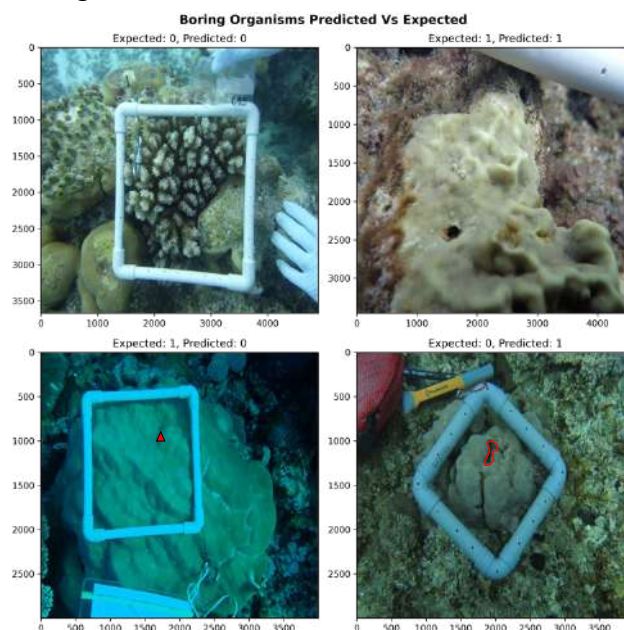


Figure 26: Results from machine for Boring Organisms showing an example of predicted and expected results for each class

Novel dataset confirms ability of machines to discern coral colony attributes from photographs alone

For further evaluation of our machines, we tested their performance on novel photographs that did not originate from the Tara Pacific dataset. For this we took advantage of a longitudinal study (2020 – ongoing) that was set up in Moorea, French Polynesia, which was one of the islands visited during the original Tara Pacific expedition (Planes et al., 2019). The longitudinal study aims to follow specific tagged colonies over an extended period (years, decades) and builds off the Tara Pacific photographic protocol, taking three photographs (one wide-angle, one standard (colony as a whole) and one macro, (Fig S1C) of each colony. Photographs were taken yearly of each individual colony (10 per genera) at three sites (Figure 6).

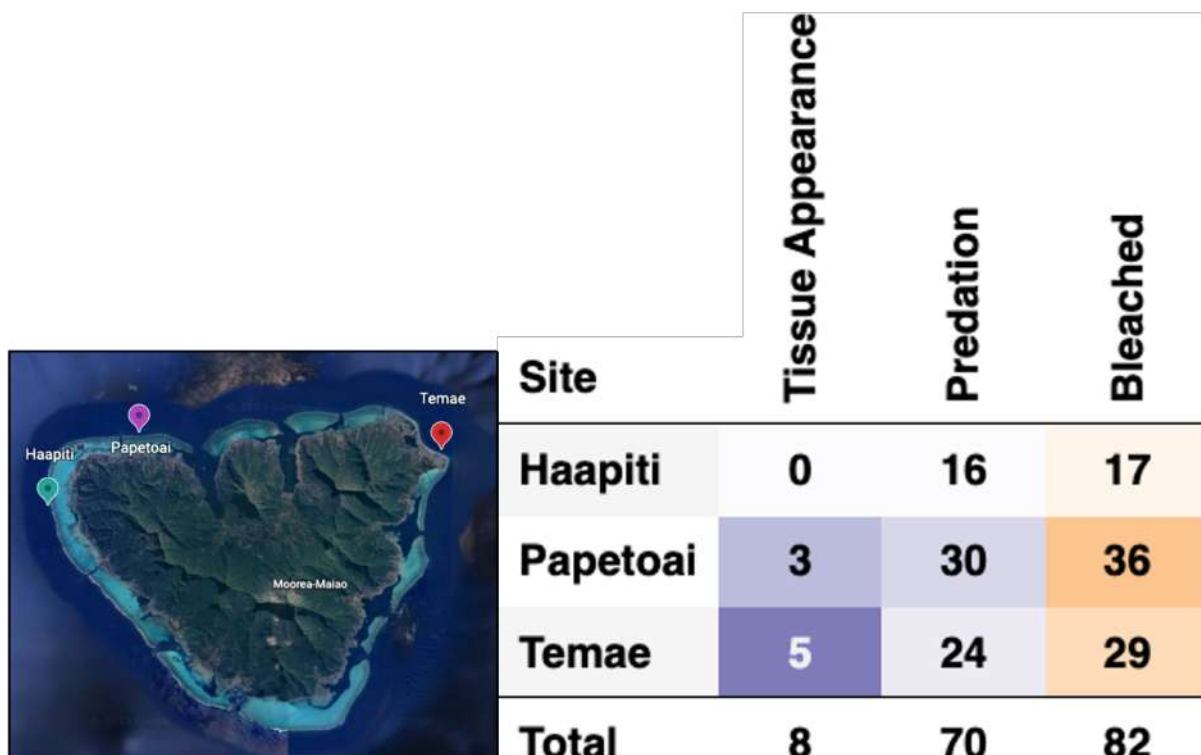
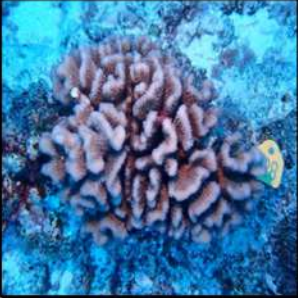



Figure 27: Map of Moorea with three sites: Haapiti, Papetoai, and Temae (left) along with overview of the number of photographs per site showing presence of the three in-water annotation categories (Tissue Appearance, Predation, and Bleached). Please note that Haapiti was added in 2022, so only has one year of data whereas Papetoai and Temae were established in 2021 and thus have two years of data.


In addition to the photographs, *in situ* surveys were conducted for each photographed colony recording information about the percent of living tissue, predation, disease, and coloration. Figure 6 shows the breakdown of photographs showing presence for the three attributes by site. The standard “colony” photographs from the Moorea study (Figure S1Cb) were used as a testing set for the machines for “Tissue Appearance”, “Bleached”, and “Predation” as these had the in-water manual annotations to use as ground truth. Given that the photographs utilized for the testing set are taken with a different protocol in respect to the Tara expedition, we investigated three pre-processing steps to see if preprocessing affected overall results. Photographic pretreatments included 1. resizing, 2. center crop and 3. resize and center crop (Table 5A, B, C).



A.RESIZE (224,224)									
	Support		F1		Precision		Recall		
	0	1	0	1	0	1	0	1	
Tissue Appearance	112	5	0.93	0.11	0.96	0.08	0.89	0.2	
Bleached	50	67	0.61	0.29	0.46	0.75	0.92	0.18	
Predation	93	24	0.87	0.42	0.85	0.47	0.89	0.38	



B. CENTERCROP (224,224)									
	Support		F1		Precision		Recall		
	0	1	0	1	0	1	0	1	
Tissue Appearance	112	5	0.07	0.07	0.8	0.04	0.04	0.8	
Bleached	50	67	0.2	0.72	0.6	0.59	0.12	0.94	
Predation	93	24	0.23	0.36	0.92	0.22	0.13	0.96	



C.RESIZE (1024) + CENTERCROP (224, 224)									
	Support		F1		Precision		Recall		
	0	1	0	1	0	1	0	1	
Tissue Appearance	112	5	0.56	0.05	0.94	0.03	0.4	0.4	
Bleached	50	67	0.48	0.62	0.49	0.62	0.48	0.63	
Predation	93	24	0.77	0.51	0.92	0.37	0.66	0.51	

Table 5: Results from a testing set for the machines Tissue Appearance, Bleached, and Predation for our machines trained on the Tara Pacific data with the best score denoted with a green square. Photos used for the testing set were novel photos never used in training or validation which came from our Moorea longitudinal study. Three different pretreatments were used with a sample image shown to the left of each table. A. shows the pretreatment “resized”, B shows “center crop” as a pretreatment and c. shows “resized and center crop”.

Based on the performances observed on running the machines on images from or generated from the Tara Pacific dataset, the AI machines we developed appeared less efficient at determining the presence (1) of attributes using this new image dataset (Table 5A, B, C). This was however attribute-specific, and across all three attributes, the F1-scores range from 0.11 (poor efficiency) to 0.72 (correct efficiency) for class one. In particular, the AI machine for determining “Tissue Appearance” reached a maximum F1 score of only 0.11 for class 1 compared to .51 for predation and .51 for bleached (Table 5A)

Of the three attributes, Tissue appearance consistently had the lowest F1 scores for class 1 but was also the most skewed with only 5 examples of presence compared to 112 examples of absence. Such a small class size means that even if all five samples were correctly classified (true positives), any misclassifications (false positives) of the other 112 samples would reduce the F1 score as the weight of the true positives is so small. On the other hand, the AI machine to determine “Predation” worked significantly better, *i.e.*, F1-score of 0.42 after resizing, and even 0.51 after resizing and center crop (Table 5A, C)-AI machine on this Moorea image dataset was the one to determine the bleaching level. In fact, following resizing and center crop the F1-score was 0.62, and with center crop alone increased to 0.71 (Table 5C, B). The

distribution of samples showing presence and absence was the most balanced for the attribute “Bleached” and achieved the highest F1 score of the three variables.

Overall, this new testing set shows us that the AI machines we developed perform very well on the Tara Pacific image Dataset. This is further highlighted by the differences in machine performance based on pretreatment when using photographs from the Moorea dataset. Although, less performant than on the Tara Pacific images, the AI machines were still able to recognize at a significant level the presence of “bleaching” and or “predation” on datasets even with a limited number of images (Table 5).

Discussion:

In this study, photographs from the Tara Pacific Expedition were manually annotated for specific attributes that may be linked to physiological states of the coral colony, these images were then curated to create a refined dataset adapted for an AI approach. This process was time consuming, taking approximately 4 months for the annotations alone with several additional months necessary to clean and refine the dataset. Implementing artificial intelligence for image analysis provides the possibility to resolve the current bottleneck that exists in photographic monitoring. By creating an AI model to determine visual cues relevant to coral colony physiology / health, a decrease in both cost and time can be achieved as well as the creation of an accessible tool for various members of the conservation and scientific community. To decrease annotation time, we used our refined dataset along with corresponding annotations to develop AI machines that identify these annotations automatically. After training, which takes approximately 15 hours per machine, new photographs can be analyzed in seconds. Our model provides the user with visual information about the colony that would otherwise need to be monitored *in situ* or annotated post dive as was done for the photos from the Tara Pacific Expedition.

Gathering information about individual coral colonies can have substantial expenses in terms of time and/or money (Durden et al., 2016) (Langenkämper et al., 2017) (González-Rivero et al., 2020) (Leujak & Ormond, 2007). Examples of financial costs could be salaries/personnel costs or equipment such as SCUBA gear, camera(s), software, and or surveying tools. Temporal expenditures depend on the methodology chosen and whether annotations are conducted during data collection or post data collection. For example, monitoring colonies *in situ* may require more dives or longer bottom time incurring higher costs (temporal and monetary) during data collection but lower costs for data analysis. Inversely, collecting photographic data would result in lower expenditures during data collection and higher expenditures for analysis. As a concrete example, González-Rivero et al. (2020) estimated that expert analysis of a single image costs around 5\$ AUD which would result in high analysis costs for large datasets. The use of our AI model would result not only in time saved during data collection by taking photographs instead of individual inspection of each colony but also time saved post data collection via machine analysis instead of through a human inspector. Combined, this would result in decreased costs by removing the need to hire a human expert annotator and minimizing time in the field and related costs. Thus, we propose a tool that capitalizes on the efficiency of photographic monitoring without the expensive and time-consuming commitment of manual annotation.

Our current model which exploits the visual information from colony photographs is the first step towards the development of more comprehensive models to determine health status. For such a model, health needs to be clearly defined with specific metrics for assessment. The manner with which we define health will help to define a research focus. For example, if the

definition of coral health is centered around the presence or absence of disease this would lead to a focus on eliminating and treating diseases. Thus, we would like to highlight the need for a global picture of coral health which combines various factors both visually apparent and molecular and would result in a focus on the conservation and preservation of coral reefs as a whole. As the photographs that were used to create these machines were generated during the Tara Pacific Expedition, each photographed colony is linked to the wealth of data that was collected during this Expedition including environmental (historical and contextual) and physiological data (telomere length and stress biomarkers).

Thus, future research aims include incorporating the molecular and environmental data of each individual colony with the visual cues annotated from the photographs to gain a better understanding of the state of each individual colony at the time of sampling and to create next-generation machines linking physiological state with visual cues. This will be done by exploiting artificial intelligence tools to match molecular and cellular parameters to photographic information. Furthermore, individual colonies will be followed across time points as changes in condition over time is an important measure of health. This data will be generated from the longitudinal study currently underway in Moorea, French Polynesia where tagged colonies will be surveyed annually utilizing the Tara methodology including both photographic and physical sampling. Thus, changes in health status as measured by mortality or partial mortality can be followed for individual coral colonies to determine which cues both physical and/or molecular are better predictors of colony survival.

To further improve upon the machines, a key step will be methodology optimization. One of the major challenges we faced with our dataset arose from disparities within the photos. These photographs were not taken with the objective of creating an artificial intelligence model. Thus, the protocol for taking pictures was not appropriately strict and resulted in incongruous photographs, creating noise. This is demonstrated in Figure S1Ba-f which shows examples of how the appearance of target colonies differed between photographs in terms of distance from the camera, placement within the camera frame (centering), presence/placement of the quadrat and camera orientation (birds eye view, side-view, at an angle...*etc.*). If we look at the classic AI Datasets (MNIST (fashion), MNIST, CIFAR, IMAGENET) the photos are homogenous with the object of interest typically centered or in the same location, and all photos being the same size and resolution. This is further exemplified in AI for medical imagery where scans produce very similar output facilitating “learning” by the machine.

This has been confirmed within our own study by our use of the Moorea longitudinal dataset. These images were more homogenous than the Tara dataset and our machines achieved lower F1 scores when classifying these images and varied depending on pretreatment. Taking the attribute bleached as an example, the increase in F1 score when center crop was used as a pretreatment for class 1 could be due to the stricter photo taking methodology used during the Moorea longitudinal study. Here, divers were instructed to center the colony and try to have it take up most of the frame meaning that a center crop pretreatment would have more of a chance of focusing on the colony of interest and eliminating noise than it would have from a Tara Pacific photograph (Table S1B). This might be especially relevant to a category like bleaching where colony color is important. When looking at F1 scores for class 1 for the attribute predation, photographs that were resized and center cropped achieved the best results, followed by resize and then center crop (Table 5). This dataset was slightly skewed towards absence but achieved decent results for resize and resize and center crop indicating that perhaps there is important information at the edges of the colonies that may get cut off with center crop alone. Thus, highlighting the need for a strict photographic methodology to choose a proper

pretreatment for all images and increase learning and overall performance by the machines when presented with novel images.

In addition to these discrepancies within our dataset, it is important to note that underwater photography already poses its own set of challenges especially in terms of light and color. Light is principally affected by depth and visibility but can also vary depending on other factors such as: time of day, surface conditions, location, and weather (Pateman, 2009). In terms of color, as depth increases different lengths of wavelengths are absorbed resulting in color loss with reds and oranges as the first to be lost and blues and violets as the last (Pateman, 2009). This means that colors are perceived differently across depths and the same subject can photograph very differently depending not only on depth but also on the above-mentioned factors. Due to this, certain annotations, specifically bleaching (light, dark, pale, or bleached) were exceedingly difficult to conduct based on the photographs alone and ended up being more subjective (Figure 2: Ac-f & Bc-f). To counter this, we incorporated a different methodology for annotating colony coloration (bleaching). First, coloration annotations were all conducted *in situ* to remove confounding variables such as weather, light conditions, photoquality, and the effects of depth on photo color. Secondly, the Coral Watch Color Card (Siebek et al., 2006) was used to determine bleaching on a six-point scale to minimize individual bias and increase replicability. With these changes in place, we plan to utilize our new photo dataset to reduce individual bias and improve classification results for the attribute bleaching.

Furthermore, through analysis of the False Positives and False Negative as predicted by the machines (Figure 5), we will better understand the machine predictions to further refine methodology and increase performance. For example, figure 5 shows machine predictions on the cleaned Tara datasets for boring organisms. The false negative (bottom left) shows a boring organism that the machine missed in the center of the photograph (as indicated by a red arrow). For the original annotators, the annotator (MC) often zoomed in on the images to fully evaluate the colony, the boring organism present in this photo could easily be overlooked without close inspection and perhaps was mistaken for a shadow or confounded by the noise in this photo (clipboard, carabiner... *etc.*). The false positive shows an algae patch that appears dark and is in the form of a figure -8 which are also the characteristics of bivalves (for a human annotator) and could be one explanation as to why the machine predicted bivalve. Further evaluation of the False Positives and False negatives will be conducted to extract more information about machine performance.

Finally, marine life has proven to be a difficult subject in terms of object detection due to intraclass differences, species/subjects spanning large areas, and camouflaged appearance (Fu et al., 2023). This highlights the need to develop a strict methodology that focuses on the conditions and limitations of working in a liquid medium while also keeping the use of AI tools model in mind. An example of keeping AI tools in mind would be the grouping of boring organisms or algal genera which facilitates learning and creates a higher-performing machine rather than individual machines specific to each boring organism or algal genera. In terms of photographic methodology, we have developed a stricter photographic methodology that attempts to consider camera angles and distances when taking the photographs ensuring that the target colony is centered which could lead to more effective machines resulting from image pretreatment (*i.e.*, center crop). Moving forward, as we continue to collect more training samples, we hope to create a larger more homogenous training dataset, that is less skewed which will not only increase the efficiency of the machines but also its efficiency in analyzing novel photographs.

Overall, this study has demonstrated that AI tools can be utilized on a photographic dataset which was not taken with an AI purpose in mind. However, to optimize the machines, a large homogenous dataset with an AI purpose in mind will need to be developed. In conclusion, this study paves the way for the development of a non-invasive machine that utilizes AI tools to assess coral colony health.

References

- Aeby GS, Williams GJ, Franklin EC, Kenyon J, Cox EF, Coles S, et al (2011) Patterns of Coral Disease across the Hawaiian Archipelago: Relating Disease to Environment. *PLoS ONE* 6(5): e20370. <https://doi.org/10.1371/journal.pone.0020370>
- Ainsworth TD, Kramasky-Winter E, Loya Y, Hoegh-Guldberg O, Fine M (2007) Coral disease diagnostics: what's between a plague and a band? *Appl Environ Microbiol.* 73(3):981-92. <https://doi.org/10.1128/AEM.02172-06>.
- Anthony K, & Larcombe P (2000) Coral reefs in turbid waters: sediment-induced stresses in corals and likely mechanisms of adaptation. 1:239-244.
- Baird AH, and Marshall PA (2002) Mortality, growth, and reproduction in scleractinian corals following bleaching on the Great Barrier Reef. *Marine Ecology Progress Series* 237:133-141. <https://doi.org/10.3354/meps237133>
- Barott KL, & Rohwer FL (2012) Unseen players shape benthic competition on coral reefs. *Trends in microbiology* 20(12):621–628. <https://doi.org/10.1016/j.tim.2012.08.004>
- Barott K, Smith J, Dinsdale E, Hatay M, Sandin S, Rohwer F (2009) Hyperspectral and Physiological Analyses of Coral-Algal Interactions. *PLoS ONE* 4(11): e8043. <https://doi.org/10.1371/journal.pone.0008043>
- Beijbom, Oscar & Edmunds, Peter & Kline, David & Mitchell, B. & Kriegman, David. (2012). Automated Annotation of Coral Reef Survey Images. *Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* 1170-1177. [10.1109/CVPR.2012.6247798](https://doi.org/10.1109/CVPR.2012.6247798).
- Bindoff NL, Cheung WWL, Kairo JG, Arístegui J, Guinder VA, Hallberg R, Hilmi N, Jiao N, Karim MS, Levin L, O'Donoghue S, Purca Cuicapusa SR, Rinkevich B, Suga T,
- Brownlee, J. (2020). *Imbalanced classification with Python: Better Metrics, Balance Skewed Classes, Cost-Sensitive Learning, Machine Learning Mastery.*
- Tagliabue A, and Williamson P (2019) Changing Ocean, Marine Ecosystems, and Dependent Communities. *IPCC Special Report on the Ocean and Cryosphere in a Changing World.* Cambridge University Press:447-587. <https://doi.org/10.1017/9781009157964.007>.
- Breslow L (1972). A Quantitative Approach to the World Health Organization Definition of Health: Physical, Mental and Social Well-being. *International Journal of Epidemiology*, 1(4): 347–355. <https://doi.org/10.1093/ije/1.4.347>
- Burns JHR, Weyenberg G, Mandel T, Ferreira SB, Gotshalk D, Kinoshita CK, ... Pelayo, R (2020) A Comparison of the Diagnostic Accuracy of in-situ and Digital Image-Based Assessments of Coral Health and Disease. *Frontiers in Marine Science* 7. <https://doi.org/doi:10.3389/fmars.2020.00304>

Cohen, J. (1960) A Coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20, 37-46.

Cetz-Navarro NP, Espinoza-Avalos J, Hernández-Arana HA, Carricart-Ganivet JP (2013) Biological Responses of the Coral *Montastraea annularis* to the Removal of Filamentous Turf Algae. *PLoS ONE* 8(1): e54810. <https://doi.org/10.1371/journal.pone.0054810>

Chow MH, Tsang RHL, Lam E & Ang P (2016) Quantifying the degree of coral bleaching using digital photographic technique. *Journal of Experimental Marine Biology and Ecology*. 479:60-68. <https://doi.org/10.1016/j.jembe.2016.03.003>.

Cox EF (1986) The effects of a selective corallivore on growth rates and competition for space between two species of Hawaiian corals. *Journal of Experimental Marine Biology and Ecology* 101: 161-174. [https://doi.org/10.1016/0022-0981\(86\)90047-X](https://doi.org/10.1016/0022-0981(86)90047-X)

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database.

Douglas AE (2003) Coral bleaching--how and why?. *Marine pollution bulletin* 46(4): 385–392. [https://doi.org/10.1016/S0025-326X\(03\)00037-7](https://doi.org/10.1016/S0025-326X(03)00037-7)

Durden JM, Bett BJ, Schoening T, Morris KJ, Nattkemper TW, Ruhl HA (2016) Comparison of image annotation data generated by multiple investigators for benthic ecology. *Mar Ecol Prog Ser* 552:61-70. <https://doi.org/10.3354/meps11775>

Ezzat L, Lamy T, Maher RL, et al. (2020) Parrotfish predation drives distinct microbial communities in reef-building corals. *anim microbiome* 2:5 <https://doi.org/10.1186/s42523-020-0024-0>

Fitt WK, McFarland F, Warner ME, & Chilcoat GC (2000) Seasonal patterns of tissue biomass and densities of symbiotic dinoflagellates in reef corals and relation to coral bleaching. *Limnology and Oceanography*, 45:677-685 <https://doi.org/10.4319/lo.2000.45.3.0677>

Flower J, Ortiz JC, Chollett I, Abdullah S, Castro-Sanguino C, Hock K, Lam V, Mumby PJ (2017) Interpreting coral reef monitoring data: A guide for improved management decisions. *Ecological Indicators*. 72:848–869. <https://doi.org/10.1016/j.ecolind.2016.09.003>

Fried LP, Cohen AA, Xue QL, Walston J, Bandeen-Roche K, & Varadhan R (2021) The physical frailty syndrome as a transition from homeostatic symphony to cacophony. *Nature Aging* 1(1):36–46. <https://doi.org/10.1038/s43587-020-00017-z>

Fu, C, Liu R, Fan X, Chen P, Fu H, Yuan W, ... & Luo Z (2023). Rethinking general underwater object detection: Datasets, challenges, and solutions. *Neurocomputing*, 517: 243-256.

GCRMN (n.d.) The Sixth Status of Corals of the World: 2020 Report. Retrieved July 29, 2022 from <https://germn.net/2020-report/>

Glynn PW (1993) Coral reef bleaching: ecological perspectives. *Coral Reefs* 12:1–17 <https://doi.org/10.1007/BF00303779>

- González-Rivero M, Beijbom O, Rodríguez-Ramírez A, Bryant DEP, Ganase A, González-Marrero Y,... Hoegh-Guldberg O (2020) Monitoring of Coral Reefs Using Artificial Intelligence: A Feasible and Cost-Effective Approach. *Remote Sensing* 12(3):489. <https://doi.org/10.3390/rs12030489>
- Grupstra CGB, Rabbitt KM, Howe-Kerr LI, et al. (2021) Fish predation on corals promotes the dispersal of coral symbionts. *anim microbiome* 3:25 <https://doi.org/10.1186/s42523-021-00086-4>
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep Residual Learning for Image Recognition* (arXiv:1512.03385). arXiv. <https://doi.org/10.48550/arXiv.1512.03385>
- Hoegh-Guldberg O (1999) Climate change, coral bleaching, and the future of the world's coral reefs. *Marine and Freshwater Research* 50(8): 839. <https://doi.org/10.1071/mf99078>
- Hughes T, Barnes M, Bellwood D, et al. (2017) Coral reefs in the Anthropocene. *Nature* 546:82–90. <https://doi.org/10.1038/nature22901>
- Ismail, A., & Wediawati, D.B. (2023). *Understanding the Fundamentals of Machine Learning and AI for Digital Business*. Asadel Publisher
- Kubomura T, Wee HB, & Reimer JD (2020) Investigating incidence and possible causes of pink and purple pigmentation response in hard coral genus *Porites* around Okinawajima Island, Japan. *Regional Studies in Marine Science* 101569. <https://doi.org/10.1016/j.rsma.2020.101569>
- Langenkämper D, Zurowietz M, Schoening T, & Nattkemper TW (2017) BIIGLE 2.0 - Browsing and Annotating Large Marine Image Collections. *Frontiers in Marine Science* 4. <https://doi.org/10.3389/fmars.2017.00083>
- Lenth RV, Bolker B, Buerkner P, Giné-Vázquez I, Herve M, Jung M, et al. emmeans: Estimated Marginal Means, aka Least-Squares Means. 2023. <https://cran.r-project.org/web/packages/emmeans/index.html>
- Leujak W, & Ormond RFG (2007) Comparative accuracy and efficiency of six coral community survey methods. *Journal of Experimental Marine Biology and Ecology*, 351(1-2):168–187. <https://doi.org/10.1016/j.jembe.2007.06.028>
- Lirman, D (2001) Competition between macroalgae and corals: effects of herbivore exclusion and increased algal biomass on coral survivorship and growth. *Coral Reefs* 19:392–399. <https://doi.org/10.1007/s003380000125>
- Little MM, & Little DS (2013). The Nature of Crustose Coralline Algae and Their Interactions on Reefs. *Research and Discoveries: The Revolution of Science through Scuba* 199- 212
- Lombard F, Bourdin G, Pesant S, Agostini S, Baudena A, ... et al. (2022) Open science resources from the Tara Pacific expedition across coral reef and surface ocean ecosystems. Preprint. <https://doi.org/10.1101/2022.05.25.493210>

- López-Otín C, & Kroemer G (2020). Hallmarks of Health. *Cell*. <https://doi.org/10.1016/j.cell.2020.11.034>
- Mahmood A, Bennamoun M, An S, Sohel F, Boussaid F, Hovey R, Kendrick G, & Fisher RB (2016) Automatic annotation of coral reefs using deep learning. *IEEE, Institute of Electrical and Electronics Engineers*. <https://doi.org/10.1109/OCEANS.2016.7761105>
- Manderson T, Li J, Dudek N, Meger D, and Dudek G (2017), Robotic Coral Reef Health Assessment Using Automated Image Analysis. *J. Field Robotics*, 34: 170-187. <https://doi.org/10.1002/rob.21698>
- McCook LJ (1999) Macroalgae, nutrients and phase shifts on coral reefs: scientific issues and management consequences for the Great Barrier Reef. *Coral Reefs* 18: 357-367.
- McCook L, Jompa J, & Diaz-Pulido G (2001) Competition between corals and algae on coral reefs: a review of evidence and mechanisms. *Coral Reefs* 19:400–417. <https://doi.org/10.1007/s003380000129>
- Moberg F, & Folke C (1999) Ecological goods and services of coral reef ecosystems, *Ecological Economics*, 29(2):215-233. <https://EconPapers.repec.org/RePEc:eee:ecolec:v:29:y:1999:i:2:p:215-233>.
- Mumuni A., & Mumuni, F. (2022). Data Augmentation: A comprehensive survey of modern approaches. *Array*, 16, 100258. <https://doi.org/10.1016/j.array.2022.100258>
- Nugues MM, Smith GW, van Hooidek RJ, Seabra MI, and Bak RPM (2004) Algal contact as a trigger for coral disease. *Ecology Letters* 7: 919-923. <https://doi.org/10.1111/j.1461-0248.2004.00651.x>
- Page C, Field S, Pollock F, Lamb J, Shedrawi G, & Wilson S (2016) Assessing coral health and disease from digital photographs and in situ surveys. *Assessing coral health and disease from digital photographs and in situ surveys*. 189. <https://doi.org/10.1007/s10661-016-5743-z>.
- Palmer CV, Modi CK, Mydlarz LD (2009) Coral Fluorescent Proteins as Antioxidants. *PLoS ONE* 4(10): e7298. <https://doi.org/10.1371/journal.pone.0007298>
- Palmer C, Mydlarz L, & Willis B (2008) Evidence of an inflammatory-like response in non-normally pigmented tissues of two scleractinian corals. *Proceedings. Biological sciences / The Royal Society*. 275:2687-93. <https://doi.org/10.1098/rspb.2008.0335>.
- Pateman V, (2009) Color correction for underwater photography, *Graphic Communication Department, College of Liberal Arts, California polytechnic state University*.
- Planes S, Allemand D, Agostini S, Banaigs B, Boissin E, et al. (2019) The Tara Pacific expedition—A pan-ecosystemic approach of the “-omics” complexity of coral reef holobionts across the Pacific Ocean. *PLOS Biology* 17(9): e3000483. <https://doi.org/10.1371/journal.pbio.3000483>

Pratte ZA, Longo GO, Burns AS, et al. (2018) Contact with turf algae alters the coral microbiome: contact versus systemic impacts. *Coral Reefs* 37: 1–13. <https://doi.org/10.1007/s00338-017-1615-4>

R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2020. <https://www.R-project.org/>.

Raymundo LJ, Couch CS, Harvell CD, Raymundo J, Bruckner AW, Work TM et al. (2008) Coral disease Handbook: Guidelines for Assessment, Monitoring, and Management. Technical report, Coral Reef Targeted Research & Capacity Building for Management Program

Sammarco PW & Risk MJ (1990) Large-scale patterns in internal bioerosion of Porites: cross continental shelf trends on the Great Barrier Reef. *Marine Ecology Progress Series*:145-156

Scaps P, & Denis V (2008) Can organisms associated with live scleractinian corals be used as indicators of coral reef status? *Atoll Research Bulletin* 566. <https://doi.org/10.5479/si.00775630.566.1>

Scott PJB, & Risk MJ (1988) The effect of Lithophaga (Bivalvia: Mytilidae) boreholes on the strength of the coral Porites lobata. *Coral reefs* 7:145-151.

Siebeck, U. E., Marshall, N. J., Klüter, A., & Hoegh-Guldberg, O. (2006). Monitoring coral bleaching using a colour reference card. *Coral Reefs*, 25(3), 453–460. <https://doi.org/10.1007/s00338-006-0123-8>

Smith S, Harriot V (1998) Tube-building polychaete worms smother corals in the Solitary Islands Marine Park, northern NSW, Australia. *Coral Reefs* 17:342. <https://doi.org/10.1007/s003380050136>

Stokes MD, Deane GB. 2009. Automated processing of coral reef benthic images. *Limnol Oceanography: Methods* 7: 157–168. Syms C, Jones GP. 2000. Disturbance, habitat structure, and the dynamics of coral-reef fish community. *Ecology* 81: 2714–2729.

Swanson D, Bailey H, Schumacher B, Ferguson M, & Vargas-Ángel B (2018) Ecosystem Sciences Division Standard Operating Procedures: Data Collection for Rapid Ecological Assessment Benthic Surveys. NOAA Tech. Memo. NMFS-PIFSC-71:1-63. <https://doi.org/10.25923/39jh-8993>

Rogers CS (1990) Responses of coral reefs and reef organisms to sedimentation. *Marine Ecology Progress Series*, 62:185-202.

Rogers CS, & Miller J (2013) Coral Diseases Cause Reef Decline. *Science* 340(6140):1522–1522. <https://doi.org/10.1126/science.340.6140.1522-a>

Rotjan R & Lewis S (2008) Impact of coral predators on tropical reefs. *Marine Ecology Progress Series* 367:73-91. <https://doi.org/10.3354/meps07531>.

Rushmore ME, Ross C, & Fogarty ND (2021) Physiological responses to short-term sediment exposure in adults of the Caribbean coral *Montastraea cavernosa* and adults and recruits of *Porites astreoides*. *Coral Reefs* 40:1579–1591. <https://doi.org/10.1007/s00338-021-02156-0>

Vermeij M, Dailer M, & Smith C (2011) Crustose coralline algae can suppress macroalgal growth and recruitment on Hawaiian coral reefs. *Marine Ecology Progress Series*. 422:1-7. <https://doi.org/10.3354/meps08964>.

Ward S, Harrison P, & Hoegh-Guldberg O (2000) Coral bleaching reduces reproduction of scleractinian corals and increases susceptibility to future stress. *Proc. 9th Int. Coral Reef Symp.*2.

Wickham H, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke C, et al. *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. 202 <https://cran.r-project.org/web/packages/ggplot2/index.html>

Weiss A, & Martindale RC (2017) Crustose coralline algae increased framework and diversity on ancient coral reefs. *PLoS One* 12(8): e0181637.

WHO (1948) Constitution of the World Health Organization. http://www.who.int/governance/eb/who_constitution_en.pdf

Wild C, Jantzen C, & Kremb SG (2014). Turf algae-mediated coral damage in coastal reefs of Belize, Central America. *PeerJ* 2:e571. <https://doi.org/10.7717/peerj.571>

Zhao, W. (2017). Research on the deep learning of the small sample data based on transfer learning. *AIP Conference Proceedings*, 1864(1), 020018. <https://doi.org/10.1063/1.4992835>

Supplementary Information:

The following information may be found below or attached to this article:

Github Repositories:

<https://github.com/UCA-MSI/tara-classification> : Code related to transfer learning and classification

<https://github.com/UCA-MSI/tara-augment>: Code related to photographic augmentation

Attached:

Figure S1. Example Photographs from the Tara Pacific Expedition.

Table S1. Table showing summary statistics per island of cleaned photographic dataset.

Table S2. Table showing Cosine similarity values comparing the final three layers for each machine.

Code S1. MATLAB code used to conduct the manual photographic annotations.

3.3 Supplementary Information

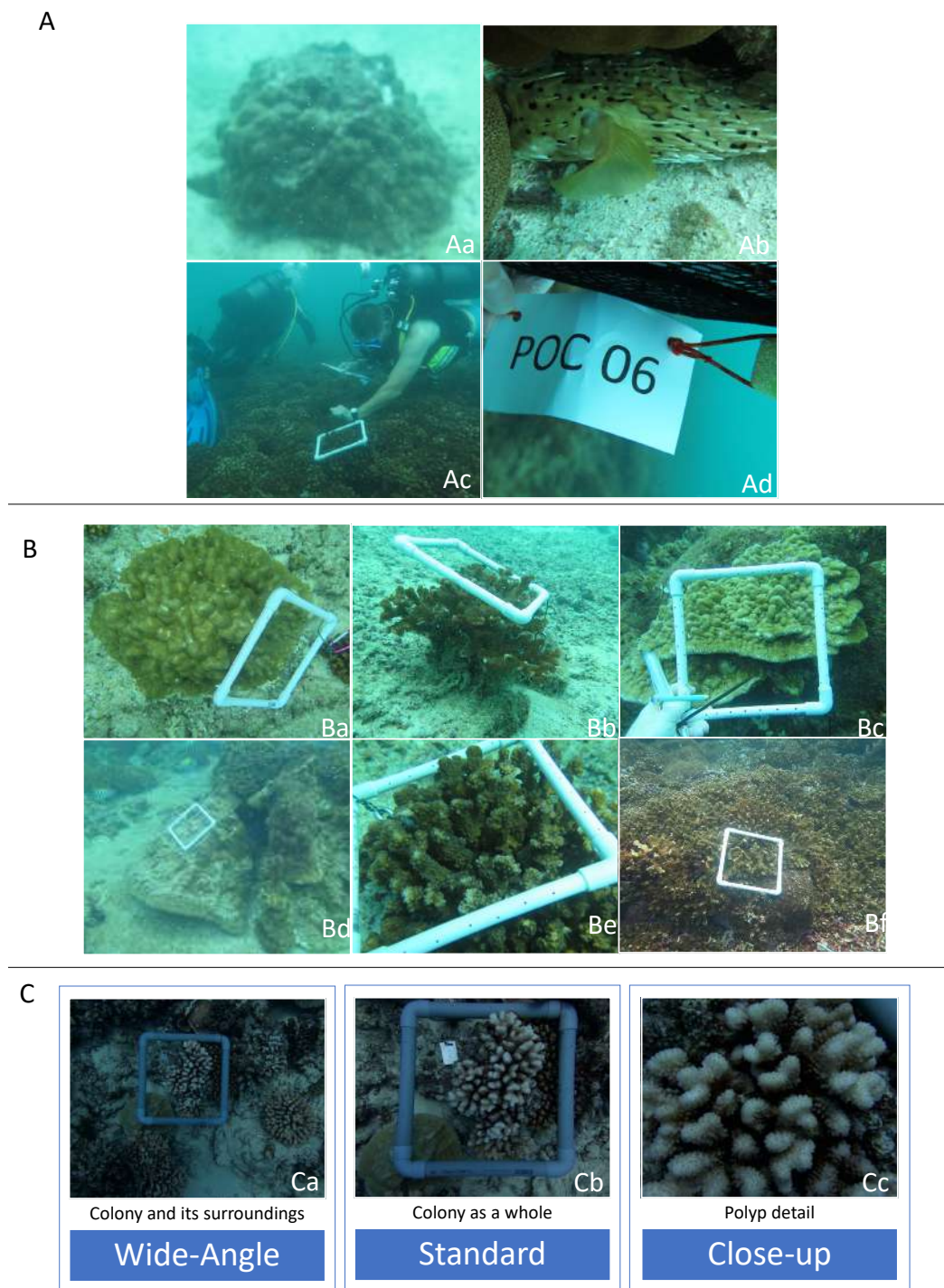


Figure S 1: Example of photographs utilized during this work. Photos in Panels A and B represent photos taken during the Tara Pacific Expedition. Photos from panel A represent photos that were excluded during data cleaning: poor photo quality/blurry (Aa), photo showing a non-coral target (Ab), photo with too much noise (Ac), photo of the colony ID tag (Ad). Photos in panel B (Ba-Bf) demonstrate the variability among images in terms of camera angle, quadrat placement and distance from camera to target colony. Photos in Panel C represent the three types of photos taken during the Moorea Longitudinal study: Aa: Wide angle photo showing the colony and its surroundings Ab: Colony photo showing the colony as a whole and Ac: Close up photo showing individual polyp detail

Summary Statistics of Photos per Colony per Island

	Island	Min Photos	Max Photos	Avg Photos	Total Photos	Total Colonies
0	01	3	33	8	513	24
1	02	1	15	6	494	37
2	03	1	3	2	70	23
3	04	1	2	2	103	20
4	05	1	4	2	117	21
5	06	2	2	2	174	30
6	07	1	3	2	179	30
7	08	1	4	2	183	31
8	09	1	3	2	176	31
9	10	1	4	2	183	32
10	11	2	7	3	228	32
11	12	1	8	2	158	46
12	13	1	10	2	190	52
13	14	1	4	2	221	72
14	15	1	4	2	235	66
15	16	1	8	2	207	50
16	17	2	7	4	324	30
17	18	1	3	2	173	30
18	19	1	6	2	266	38
19	20	1	6	3	280	47
20	21	1	5	3	289	40
21	22	2	2	2	90	15
22	23	2	2	2	92	16
23	24	1	2	2	89	15
24	25	1	9	3	228	31
25	26	1	4	2	95	15
26	28	1	3	2	198	42

Table S 1: Table showing summary statistics for photographic data per island from the cleaned Tara Pacific photographic dataset.

	Tissue Appearance	Algal Contact	CCA	Pigmentation	Predation	Turf Overgrowth	Sediment Contact	Bleached
Boring Organism	0.039800	0.000600	0.020100	0.021600	-0.016600	0.000800	0.025400	-0.002600
Tissue Apperance		0.019400	0.004200	0.010100	0.010400	-0.029900	0.023300	0.013800
Algal Contact			-0.007000	-0.020400	0.030600	0.042000	0.010000	0.005900
CCA				-0.006400	-0.010700	-0.002700	0.017000	0.030600
Pigmentation					0.005800	0.006300	0.013200	0.014100
Predation						0.005600	0.012600	0.006600
Turf Overgrowth							-0.040100	0.031800
Sediment Contact								0.002100

Table S 2: Table comparing the CSV between all the individual machines

SUPPLEMENTARY INFO: CODE S1

MATLAB code developed specifically for Tara Pacific Photographic annotation.



%% requires that you follow instructions at the following page to install listdlg2, except use my custom listdlg3 script
<https://www.mathworks.com/matlabcentral/answers/97420-how-do-i-specify-the-location-of-a-listdlg>

```
refresh toolboxcache clear all
```

```
close all
```

```
yesno = [true(),false(),NaN];
standardshots = ["standard_wideangle" "standard_closeup" "very_wide"
"other"];
borers = ["bivalve" "Spirobranchus" "Tridacna" "boringurchin"
"other_polychaete" "sponge" "other"];
algae = ["Halimeda" "Turbinaria" "Dictyota" "Lobophora" "cca" "Galaxaura"
"Sargassum" "other"];
targets = ["Porites","Pocillopora","Millepora","other"];
bleached = ["normal" "light" "bleached" "dark"];
othernotable = ["no" "other" "gallcrabs" "ascidians"];
poc_morphotypes = ["grandis" "elegans" "meandrina" "verrucosa" "damicornis"
"acuta" "capitata" "other"];
por_morphotypes = ["lobata" "lutea" "columnar" "knobby" "other"];
mil_morphotypes = ["platyphylla" "dichotoma" "other"];
quadrats =
["large_thin_white","small_fat_grey","large_mid_white","hammer","other","no
ne"];
quadrat_sizes = [21.75,19^2,20.75^2,28.5*10]; %% in cm2
quadrat_sizes_size = size(quadrat_sizes);

colnames = ["filename" "isntjunk" "hastarget" "widevsclose" "target"
"morphotype" "quadrat" "parrotfish_scars" "bivalve" "Spirobranchus"
"Tridacna" "boringurchin" "other_polychaete" "sponge" "otherborers"
"Halimeda" "Turbinaria" "Dictyota" "Lobophora" "cca" "Galaxaura"
"Sargassum" "otheralgae" "bleached" "unhealthy" "turf_over" "sediment"
"colsize_quadratio" "colsize_cm2" "porites_lumps" "porites_ridges"
"pigmentation" "trematodiasis" "annotationssession" "annotator" "gallcrabs"
"ascidians" "notes"];

coltypes =
["string","logical","logical","string","string","string","string","logical"
,
"logical","logical","logical","logical","logical","logical","string","logica
l",
"logical","logical","logical","logical","logical","logical","string","s
tring"
,"logical","logical","logical","double","double","logical","logical","logi
ca l",
"logical","string","string","logical","logical","string"];
varsize = size(colnames);

imgpath = uigetdir(); %% define folder of input images
cd(imgpath) %% move to that folder
names = dir('*CORAL*.jpg'); %% put the names of all the jpgs in the folder
into an object (careful when expanding to full dataset because there could
be variable file types?)
names2 = string({names.name}); %% extract a simple list of names from the
object, transpose it, and convert from character to string

annotationssession = datestr(clock,'yyyymmddHHMM');
annotator = string(inputdlg('What are your initials?'));
prompt = {'Do you want to take quantitative measurements?'};
doingquantitative =
```

SUPPLEMENTARY INFO: CODE S1

MATLAB code developed specifically for Tara Pacific Photographic annotation.



```

yesno(listdlg('PromptString',prompt,'ListSize',[300,300],'ListString',{'yes','no'}));

prompt = {'Are you adding to existing annotations?'};
if
yesno(listdlg('PromptString',prompt,'ListSize',[300,300],'ListString',{'yes','no'}))

[temp,existingannotations] = uigetfile('*','select existing annotation file');

existingannotations = [existingannotations,temp];
opts = detectImportOptions(existingannotations, 'Delimiter', '\t');
varNames = opts.VariableNames; % variable names
if length(varNames) ~= length(coltypes)

uiwait(msgbox('Old annotation fields do not match new ones. Output file will only contain new annotations', 'note', 'modal'));

oldresults = readtable(existingannotations); names_rand = string.empty;
for i = 1:length(names2)

if not(any(strcmpi(extractBefore(names2(i), '.'),
extractBefore(oldresults.filename, '.'))))

names_rand = [names_rand; names2(i)];

end end

names_rand = names_rand(randperm(length(names_rand))); %% randomize the order

results = table('Size', [0 varsize(2)], 'VariableTypes', coltypes,
'VariableNames', colnames);

resstartsize = 0;

else

varTypes = coltypes;
varTypes(strcmpi('logical', varTypes)) = 'double'; opts =
setvartype(opts,varNames,varTypes); oldresults =
readtable(existingannotations, opts); names_rand = string.empty;
for i = 1:length(names2)

if not(any(strcmpi(extractBefore(names2(i), '.'),
extractBefore(oldresults.filename, '.'))))

names_rand = [names_rand; names2(i)];

end end

names_rand = names_rand(randperm(length(names_rand))); %% randomize the order

results = oldresults;

resstartsize = length(results.filename);

```


SUPPLEMENTARY INFO: CODE S1

MATLAB code developed specifically for Tara Pacific Photographic annotation.



```

end

standardshots = unique([standardshots, rmissing(results.widevsclouse)],
'stable');

targets = unique([targets, rmissing(results.target)], 'stable');

poc_morphotypes = unique([poc_morphotypes,
rmissing(results(strcmp(results.target, 'Pocillopora'),:).morphotype)],
'stable');

por_morphotypes = unique([por_morphotypes,
rmissing(results(strcmp(results.target, 'Porites'),:).morphotype)],
'stable');

mil_morphotypes = unique([mil_morphotypes,
rmissing(results(strcmp(results.target, 'Millepora'),:).morphotype)],
'stable');

quadrats = unique([quadrats, rmissing(results.quadrat)], 'stable'); %%
there might be a fourth or even fifth quadrat to watch out for
else

results = table('Size', [0 varsize(2)], 'VariableTypes', coltypes,
'VariableNames', colnames);

names_rand = names2(randperm(length(names2))); %% randomize the order

resstartsize = 0;

end % if adding to preexisting annotations

h = msgbox([' ' char(int2str(length(names_rand))) ' photos to go!']);
uiwait(h)
%% start looping through photos. use the cancel button on any prompt to
either end the session or redo the active photo from the beginning
(technically any error will lead you to this choice)

i = resstartsize + 1;
while i < (resstartsize + length(names_rand))

try

results = [results; array2table(nan(1,varsize(2)), 'VariableNames',
colnames)]; %% append a new row to the results table that corresponds to
this photo

results.filename(i) = names_rand(i - resstartsize);
results.annotationsession(i) = annotationsession; results.annotator(i) =
annotator;

close all

currentimage = imread(char(results.filename(i))); imshow(currentimage)
set(gcf, 'units', 'normalized', 'outerposition', [0 0
set(gcf, 'units', 'normalized', 'innerposition', [0 0

%% 1)

```

SUPPLEMENTARY INFO: CODE S1

MATLAB code developed specifically for Tara Pacific Photographic annotation.



```

prompt = {'Is this a useful photo?'};

%% open the image

i 1}); i 1});

results.isntjunk(i) =
yesno(listdlg3('PromptString',prompt,'ListString',{'yes','no'},'position',[
60 0 0 0 0]));

if not(results.isntjunk(i))
prompt = {'Anything notable?'}; if

yesno(listdlg3('PromptString',prompt,'ListString',{'yes','no'},'position',[
60 0 0 0 0]))

results.notes(i) = inputdlg('Describe what's notable'); end

        close all
        i = i + 1;
        continue

end

%% 2)

prompt = {'Is there a target colony in the photo?'};

results.hastarget(i) =
yesno(listdlg3('PromptString',prompt,'ListString',{'yes','no'},'position',[
60 0 0 0 0]));

if not(results.hastarget(i)) prompt = {'Anything notable?'}; if

yesno(listdlg3('PromptString',prompt,'ListString',{'yes','no'},'position',[
60 0 0 0 0]))

results.notes(i) = inputdlg('Describe what's notable'); end

        close all
        i = i + 1;
        continue

end

%% 2.5)

prompt = {'Is this a \"standard\" wideangle or closeup shot?'}; % standard
meaning when we were doing just one of each, with quadrat in the wideangle
and showing the entire colony, and no scale but showing polyp details in
the closeup

results.widevsclose(i) =
standardshots(listdlg3('PromptString',prompt,'ListSize',[300,300],'ListStri
ng ','standardshots','position',[600 0 0 0]));

```


SUPPLEMENTARY INFO: CODE S1

MATLAB code developed specifically for Tara Pacific Photographic annotation.



```

if strcmpi(results.widevsclose(i), "other")
results.widevsclose(i) = inputdlg('Describe what''s nonstandard
about the photo');
standardshots = [standardshots, results.widevsclose(i)];

end

%% 2.5.5)

prompt = {'Which quadrat is in the photo?'};

whichQuad =
listdlg3('PromptString',prompt,'ListString',quadrats,'position',[600 0 0
0]);

results.quadrat(i) = quadrats(whichQuad); if strcmpi(results.quadrat(i),
"other")

results.quadrat(i) = string(inputdlg('Create a quadrat identifier'));

quadrats = [quadrats, results.quadrat(i)];

end

%% 5)

prompt = {'Are there parrotfish scars on the target?'};

results.parrotfish_scars(i) =
yesno(listdlg3('PromptString',prompt,'ListString',{'yes','no','unsure/maybe
'}, 'position',[600 0 0 0]));

prompt = {'Are there boring organisms growing in the target?'};
results.bivalve(i) = false();
results.Spirobranchus(i) = false();
results.Tridacna(i) = false();

results.boringurchin(i) = false(); results.other_polychaete(i) = false();
results.sponge(i) = false(); results.otherborers(i) = "none";

if

yesno(listdlg3('PromptString',prompt,'ListSize',[300,300],'ListString',{'ye
s','no'},'position',[600 0 0 0]))

command)');

prompt = {'Which borers? (you can select multiple using

currborers =
borers(listdlg3('PromptString',prompt,'ListSize',[300,300],'ListString',bor
ers,'position',[600 0 0 0]));

borers?');
end

```

SUPPLEMENTARY INFO: CODE S1

MATLAB code developed specifically for Tara Pacific Photographic annotation.



```

if any(strcmpi(currborers, "bivalve")) results.bivalve(i) = true();

end
if any(strcmpi(currborers, "Spirobranchus"))

results.Spirobranchus(i) = true();

end
if any(strcmpi(currborers, "Tridacna"))

results.Tridacna(i) = true();

end
if any(strcmpi(currborers, "urchin"))

results.boringurchin(i) = true();

end
if any(strcmpi(currborers, "other_polychaete"))

results.other_polychaete(i) = true();

end
if any(strcmpi(currborers, "sponge"))

results.sponge(i) = true();

end
if any(strcmpi(currborers, "other"))

results.otherborers(i) = inputdlg('What're the other end

prompt = {'Is the target in direct contact with algae?'};
results.Halimeda(i) = false();
results.Turbinaria(i) = false();
results.Dictyota(i) = false();

results.Lobophora(i) = false(); results.cca(i) = false();
results.Sargassum(i) = false(); results.Galaxaura(i) = false();
results.otheralgae(i) = "none"; if

yesno(listdlg3('PromptString',prompt,'ListSize',[300,300],'ListString',{'ye
s','no'},'position',[600 0 0 0]))

prompt = {'Which algae? (you can select multiple holding

command)'};

curralg =
algae(listdlg3('PromptString',prompt,'ListSize',[300,300],'ListString',alga
e,'position',[600 0 0 0]));

if any(strcmpi(curralg, "Halimeda")) results.Halimeda(i) = true();

end
if any(strcmpi(curralg, "Turbinaria"))

```

SUPPLEMENTARY INFO: CODE S1

MATLAB code developed specifically for Tara Pacific Photographic annotation.



```

results.Turbinaria(i) = true();

end
if any(strcmpi(curralg, "Dictyota"))

results.Dictyota(i) = true();

end
if any(strcmpi(curralg, "Lobophora"))

results.Lobophora(i) = true();

end

algae?');
    end

if any(strcmpi(curralg, "cca")) results.cca(i) = true();

end
if any(strcmpi(curralg, "Sargassum"))

results.Sargassum(i) = true();

end
if any(strcmpi(curralg, "Galaxaura"))

results.Galaxaura(i) = true();

end
if any(strcmpi(curralg, "other"))

results.otheralgae(i) = inputdlg('What're the other end

prompt = {'Is the target in contact with sediment?'};

results.sediment(i) =
yesno(listdlg3('PromptString',prompt,'ListSize',[300,300],'ListString',{'ye
s','no','unsure/maybe'},'position',[600 0 0 0]));

prompt = {'Does the target have patches of turf overgrowth?'};

results.turf_over(i) =
yesno(listdlg3('PromptString',prompt,'ListSize',[300,300],'ListString',{'ye
s','no','unsure/maybe'},'position',[600 0 0 0]));

prompt = {'Is the target bleached?'};

results.bleached(i) =
bleached(listdlg3('PromptString',prompt,'ListString',bleached,'position',[6
00 0 0 0]));

prompt = {'Does the target have a pigmentation response? (usually pink)'};

results.pigmentation(i) =
yesno(listdlg3('PromptString',prompt,'ListSize',[300,300],'ListString',{'ye
s','no','unsure/maybe'},'position',[600 0 0 0]));

```

SUPPLEMENTARY INFO: CODE S1

MATLAB code developed specifically for Tara Pacific Photographic annotation.



```

if not(isnan(results.pigmentation(i))) if results.pigmentation(i)

prompt = {'Is it trematodiasis'};

results.trematodiasis(i) =
yesno(listdlg3('PromptString',prompt,'ListSize',[300,300],'ListString',{'yes',
's','no','unsure/maybe'},'position',[600 0 0 0]));

end end

prompt = {'Is the target otherwise unhealthy or recently losing tissue?'};

results.unhealthy(i) =
yesno(listdlg3('PromptString',prompt,'ListSize',[300,300],'ListString',{'yes',
's','no','unsure/maybe'},'position',[600 0 0 0]));

%% 3)

prompt = {'Which target genus is the colony?'};

results.target(i) =
targets(listdlg3('PromptString',prompt,'ListString',targets,'position',[600
0 0 0]));

if strcmpi(results.target(i), "other")
results.target(i) = inputdlg('What other target is it?'); targets =
[targets, results.target(i)];

end

prompt = {'Anything else notable? (you can select multiple holding
command)'};

results.gallcrabs(i) = false(); results.ascidians(i) = false();
results.notes(i) = 'none'; notablenow =

othernotable(listdlg3('PromptString',prompt,'ListSize',[300,300],'ListStrin
g',othernotable,'position',[600 0 0 0]));

if not(strcmpi(notablenow, "no"))
if any(strcmpi(notablenow, "gallcrabs"))

results.gallcrabs(i) = true();

end
if any(strcmpi(notablenow, "ascidians"))

results.ascidians(i) = true();

end
if any(strcmpi(notablenow, "other"))

results.notes(i) = inputdlg('Describe what''s notable'); end

end
if strcmpi(results.target(i), "Pocillopora")

```


SUPPLEMENTARY INFO: CODE S1

MATLAB code developed specifically for Tara Pacific Photographic annotation.



```

%% 4)

prompt = {'Which morphotype is the colony?'};

results.morphotype(i) =
poc_morphotypes(listdlg3('PromptString',prompt,'ListString',poc_morphotypes
, 'position', [600 0 0 0]));

if strcmpi(results.morphotype(i), "other") results.morphotype(i) =
string(inputdlg('What other
morphotype is it?'));
poc_morphotypes = [poc_morphotypes, results.morphotype(i)];

end
if doingquantitative

%% 5)

title('click on the inside corners of the quadrat, clockwise from the
bottom left')

movingPoints = [ginput(1); ginput(1); ginput(1); ginput(1)];

%% identify the inner corners of the quadrat
fixedPoints = [0 1000; 1000 1000; 1000 0; 0 0]; %% not sure

why these are the proper coordinates; I'm guessing MatLab must have the
vertical axis with positive numbers downward?? Since the first coordinate
is positive on the vertical axis instead of zero?

tform = fitgeotrans(movingPoints, fixedPoints, 'Projective'); %% find
transformation to squish the points from the original photo into the
'fixedPoints' defined above.

entirely sure
transform the photo

RA = imref2d([1000 1000], [0 1000], [0 1000]); %% not [I,r] =
imwarp(currentimage, tform, 'OutputView', RA); %%

I = rgb2gray(I); %% go to greyscale. this discards info - would be nice to
use all color channels somehow

[I,temp] = histeq(I); %% standardize brightness and contrast. important for
consistent segmentation

imshow(I)

[L,N] = superpixels(I,100000); %% sort of a rough segmentation; finds
regions of low contrast and merges pixels to smooth things out. I actually
found this an important step to sort of standardize the inputs - since each
photo will have a different number of pixels inside the cropped quadrat,
the pixels will be at different realworld scales, which can introduce a
bias. this step normalizes each image so that the region inside the quadrat
always has the same number of 'pixels'.

```

SUPPLEMENTARY INFO: CODE S1

MATLAB code developed specifically for Tara Pacific Photographic annotation.



```

outputImage = zeros(size(I), 'like', I); idx = label2idx(L);
numRows = size(I,1);
numCols = size(I,2);

for labelVal = 1:N
redIdx = idx(labelVal); outputImage(redIdx) = mean(I(redIdx));
end

IS = outputImage;
imshow(IS)

gmag = imgradient(IS); %% next few sections from
https://www.mathworks.com/help/images/marker-controlled-watershed-segmentation.html

imshow(gmag,[])
se = strel('disk',20);

Ie = imerode(I,se);
Iobr = imreconstruct(Ie,IS); imshow(Iobr)

Iobrd = imdilate(Iobr,se);

Iobrcbr = imreconstruct(imcomplement(Iobrd),imcomplement(Iobr));

Iobrcbr = imcomplement(Iobrcbr); imshow(Iobrcbr)
title('Opening-Closing by Reconstruction')

fgm = imregionalmax(Iobrcbr); imshow(fgm)
I2 = labeloverlay(I,fgm); imshow(I2)

se2 = strel(ones(5,5)); fgm2 = imclose(fgm,se2); fgm3 = imerode(fgm2,se2);
fgm4 = bwareaopen(fgm3,20); I3 = labeloverlay(I,fgm4); imshow(I3)

T = adaptthresh(Iobrcbr,0.85,'NeighborhoodSize',
2*floor(size(Iobrcbr)/32)+1, 'Statistic', 'median');

bw = imbinarize(Iobrcbr, T); %% this is dependent on image scale! a cropped
image with fewer pixels will have more sensitivity relative to pixels; the
same sensitivity if the cropped image has the same realworld scale

%imshow(bw)

D = bwdist(bw);

DL = watershed(D);
bgm = DL == 0;
%imshow(bgm)
%title('Watershed Ridge Lines')

gmag2 = imimposemin(gmag, bgm | fgm4); L = watershed(gmag2);

%labels = imdilate(L==0,ones(3,3)) + 2*bgm + 3*fgm4; %I4 =
labeloverlay(I,labels);
%imshow(I4)
%title('Markers and Object Boundaries Superimposed on

```


SUPPLEMENTARY INFO: CODE S1

MATLAB code developed specifically for Tara Pacific Photographic annotation.



```

Original Image')

imshow(I)
set(gcf, 'units', 'normalized', 'outerposition', [0 0 1 1]); set(gcf,
'units', 'normalized', 'innerposition', [0 0 1 1]); hold on
Lrgb = label2rgb(L,'jet','w','shuffle');
himage = imshow(Lrgb);
himage.AlphaData = 0.2;
title('Left-click on all the segments that correspond to

target coral branches. Right click to stop; ESC to quit')

particproperties = regionprops(L); particname=[];
for j = 1:size(particproperties,1)

particcentroids = [particproperties(j).Centroid];

particname = [particname, L(round(particcentroids(2)),
round(particcentroids(1)))];

plot(round(particcentroids(1)), round(particcentroids(2)), 'ko')

text(round(particcentroids(1)), round(particcentroids(2)),
num2str(double(particname(j)))); %% label each

segmented region

end

totregion = zeros(1000,1000);
moreimage = imshow(label2rgb(ones(1000,1000),'jet',[0 0

moreimage.AlphaData = 0; datmat2 = uint8.empty; button = 1;
while button==1

[xi,yi,button] = ginput(1);
xi = round(xi);
yi = round(yi);
selected = L(round(yi),round(xi)); region = L==selected;

totregion = totregion+region; moreimage.AlphaData = 0.5 * totregion;

if(button == 3) %If right button is pressed finish

% h = msgbox('Stop selecting! '); % uiwait(h)

0], 'shuffle'));

digitising

Interrupted.....');

elseif(button == 27) %If ESC is pressed h = msgbox(' Stop! Digitalization

                uiwait(h)
                close
                clear
    
```

SUPPLEMENTARY INFO: CODE S1

MATLAB code developed specifically for Tara Pacific Photographic annotation.



```

        break;

    else % if left button is used plot(xi,yi,'b+')
        datmat2 = [datmat2; xi yi selected];

    end %if button3
end %while button1 %% select the relevant regions

totregion = totregion==0; %%binary mask; if a pixel belonged of the clicked
regions it will now belong to this superregion

results.colsize_quadratio(i) = sum(sum(totregion)) /

(1000*1000); %%
the number of pixels that belong to the mask. we define the number of
pixels for the image above as 1000*1000, so this quantity is the percentage
of pixels in the cropped photo that corresponds to the coral colony

if whichQuad <= quadrat_sizes_size(2) results.colsize_cm2(i) =
results.colsize_quadratio(i) *

quadrat_sizes(whichQuad); %% multiply the percent of the image
corresponding to coral by the real-world scale of the image to get the
real-world size of the colony. this is a container that accumulates the
results for each image as the loop goes on

end

close all
end % if doingquantitative

elseif strcmpi(results.target(i), "Porites") %% 4)

prompt = {'Which morphotype is the colony?'};

results.morphotype(i) =
por_morphotypes(listdlg3('PromptString',prompt,'ListString',por_morphotypes
,'position', [600 0 0 0]));

if strcmpi(results.morphotype(i), "other") results.morphotype(i) =
string(inputdlg('What other

morphotype is it?'));
por_morphotypes = [por_morphotypes, results.morphotype(i)];

end

prompt = {'Does the colony have round lumpy features?'};

results.porites_lumps(i) =
yesno(listdlg3('PromptString',prompt,'ListString',{'yes','no','unsure/maybe
'}, 'position', [600 0 0 0]));

prompt = {'Are there linear ridges on the surface of the

colony?'};

```

SUPPLEMENTARY INFO: CODE S1

MATLAB code developed specifically for Tara Pacific Photographic annotation.



```

results.porites_ridges(i) =
yesno(listdlg3('PromptString',prompt,'ListString',{'yes','no','unsure/maybe
'}, 'position', [600 0 0 0]));

elseif strcmpi(results.target(i), "Millepora") %% 4)

prompt = {'Which morphotype is the colony?'};

to at least one

since the mask is a binary array, the sum of all elements is

results.morphotype(i) =
mil_morphotypes(listdlg3('PromptString',prompt,'ListString',mil_morphotypes
,' position', [600 0 0 0]));

if strcmpi(results.morphotype(i), "other") results.morphotype(i) =
string(inputdlg('What other
morphotype is it?'));
mil_morphotypes = [mil_morphotypes, results.morphotype(i)];

end
end % if strcmpi(results.target(i), "Pocillopora")

i = i + 1;

catch

results(i,:) = [];
prompt = {'Do you want to end the session or redo this photo?'}; if
yesno(listdlg('PromptString', prompt, 'ListString', {'end
session', 'redo photo'})) break

end end % try

writetable(results, ['results_' annotationssession '.txt'], 'Delimiter',
'\t')

end % while i < (resstartsize + length(names_rand))
close all
writetable(results, ['results_' annotationssession '.txt'], 'Delimiter',
'\t')

if i == (resstartsize + length(names_rand))
h = msgbox(['Congratulations! You''re done! (I think)']); uiwait(h)

end

```

4 Article 3: Evaluation of Site Type Based on Benthic Composition, Hard Coral Diversity, and Coral Physiological Attributes (in prep)

4.1 Chapter 4 foreword

This chapter presents an article currently in preparation for publication. Here, I focused on the foundational data for the utilization of the AI machines developed and presented in Chapter 3 (Clampitt et al(a)., in prep) in an ecological context, *i.e.*, to determine the type/state of coral reefs. In this study the goal was to compare the frequency of the visual cues (Sediment Contact, Boring Organisms, Macroalgal Contact, CCA, Turf Overgrowth, Pigmentation, Bleaching, Tissue Appearance,) between four* types of sites (Pristine, Damaged, Restoration, and Polluted) to investigate if the visual cues differ between the sites and if they can be used as proxies for evaluating overall reef health of a particular site.

The sites of interest were located within the Birds Head Seascape in Indonesia with site visits occurring in May and October*. Sites were divided into four categories *a priori*:

1. **Damaged** sites or sites suffering from the effects of dynamite fishing
2. **Restoration** sites or sites damaged by dynamite fishing where restoration structures have been put in place
3. **Pristine** sites or sites lacking obvious anthropogenic influences that are typically representative of the bird's head seascape
4. **Polluted*** sites or sites with known anthropogenic waste or where multiple examples of trash were present among the substrate

Line Intercept Transect (LIT) benthic surveys were conducted to identify and calculate ecological information about the site relative to the flora, fauna, and abiotic factors (*i.e.*, substrate) present enabling us to calculate coral diversity (to the genus level), and additional

* The current state of the article does not include all the results. Site visits occurred in October 2022 and May 2023. Site visits differed in that additional sites and site types were added during the May 2023 visit which included polluted site types and replicates of certain sites at different depths which were not included in the October analysis. Analysis for data collected in May is currently underway and we plan to include this information for the final paper but will not be fully presented in this chapter. Preliminary results can be found in the annex immediately following this chapter. Here, I will focus mainly on data from the October 2022 visit which includes three types of sites: Pristine, Damaged, and Restoration.

benthic coverage metrics such as the percentage and composition of algae, substrate, and fauna. Data from the LIT surveys was used to perform a hierarchical clustering analysis to designate types of site in an unbiased manner. Multi-factorial analyses were then used to tease out benthic categories which may be drivers of site type designations. Furthermore, hard coral biodiversity indexes were calculated which provided strong support for site classifications. Overall, these analyses indicate that coral diversity may be a more relevant predictor of site type than coral coverage with replicates falling into specific biodiversity ranges that correspond nicely to site types. These results suggest that simplifying methodology to only collect benthic information relevant to hard corals, specifically hard coral diversity may be sufficient to determine site type.

Additionally, in October individual coral colonies of three genera (*Pocillopora*, *Porites*, and *Acropora*) were photographed and surveyed following the methodology explained in Chapter 3 (Clampitt et al(a), in prep). In May, novel methodology was utilized to incorporate all 9 visual cues into the *in-situ* survey, surveyors noted the absence or presence of each of the 9 attributes as well as the percentage of living tissue. Each surveyed colony was photographed to show the colony as well as macro details of each attribute, resulting in 20 photos per colony.

Correlations between visual cues and the four* different types of sites were analyzed to determine if certain visual cues were more or less prevalent in specific types of sites. Results showed no relationship between visual cues and type of site. However, these results appeared to be heavily influenced by the photographic monitoring protocol, both of which were aimed towards optimizing data collection for the improvement of our AI machines.

Finally, the new images and attribute datasets were used to further test the AI machines developed in chapter 3 (Clampitt et al, in prep). F1 scores improved after the May site visit indicating that the novel methodology is more effective. When comparing machine performance on colony level vs macro level pictures, machines performed slightly better on the colony level photographs indicating that the macro level photographs were more different to the original training photographs than the colony level ones. Using the data generated from this study, we aimed to better understand and quantify the differences between the four* different types of sites as well as gain insight into how the visual cues may differ between sites.

* Data presented in this chapter focuses on three types of sites: pristine, restoration, and damaged

4.2 Evaluation of Site Type based on Benthic Composition and Hard Coral Diversity

Megan Clampitt^{1-3,5*}, Florentin Remot^{1-3*}, Marco Milanese⁵, Paul Boli⁶, Ridwan Sala⁶, Virly Yuriken⁷, Ayub Markus⁷, Eric Röttinger^{1-3@}, Eric Gilson^{1-4@}

¹ Université Côte d'Azur-CNRS- Inserm - Institute for Research on Cancer and Aging, Nice (IRCAN), France.

² Université Côte d'Azur, Institut Fédératif de Recherche - Ressources Marines (IFR MARRES)

³ Laboratoire International Associé Université Côte d'Azur - Centre Scientifique de Monaco (LIA ROPSE), France, Monaco.

⁴ Department of Medical Genetics, CHU, Nice, France.

⁵ Université Côte d'Azur, Center of Modeling, Simulation, & Interaction, France

⁶ University of Papua (UNIPA), Manokwari, Indonesia

⁷ Misool Foundation, Sorong, Indonesia

*These authors contributed equally to this work

@Corresponding authors

Statements and Declarations:

The authors have no competing interests to declare that are relevant to the content of this article.

Acknowledgements:

This work was supported by ANR-15-IDEX-01 via the IDEX UCA^{jedi} (MSI), the Ecole Doctorale Sciences de la Vie et de la Santé de Université Côte d'Azur, the LIA-ROPSE, The Université Côte d'Azur, la Région SUD, CNRS INSB. The authors are grateful to the Misool Foundation with specific thanks to the Misool ranger patrol for their hospitality and warmth.

Abstract

Coral cover is a commonly used metric for defining the current state of different reef ecosystems. Typically, reefs are measured on a quartile scale with 0-25% coral coverage classified as poor and 76-100% coral coverage classified as very good or excellent. With worldwide reef degradation and shifting baselines, evaluation of coral reef ecosystems may need to be rethought with increasing demand for a simple, rapid, and reproducible methodology. Here we evaluated the benthic coverage and coral diversity of four* types of sites: Pristine, Restoration, Damaged, and Polluted while also investigating physiological hallmarks present in individual colonies through in water annotations and photographs. Furthermore, photographs were run through a newly developed AI pipeline that automatically annotates physiological hallmarks from coral colony images to assess the efficiency of these machines on a novel dataset obtained using an updated photographic methodology. Results indicate that biodiversity is a more relevant indicator of site type than coral coverage with the biodiversity indexes falling into ranges which corresponded to the three* types of sites. Visual cues / physiological attributes were not found to be predictors of site type, but methodology changes are required to further investigate this trend. Finally, machine performance was adequate when classifying novel photographs with a higher average F1 score for classifying presence when using colony photos as compared to macro photos and higher scores were achieved with updated methodology. Overall, this study presents a link between biodiversity and site determination and is the foundational study for the development of AI machines capable of categorizing colony level photographs into site designations.

Key Words: (4-6)

Coral, Benthic Surveys, Biodiversity, Site differentiation

* Data presented in this chapter focuses on three types of sites: pristine, restoration, and damaged

Introduction:

The coral triangle is the epicenter of global marine biodiversity for major taxa such as corals, fish, and crustaceans and showcases the remarkable richness found in coral reef ecosystems (Bruno & Selig, 2007). Indonesian reefs situated in the heart of the coral triangle, harbor the highest biodiversity of coral species, with nearly 70% of all known hard coral species found in this region (Hadi et al., 2020). Corals, recognized as foundational species, play a vital role in facilitating the establishment of other associated species, indirectly supporting their survival and continuity by providing refuge from environmental pressures and predators, enhancing propagule availability and retention, and augmenting resource availability (Bruno & Bertness, 2001). As an example, mortality of hard corals has led to reduction and even local extinction of certain fish species who rely heavily on live coral throughout their life for shelter, diet and recruitment purposes (Graham et al., 2006). Furthermore, a positive correlation exists between biodiversity and live coral cover as sites with higher coral cover also demonstrate greater diversity (Edinger et al., 1998). It has even been found that MPA designations were more effective in terms of increased habitat quality when coral cover was considered (Vercammen et al., 2018). The significance of living coral cover as a key measure of reef habitat quality and quantity is highlighted by its widespread use as a metric to evaluate reef condition (Bruno & Selig, 2003) (Vercammen et al., 2018) (Permana et al., 2020).

Globally, an estimated 25% of reefs have already been destroyed or severely damaged as a result of climate change and 58% of coral reefs are threatened by human activities like overfishing, deforestation, agricultural practices or terrestrial development which can disturb the balance of these ecosystems and introduce sediment, nutrients and pollutants into coastal environments (Roberts et al., 2002). A study conducted in 2003 revealed that coral cover on Indo-Pacific reefs was lower and more uniform than previously estimated, with a region-wide average of 22.1% suggesting that the decline of corals in the Indo-Pacific has been underestimated (Bruno & Selig, 2003). Moreover, by considering the average coral cover of Indo-Pacific reefs in the 1980's, which was around 42.5%, researchers were able to estimate the average annual loss over a 20-year period which amounted to approximately 1% or 1,500km², a figure similar to the losses of 1.5% observed in the Caribbean during this same time period (Bruno & Selig, 2003). Currently, less than 7% of Indonesia's coral reefs are classified as excellent (76-100%) in terms of coral cover, indicating the urgent need for conservation efforts (Permana et al., 2020).

Coral reefs hold immense ecological, economic, and cultural importance for the people of Indonesia. Indonesia stands as the fourth most populated country, with 80% of its habitants residing along the coast thus implying a reliance on coastal resources (Habibi & Sartin, 2007) (Hadi et al., 2020). But, despite their value and richness, Indonesian reefs remain threatened. Roberts et al. (2002) found that regions characterized by high species richness face a greater degree of vulnerability to human impacts compared to less diverse areas. Certain sites in Indonesia have seen a 25% decrease in coral generic diversity over the past 15 years (Edinger et al., 1998) most likely resulting from land-based pollution. Compounded stressors such as environmentally damaging fishing practices, overfishing, sedimentation, waste disposal, and coral bleaching pose serious risks to these valuable ecosystems. Overfishing and destructive fishing practices alone negatively impact over 80% of Indonesia's coral reefs (Burke et al., 2012).

Indonesia and other areas of Asia have long suffered from the devastating effects of dynamite fishing which can leave reefs in ruin and double or even triple the time of recovery as compared to a natural disturbance (Fox & Caldwell, 2006). As foundational species, corals often produce large three-dimensional structures which act as a refuge from predators while also reducing

flow velocity thereby stabilizing the substrate and enhancing settlement (Bruno & Bertness, 2001). One of the main detrimental effects of dynamite fishing is the destruction of corals, transforming the once stable, complex 3-dimensional structure of the benthos into an unstable, flat and shifting substratum known as rubble. To address these destructive effects, coral restoration techniques are often employed. One such method is substrate stabilization, wherein secure substrates (such as grates or metal structures) are put onto the reef and coral colonies are attached to them by divers. In such a way, restoration practitioners create stable substrate for coral colony growth and settlement while also increasing abundance and species richness of reef fish, increasing coral cover, and favoring coral recruitment with the added benefit of raising awareness of reef conservation issues (Fadli et al., 2012) (Williams et al., 2019).

Apart from coral cover, we can glean insight into the current state of reefs through the investigation of visual attributes relating to individual coral colonies. In Clampitt et al., in prep we denote 9 physiological hallmarks (bleaching, predation, boring organisms, macroalgal contact, pigmentation, sedimentation, turf overgrowth, CCA, and tissue appearance) that can give us information relevant to the current physiological state of individual colonies. These physiological hallmarks and their potential use as proxies for determining reef health are explained in more detail in Clampitt et al(a)., in prep.

Without detailing all hallmarks but to illustrate this point, predation, for example, can require energetic demands from the coral while also providing information about herbivory which can be linked back to the overall state of the reef (McField et al., 2007). The effects of predation on an individual colony can be detrimental or minor in terms of mortality while also affecting growth and/or fitness (Rotjan & Lewis, 2008). Another example is algal growth. In fact, due to limited benthic space, the ratio between corals and algae can be a primary indicator of reef state (Mcfield et al., 2007). When corals die, their skeletal structure is often invaded by algae which in incidences of mass coral mortality can lead to a phase-shift from a coral dominated reef to an algal dominated reef and thus overall degradation in the state of the reef (Done, 1998). Another hallmark, sedimentation, can also be used for individual colony as well as reef wide assessment. In terms of an individual colony, sedimentation can interfere with photosynthesis and active feeding while also requiring an energetic expense by the colony for removal (McField et al., 2007). As for the reef in a global sense, high levels of sedimentation can reduce growth rates, species richness and zonation patterns and can be indicative to the presence of stressors or threats (McField et al., 2007). For example, sedimentation can directly affect coral coverage with reefs experiencing high sedimentation showing little or no *Acropora* cover (Edinger et al., 1998). Boring organisms are bioeroders to coral colonies which affect the competing processes of reef growth and erosion and thus the long-term sustainability of reefs (Mcfield et al., 2007). Rice et al., (2020) found that increased bioerosion was correlated to increased eutrophication. Taken together, visual cues of individual coral colonies as presented in Clampitt et al., in prep could provide information not only on the physiological stage of the colony, but also about the overall state of the reef.

In this study we surveyed coral reef sites experiencing differing anthropogenic influences in two different locations within the coral triangle in Indonesia. Site selection was pre categorized into the following four* categories: (1) Damaged – defined as sites suffering from the aftereffects of dynamite fishing, characterized by large rubble fields (2) Restoration – defined as sites currently undergoing active restoration activities (3) Pristine - or more remote sites

* Data presented in this chapter focuses on three types of sites: pristine, restoration, and damaged

lacking obvious anthropogenic influences. and (4) Polluted – as identified by the presence of trash or anthropogenic waste Using benthic survey data, we used a non-biased approach to differentiate the sites to gain a better understanding of what elements may predict the type of site and how coral cover and diversity plays a role in evaluating the state of a reef. Colonies at each site were also photographed in order to evaluate the nine visual cues described in Clappitt et al., in prep. Finally, by combining ecological information with colony level physiological hallmarks, we aimed at deepening our understanding of how visual assessments of coral colonies may be affected by ecological factors.

Materials & Methods:

a priori site determination: Sites were chosen from the Birds Head Seascape in Indonesia with sites selected to represent differing anthropogenic pressures. Sites were divided into four* categories: pristine, restoration, damaged, and polluted (Figure 1A-C) and sampled between at either 5 or 10m depth. Pristine sites were defined as sites with little to no anthropogenic influence (Figure 1A). Restoration sites were defined as active sites of coral restoration; with corals being on or among coral restoration structures (Figure 1B). Damaged sites were defined as sites showing high amounts of rubble resulting from dynamite fishing (Figure 1C). Polluted sites were defined as sites with known anthropogenic waste or where multiple examples of trash were present among the substrate (not shown). Two sites were also selected which didn't fit into one of the above distinctions and were unclassified. Site visits occurred in October 2022 and May 2023*.

Benthic Survey (Coral Diversity): At each site, LIT Benthic Surveys were conducted following English et al. (1997) Survey Manual for Tropical Marine Resources with appropriate updates to account for current advances in coral taxonomy (SI – Datasheet 1). Changes in the benthos were recorded for two replicate transects of 20m per site with depths ranging from 2-10m. Hard coral cover and diversity (to the genus level), other fauna (soft corals, sponges...etc.), algae, and substrate cover were the main categories investigated. Percent cover was calculated as
$$\left(\frac{\text{Total length of benthic category}}{20} \right) \times 100.$$

Photographic survey and underwater annotations of coral colonies: Additionally, photos and in situ surveys of individual colonies were taken targeting three genera: *Pocillopora sp.*, *Porites sp.*, and *Acropora sp.* Colonies were selected that were in the same area and depth range as the LIT survey. For the October 2022 survey, photographic methodology and *in-situ* surveys followed the procedure described in Clappitt et al., in prep. Three photographs were taken per colony with in-water annotations including percentage of living tissue, bleaching on a six-point scale using the Coral Watch Health card (Siebek et al., 2006) and absence and presence of predation and disease. For the May 2023 survey an updated photographic methodology and in situ annotation procedure was performed to include all the major in water annotation categories: Pigmentation Response, Bleaching, Tissue Appearance, Sediment Contact, Macroalgal contact, CCA, Turf Overgrowth, Boring Organisms, and Predation) (Clappitt et al(a)., in prep) (SI-Datasheet 2). To reflect these changes, 20 photographs were taken for each target colony one wide angle photo and one standard photo as described in Clappitt et al(a)., in prep, and 18 macro level photos with 2 photos for each of the nine annotation categories listed above (Figure 7). The first of the two photos for each attribute was taken to show absence *i.e.*, an area of the colony that did not display the annotation category (for example, a section of tissue displaying

* In this chapter, I will focus mainly on the October 2022 data, May 2023 data can be found in the annex

no predation marks for the predation category)(Figure 7). The second photo was taken to show the presence of the attribute (*e.g.*, a section of coral displaying predation for the attribute predation)(Figure 7). If a colony did not display the attribute, 2 “absence” photos were taken so that each colony had 2 photos per attribute, always totaling 20 photos per colony.

Immediately following sampling, photos were reviewed in Adobe Lightroom CC, to eliminate superfluous photos and ensure the correct number of photos were taken for each colony. If extra photos were taken, the diver selected the photo in terms of the best quality (not blurry, centered, clearly showing the attribute...*etc.*). Photo filenames were then matched to the *in-situ* annotation.

Statistical analyses:

Multiple factor analysis (MFA): To evaluate the composition and the ecological structure of our different sites, a Multiple Factor Analysis (MFA) was conducted using count and length data generated from the LIT survey. All statistical analyses were performed using the R software version 4.1.2 (R Core Team, 2020). Multivariate analyses were performed using the package FactoMineR version 2.8 (Husson et al., 2023) and data visualization was performed using the package factoextra version 1.0.7 (Kassambara & Mundt, 2020)

An MFA, in short, is a multivariate data analysis method for summarizing and visualizing a complex data table in which individuals are described by several sets of variables (quantitative and/or qualitative) and structured into groups. To define the distance between individuals, it considers the contribution of all active groups of variables. The number of variables in each group may differ, as may the nature of the variables (qualitative or quantitative), but all variables in each group must be of the same nature. We therefore split the LIT survey data into 4 categories, hard coral (41 genera of hard coral), fauna (5 variables), algae (3 variables) and substrates (8 variables)*. Raw data of all count and length was used in the analysis and the *a priori* designations of sites (pristine, restoration, damaged and polluted) was used as a supplementary variable. We decided to keep 5 dimensions in our analysis which explained 86% of the total variance for both datasets*. Results of the MFA were displayed on correlation plots between quantitative variables and the first two dimensions as well as the contributions of groups and of variables to the first two dimensions. Sites were also displayed on an individual factor map and were colored and grouped according to their *a priori* designations.

Following the MFA, to assess whether the *a priori* sites designations were biologically relevant and whether unclassified sites were more similar to pristine or damaged sites, we performed a hierarchical clustering on principal components analysis (HCPC) using the function *hcpc* from the R package FactoMineR. The HCPC was performed using the Ward algorithm and the optimal number of clusters was calculated by the function (using the argument *nb.clust* = -1). Using count data, the function found 3 clusters, therefore we set the number of clusters to three for the length dataset. Results of the HCPC were displayed on dendrograms using the function *fviz_dend* from the factoextra package.

Clustering using the hard coral composition: Due to different clusters when utilizing length vs count data from all benthic categories and to determine if hard coral composition alone was enough to differentiate between sites, clustering was computed using data from hard corals alone. Using the data of hard coral (count and length) only, we calculated the Bray-Curtis distance matrix using the *vegdist* function from the vegan package version 2.6-4 (Oksanen et

* This is specific to October 2022 data

al., 2022). Then we performed a cluster analysis using Ward's algorithm on the square-root transformation of the Bray-Curtis distance matrix (since Ward's algorithm needs metric distance) using the function *agnes* from the package *cluster* version 2.1.2 (Maechler et al., 2022). Following previous results from the HCPC, we decided to fix the number of clusters to three using the function *cutree* (argument $k = 3$)*. The same clustering analysis was performed on the length data of hard coral.

Diversity indexes calculation: To analyze the ecological diversity and to compare the diversity of the various clusters we defined before (*a priori*, hcpc and dissimilarity matrix) we calculated 3 diversity indexes for each clustering methods. Using the function *diversity* (package *vegan* (Oksanen et al., 2022)), we calculated the genera richness (i.e., the number of genera), the Shannon index calculated as $H = -\sum_{i=1}^S p_i \log \log p_i$ with S the genera richness and p_i the relative abundance of genera i and the Gini-Simpson index calculated as $GD = 1 - \sum_{i=1}^S p_i^2$. We calculated both Shannon and Simpson indexes because they can provide different results since Shannon puts more weight in richness whereas Simpson puts more weight on evenness (i.e., the genera abundance distribution of a community). Diversity indexes was calculated in two ways: the total diversity index for each cluster (similar to gamma diversity) and the mean of diversity index within clusters (similar to alpha diversity).

Photographic annotation analyses: To analyze photographic annotations, we performed a Multiple Correspondence Analysis (MCA), using the package *FactoMineR* (Husson et al., 2023). The MCA is a multivariate analysis method for summarizing and visualizing a data table containing only qualitative variables. In short, we performed an MCA to assess whether photographic annotations (percent of living tissue, bleaching on a six-point scale, presence/absence of predation and disease) were related to site type. First, this analysis was performed on all 891** colonies regardless of genus (434 *Acropora*, 237 *Pocillopora*, and 220 *Porites*) then we ran the same analysis for each genus separately.

Results:

***A priori* site designations proved to be robust based on benthic data**

In October 2022, 11 different sites were sampled and determined *a priori* as 3 Pristine sites, 3 Restoration sites, 3 Damaged sites, & 2 Unclassified sites (Figure 1A-F). Two replicate LIT surveys of 20m were conducted per site on adjacent stretches of reef, totaling 22 surveys and equating to 440m of surveyed reef. Figure 1G shows an overview of benthic composition by site and by replicate detailing coverage of hard corals, fauna, algae, and substrate. Focusing on hard corals, we can see that coral coverage varied by site ranging from 3.3% to 37% with an average of 19% across all 22 sites (Figure 1G). This corresponds to the regional average found in the 2007 study by Bruno & Selig and a 2012 study suggesting that the majority of Indo-pacific reefs had average coverage from 26-50% with this number decreasing (Burke et al., 2012). Based on the *a priori* site designations, pristine sites averaged 23.2% coral cover, restoration sites averaged 24.6% and damaged sites averaged 9.72%.

* This is specific to October 2022 data

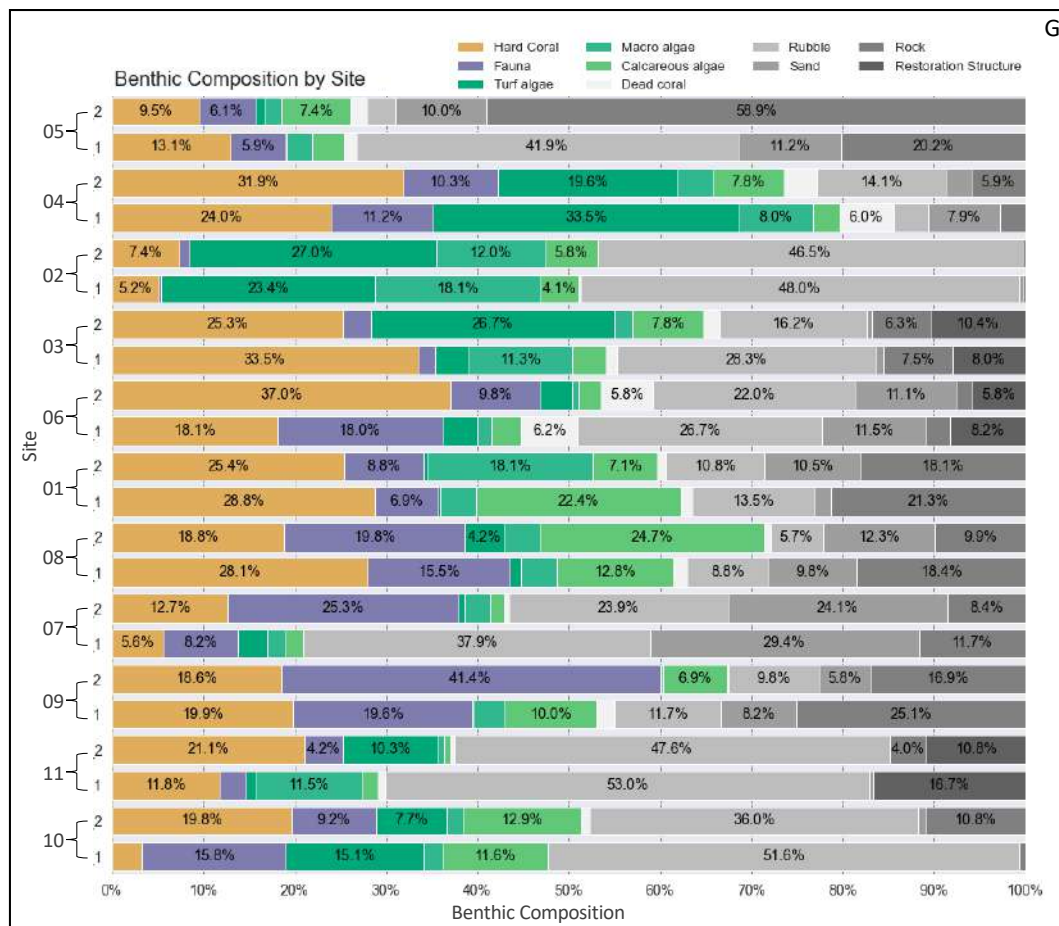
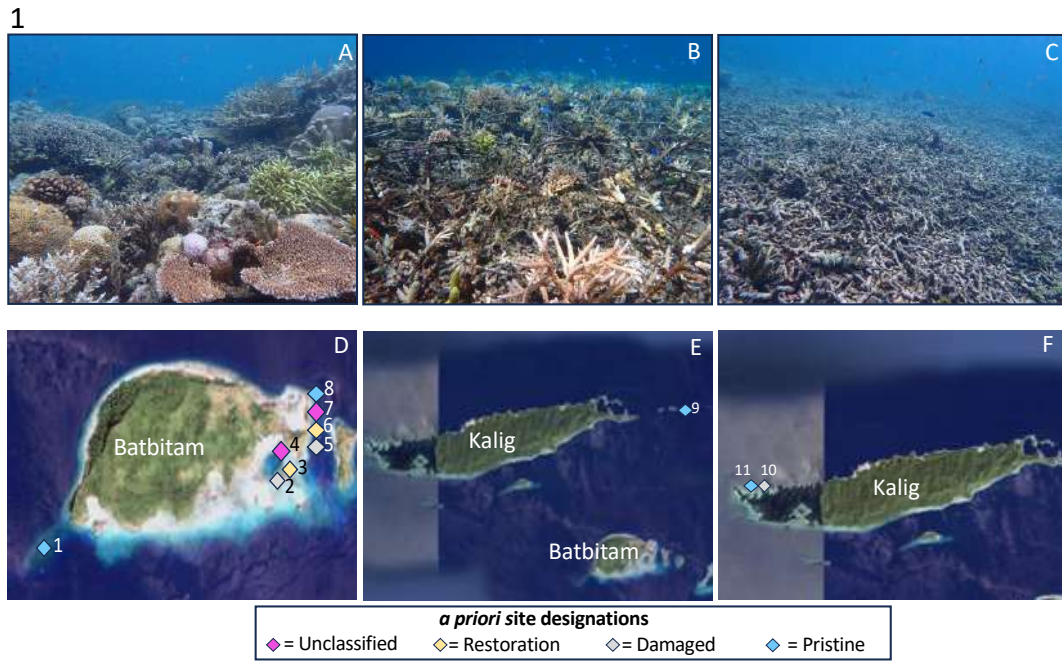


Figure 1: (A) Example topography from a pristine site (B) example topography from a restoration site (C) Example topography from a damaged site (D) Map showing Batbitam island with the eight sites surrounding this island (E) Maps showing the two main islands (Kalig & Batbitam) where sites are located as well as the location of site 9 (F) Map showing Kalig island with the two sites selected from this island. A priori site designations are denoted by different colored diamonds with pink being unclassified, yellow being restoration, gray being damaged and blue being pristine. (G) Overview of benthic composition by site showing hard corals in yellow, fauna in purple, algae in shades of green and substrate in shades of grey. Percentages are displayed unless totalling less than 4.0% for legibility purposes.

To verify our *a priori* site designations, LIT Survey data was utilized to determine site types in an unbiased manner via Hierarchical Clustering on Principal Components (HCPC) the Ward algorithm to denote three clusters. Data was divided into count data *i.e.*, the number of times a benthic category appeared in the 20m survey and length data *i.e.*, the distance a benthic category covered across the total surveyed length of 20m. Hierarchical clustering was performed for both length and count data using all benthic categories and using hard coral only data. Figure 2 shows the different dendrograms created (Figure 2A-D) with the four different analytical approaches and a consensus dendrogram (Figure 2E) created based on majority cluster membership across the four dendrograms. Table S1 shows cluster membership by site for each analytical approach. Importantly, three major clusters were determined that correspond to the *a priori site designations*. Cluster 1 (gray) corresponds to damaged sites, cluster two (yellow) corresponds to restoration sites and cluster 3 (blue) corresponds to pristine sites. The designation of clusters as pristine, restoration or damaged was determined by the majority *a priori* site type in a cluster. For example, if a cluster was made up of nine replicates total and 7 were pristine, 1 was unclassified and 1 was damaged – this cluster would be designated pristine.

Cluster membership proved stable with sites tending to group together and appear in the same clusters across the four analytical approaches. 10 out of the 22 replicates were always grouped together (01-R1, 01-R2, 08-R1, 08-R2, 09-R1, 03-R1, 06-R2, 11-02, 11-01,10-R1), appearing in the same cluster across the four analytical approaches. 11 were grouped together 3 out of 4 times (09-R2, 05-R1, 05-R2, 07-R1, 07-R2, 06-R1, 10-R2, 02-R1, 02-R2, 04-R1, 04-R2) and only 1 replicate (03-R2) was split between two clusters, appearing in cluster 1 twice and cluster 2 twice. Apart from site 03 and site 10, replicates by site always appeared in the same clusters with the *a priori site* classifications of these sites being restoration and damaged, respectively. With the new consensus dendrogram they are Restoration (03-R1) & Restoration and/or Damaged (03-R2), and Restoration (10-R2), and Damaged(10-R1). The two unclassified sites (04 and 07) are classified as damaged and pristine, respectively with the new consensus diagram. Apart from the unclassified sites, the only site that changed category from the *a priori* classifications to the *a posteriori* consensus classification was site (05) which changed from damaged to pristine (Table S1). Interestingly, if average coral coverage is calculated based on the consensus groupings pristine sites average 18.05%, restoration sites average 23.8% and damaged sites average 14.36%. To calculate these values, site 03-R2 was included in restoration sites because it has active coral restoration activities happening currently.

Overall, our *a priori* site designations corresponded to the count and length data extracted from all the benthic categories as well as the hard coral only categories. By assessing the different groupings, we were able to create a consensus tree which summarizes the cluster that best applies to each site replicate (Figure 2E).

Differentiation between restoration and pristine sites is robust and highly distinct across analytical methods with differences between unclassified and damaged sites more dependent on analytical approach.

All Benthos - Count Data: Clusters generated using all benthic count data correspond exactly to the *a priori* classification with the unclassified sites grouped with the damaged sites (Figure 2A, 3A). Thus, the count data proves our *a priori* site classifications are biologically relevant. Furthermore, using the benthic categories 86% of variation can be explained by five axes (Figure 3B, 4B) with hard coral and substrate being the main contributors for the first dimension with *Acropora sp* contributing the most followed by soft coral, crustose coralline algae, and then substrate type: restoration structure, rubble, and rock, respectively (Figure 3C,

2

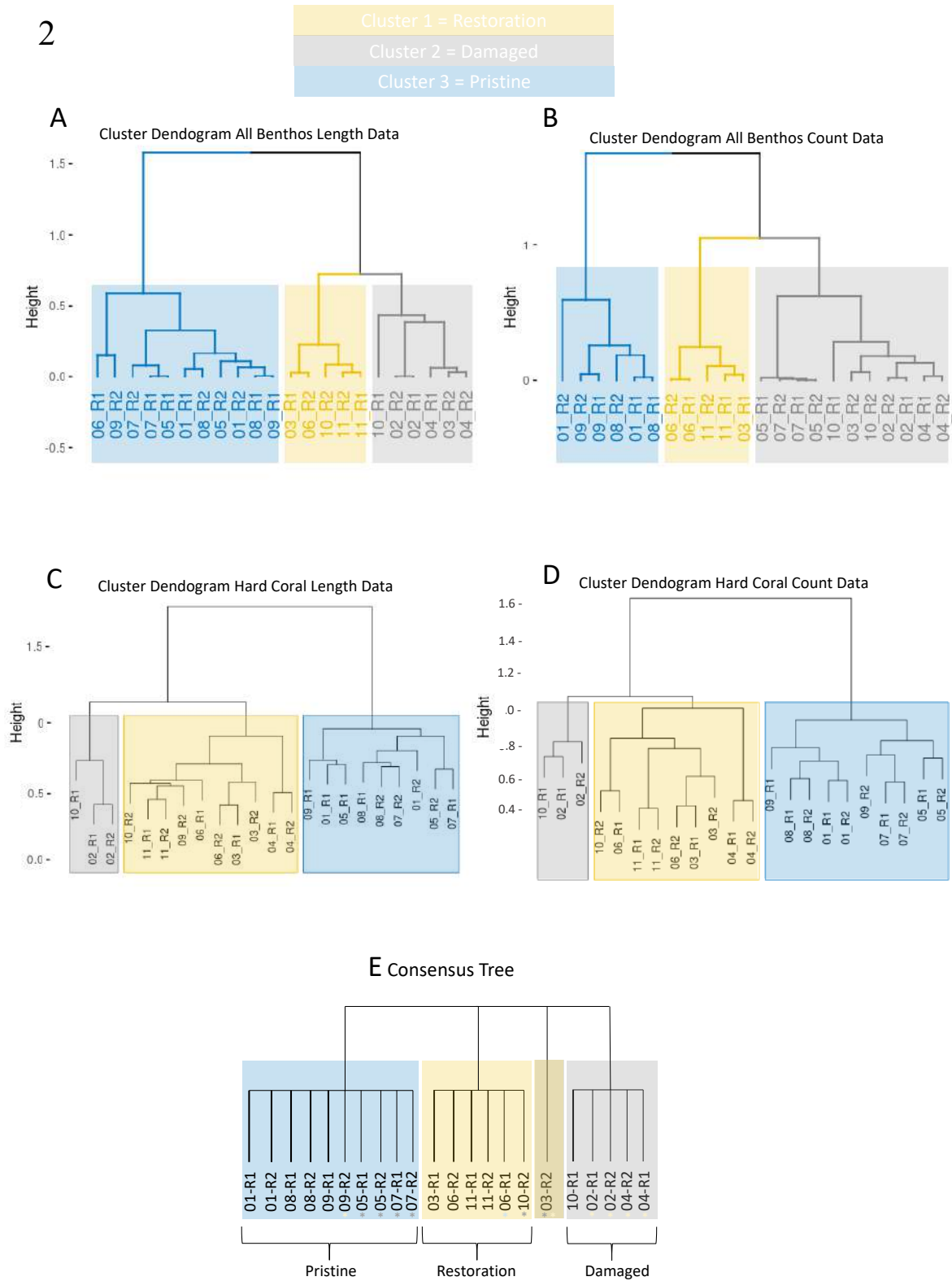


Figure 2: Dendrograms for clusters 1, 2 and 3 based on different grouping methods (A) All benthic categories length data (B) All benthic categories count data (C) Hard coral only length data (D) Hard coral only count data (E) consensus dendrogram based on majority groupings of above four dendrograms. Thicker lines represent sites that appeared in the same cluster across all four groupings, thin lines represent sites that appeared in 3/4 groupings and the thinnest line (Site 03-R2) appeared twice in one grouping and twice in another and is thus represented as its own category as it could belong to either cluster 1 (yellow) or cluster 2 (gray). The asterisks represent the other cluster that the group could be found in (see chart S1 for more details) Yellow represents cluster 1, Gray represents cluster 2 and Blue represents cluster 3 across all dendrograms.

3D). Algal composition was the most important contributor for the second dimension with macroalgae accounting for over 60% of the variance and CCA contributing the second most (Figure 3C, 3E). When plotting the *a priori* site designations on the first two axes, we can see a clear distinction between all three sites (restoration, damaged, and pristine) but specifically between restoration sites and pristine sites with damaged sites differentiated but not as markedly (Figure 3A). When looking at what factors were the most relevant for site determination, we can see that rock, CCA, soft coral, and several hard coral genera, most notably *Porites*, were the most relevant for determining pristine sites (Figure 3A, 3C). In terms of restoration sites, substrate played an important role with rubble, restoration structures and dead coral covered in algae dominating and *Acropora* proving important in terms of hard coral genera. Damaged (& unclassified) sites seemed to have less specific benthic categories used for classification with various hard coral species overlapping (Figure 3A, 3C).

All Benthos – Length Data: Here, the main contributors to the first dimension are substrate and algae (Figure 4D) with turf algae and soft coral accounting for approximately 20% each followed by *Acropora sp.*, rubble, and rock (Figure 4C, 4D,). The second dimension can be explained mostly by hard coral and fauna with *Acropora* explaining almost 45% of the variance followed by soft coral, rubble, CCA and other (Figure 4C, 4E). Here, the hierarchical clustering differs slightly from our *a priori* classifications (Table S1). The pristine sites all fall within the same cluster, both replicates of one unclassified site and one damaged site as well as one replicate from a restoration site are mixed in. The restoration sites are grouped together along with one replicate from a damaged site. The damaged grouping has three replicates from damaged sites, one unclassified site and one replicate from a restoration site.

When looking at the MFA showing the benthic categories and the sites along the first two dimensions (Figure 4A, 4C), we can see that rock and sand, in terms of substrate can be used to define pristine site as well as *Porites* in terms of hard coral genera. Restoration sites were again defined by restoration structures with *Acropora* being one of the most important coral genera. In terms of length data, rubble seems to be the most important determinant for damaged sites.

This can be visualized by looking at the MFA plots showing site groupings (Figure 3A, 4A). In both groupings (length and count) we can see that the pristine and restoration sites are always highly differentiated, specifically on the first dimension which corresponds to *Acropora*, soft coral, CCA, restoration structures, and rubble for the count data and turf algae, soft coral, *Acropora* and rubble for the length data. The main difference between the two groupings occurs within the damaged sites with overlap occurring between damaged and restoration sites as mentioned above, when using the length data. When looking at what benthic categories are most relevant for determining site, *Acropora* and *Porites* seem to be the most relevant of the different genera of hard coral with *Porites* helping to define a site as pristine and *Acropora* for defining a site as restoration type. Substrate type seemed to be important as well with rock defining pristine sites, restoration structures defining restoration sites and rubble defining damaged sites.

Overall, both MFA analyses displaying sites across the first two dimensions show a clear separation between pristine and restoration and pristine and damaged sites. Thus, regardless of analytical method, and based on benthic data pristine sites can easily be differentiated from restoration and damaged sites. The count data is more in accordance with *a priori* site designations but that does not necessarily indicate that these groupings are correct. Thus, further analysis was needed.

3

Count data for all Benthic Categories

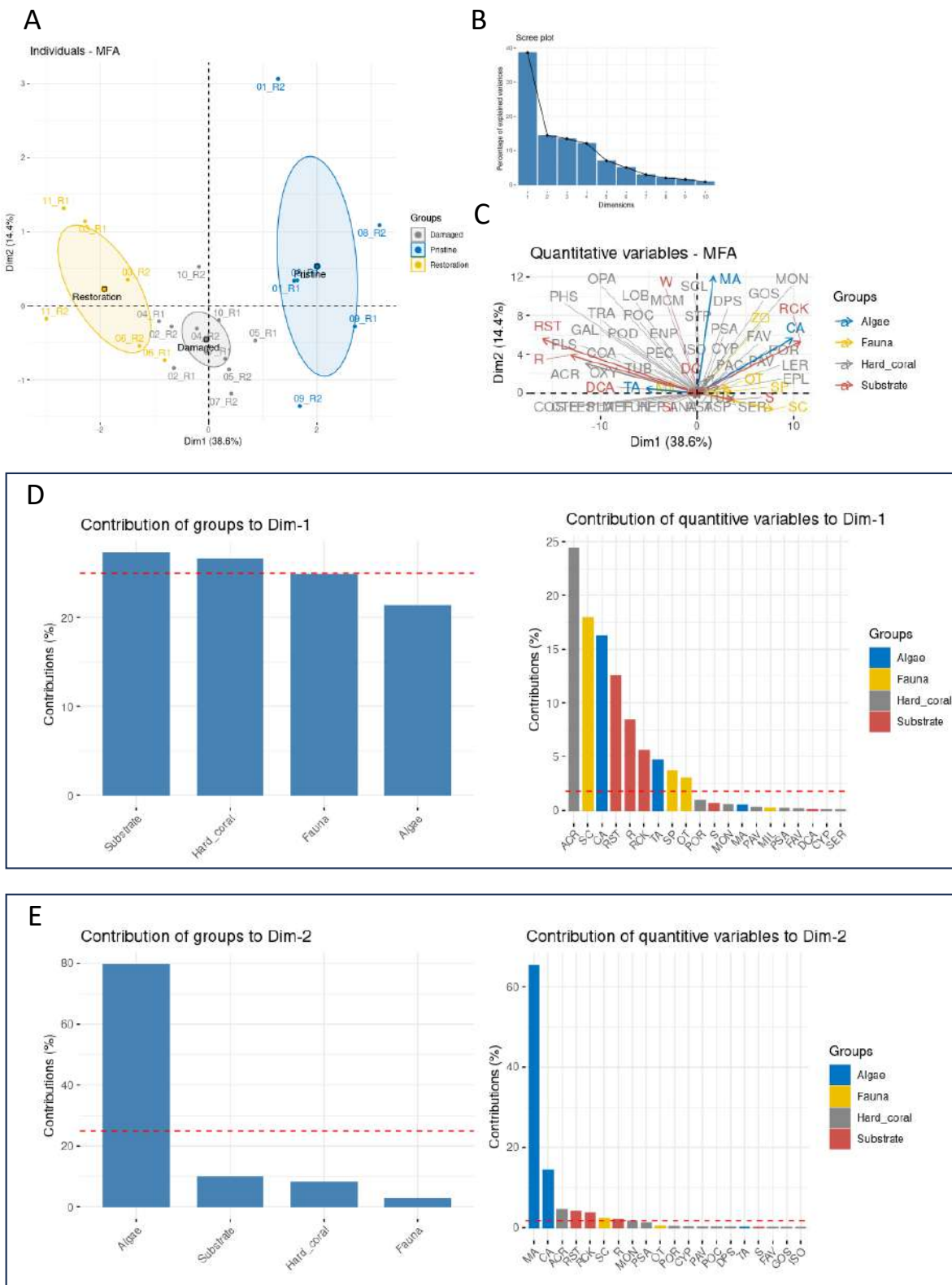


Figure 3: (A) MFA clustering for sites (B) Scree plot showing the percentage of explained variance per dimension for count data from all benthic categories (C) MFA Plot showing benthic categories across dimensions 1 and 2 (D) Graphs showing the makeup of the first dimension by the four major benthic groups (left) and by individual benthic categories (right) (E) Graphs showing the makeup of the second dimension by the four major benthic groups (left) and by individual benthic categories (right)

4

Count data for all Benthic Categories

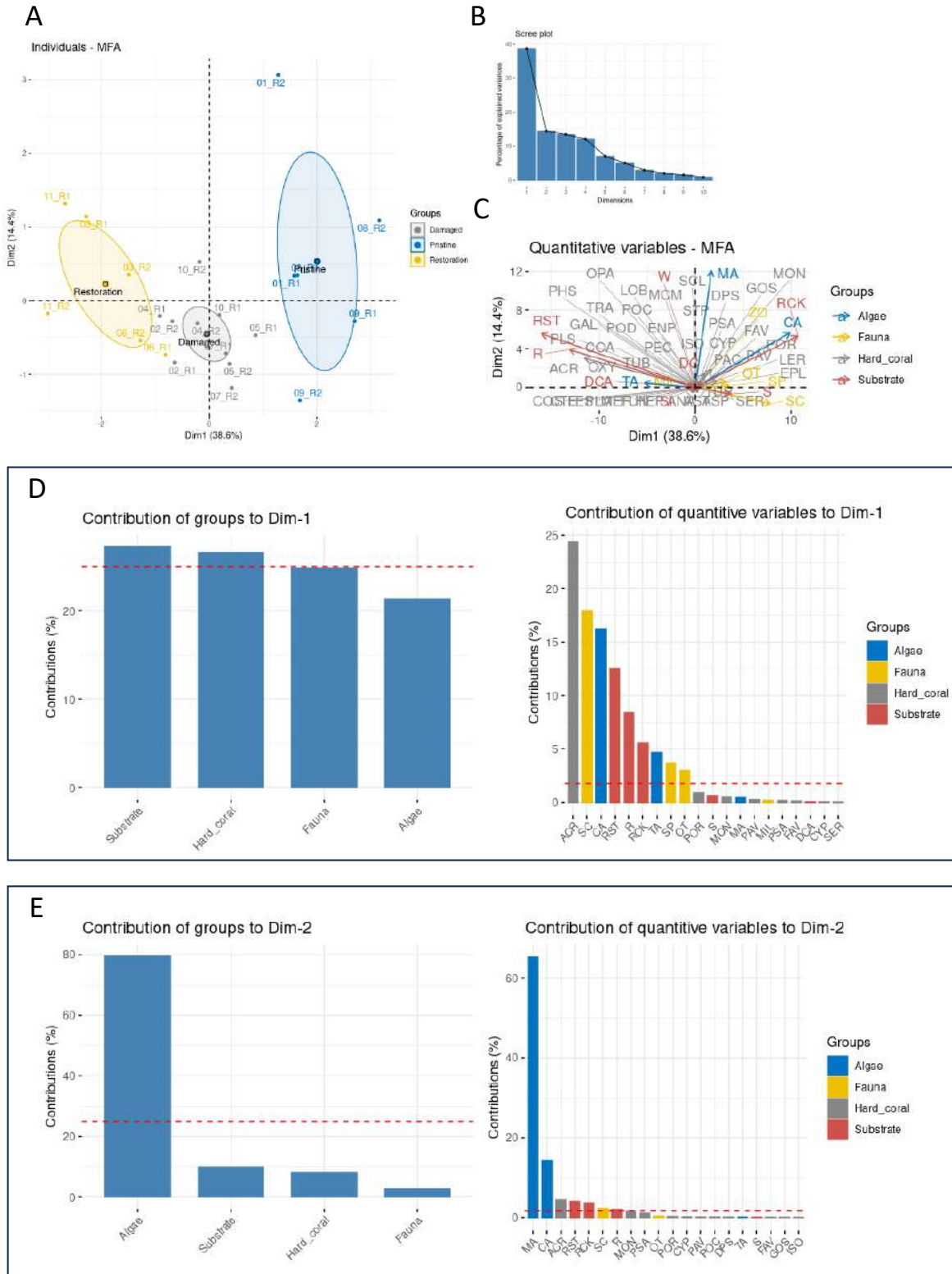


Figure 4: (A) MFA clustering for sites (B) Scree plot showing the percentage of explained variance per dimension for length data from all benthic categories (C) MFA Plot showing benthic categories across dimensions 1 and 2 (D) Graphs showing the makeup of the first dimension by the four major benthic groups (left) and by individual benthic categories (right) (E) Graphs showing the makeup of the second dimension by the four major benthic groups (left) and by individual benthic categories (right)

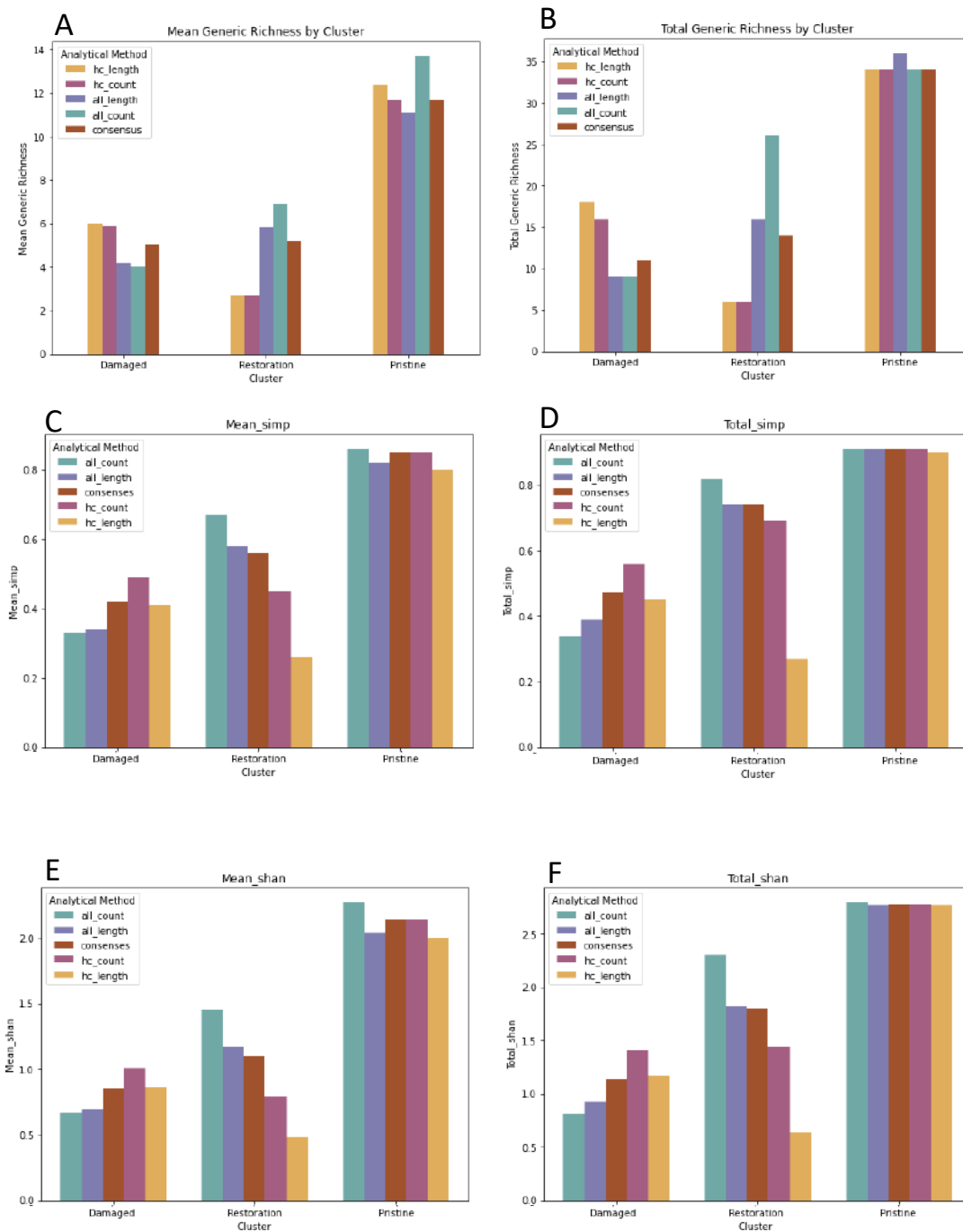


Figure 5: Bar plots showing (A) Mean generic richness by cluster across analytical approaches (B) Total generic richness by cluster across analytical approaches (C) Mean gini-simpson bidiveristy index by cluster and across analytical approach (D) Total gini-simpson bidiveristy index by cluster and across analytical approach (E) Mean shannon bidiveristy index by cluster and across analytical approach (F) Total simpson bidiveristy index by cluster and across analytical approach

Shannon and Simpson Diversity indexes confirm clustering with distinct values per site type across analytical method

To further analyze site groupings, richness of hard corals was calculated to the genus level by site type corresponding to the three clusters for each of the analytical approaches as well as the consensus groupings. For cluster 1 (restoration), mean generic richness had a range of 2.67 -5 with an average of 3.7 genera per replicate (Figure 5A). Total generic richness ranged from 6-11 with an average of 8.2 (Figure 5B). Cluster 2 (damaged) had a mean generic richness that ranged from 5.2-6.9 averaging 6 genera per replicate with a total generic richness ranging from

16-26 and with an average of 18 (Figure 5A, 5B). The mean generic richness of cluster 3 (pristine) ranged from 11.09-13.67 and averaged 12.1 with total richness ranging from 34-36 and averaging 34.4 (Figure 5A, 5B). Here, we can see that average and total generic richness is typically highest for pristine sites followed by damaged sites followed by restoration sites.

Simpson and Shannon diversity indexes were calculated per cluster per analytical approach with this trend almost always respected across all analytical approaches (Figure 5C-F). Due to similar patterns between the two indexes, moving on I will focus only on Simpson.

Biodiversity indexes are better predictor of site type than percent coral coverage

When comparing site type as determined by the consensus dendrogram to the Simpson diversity index, we can see a pattern emerge. Biodiversity of hard corals (to the genus level) as calculated with the Simpson index corresponds almost exactly to our consensus dendrogram site designations with a Simpson index of 0.0-0.6 indicating a restoration site, 0.61-.74 indicating a damaged site and 0.77 and above indicating a pristine site (Table 1). The only two exceptions to this are replicate 02_R1, which had a Simpson value of 0 and was thus not grouped with the other damaged sites, and Site 03_R2, which according to the consensus diagram could have been either restoration or damaged but was labeled as restoration since it is a site experiencing active coral restoration. Interestingly, coral coverage doesn't seem to be a good predictor of site type with the consensus dendrogram groupings having the following ranges: Damaged 3.3-31.9, Restoration:11.8-37 and pristine: 5.8-28.8 (Table 1). Overall, it appears that biodiversity as calculated using the Simpson index is a good predictor of site type.

In order to further evaluate length vs count data site designations, this same exercise was performed on these datasets with length data, more in accordance with the above-mentioned pattern. When the two outliers are ignored (Site 06 R1 and Site 02 R1) it produced the same scale as the consensus dendrogram (Table 1). Thus, insinuating that length data may be a more relevant predictor of site.

Site	Consensus	Coral Coverage	all_count	all_length	Simpson
02_R1	Damaged	5.2	Damaged	Damaged	0.00
11_02	Restoration	21.1	Restoration	Restoration	0.11
11_01	Restoration	11.8	Restoration	Restoration	0.14
03_R1	Restoration	33.5	Restoration	Restoration	0.37
06_R2	Restoration	37.0	Restoration	Restoration	0.48
06_R1	Restoration	18.1	Restoration	Pristine	0.53
10_R2	Restoration	19.8	Damaged	Restoration	0.59
10_R1	Damaged	3.3	Damaged	Damaged	0.64
03_R2	Restoration/Damaged	25.3	Damaged	Damaged	0.70
04_R2	Damaged	31.9	Damaged	Damaged	0.70
02_R2	Damaged	7.4	Damaged	Damaged	0.72
04_R1	Damaged	24.0	Damaged	Damaged	0.74
05_R1	Pristine	13.1	Damaged	Pristine	0.77
09_R2	Pristine	18.6	Pristine	Pristine	0.77
07_R1	Pristine	5.6	Damaged	Pristine	0.83
07_R2	Pristine	12.7	Damaged	Pristine	0.84
09_R1	Pristine	19.9	Pristine	Pristine	0.84
05_R2	Pristine	9.5	Damaged	Pristine	0.85
08_R2	Pristine	18.8	Pristine	Pristine	0.86
08_R1	Pristine	28.1	Pristine	Pristine	0.89
01_R1	Pristine	28.8	Pristine	Pristine	0.90
01_R2	Pristine	25.4	Pristine	Pristine	0.91

Table 1: Table depicting site type as determined by the consensus tree along with clustering derived from all benthic count and all benthic length clustering with sites shown in ascending order according to the Simpson biodiversity index (light green = lower diversity, dark green = higher diversity). Coral coverage is shown as well with lighter purple indicating lower coverage and darker purple indicating higher coverage.

Visual attributes do not explain site type when using current methodology

In order to determine if physiological attributes of individual coral colonies are sufficient to determine site type/coral reef state, individual colonies of the genera *Pocillopora*, *Acropora* and *Porites* were photographed in the same section of reef and at the same depth as where the LIT survey was conducted. For October 2022 only three attributes were recorded underwater: the tissue appearance, predation, and bleaching (Clampitt et al(a), in prep). Overall, 891 colonies were surveyed of which 434 were *Acropora* 237 were *Pocillopora* and 220 were *Porites* (data not shown). 396 colonies showed predation marks, 16 were showing disease or recent tissue loss and 493 displayed bleaching on some portion of the colony (Table 2).

Site	Tissue Appearance	Predation	Bleached
01	0	84	95
02	0	33	67
03	0	26	42
04	0	0	37
05	1	11	14
06	0	23	42
07	4	55	30
08	0	61	43
09	3	75	59
10	0	5	12
11	8	23	52
Total	16	396	493

Table 2: Summary table by site for underwater annotations showing the total number of photos showing presence. Here, "presence" of bleaching refers to our AI definition which corresponds to a 1 or 2 on the coral watch color chart for the lightest value. Tissue appearance was marked as present if the colony was showing disease or recent tissue loss.

For this, we used the manual annotations of the October 2022 survey and correlated it to the site types determined above. When performing a Multiple Correspondence Analysis, no relationship was found between site type and underwater annotations. The first and second dimension only explained 5.1% and 4.7% percent of variance respectively, and the first ten dimensions explained less than 40% of variance (Figure 6A). To see if genus may have an effect, individual analyses were conducted for each of the three genera. Results were roughly the same, with slight improvements seen for *Porites* and *Pocillopora* species but resulting in poor results indicating no trends between visual cues and site type (Figure 6B-D). A similar trend is observed for photo data from May (figure S2) which is currently being analyzed. Thus, it appears that using the current protocol, attributes may not be indicative for site types. The potential reasons for this are considered in more detail in the discussion.

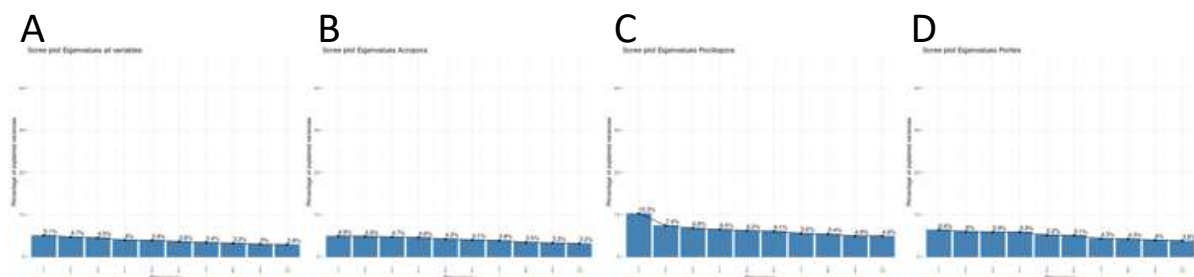


Figure 6: Scree plots for MCA analysis on physiological attributes and type of site. (A) shows all variables (B-D) are per individual genera with (A) for *Acropora* (B) for *Pocillopora* and (D) for *Porites*

Different photographic protocols affect performance on AI machines

Finally, to test the efficiency of the AI machines developed in Clampitt et al(a)., in prep on novel images, we ran these images through our machines for the abovementioned physiological attributes.

Colony Level Photos

Category	F1 Score Class 0	F1 Score Class 1	Support Class 0	Support Class 1
Bleached	0.42	0.64	348	437
Tissue Appearance	0.31	0.04	769	16
Predation	0.39	0.11	745	40

Table 3: Summary table showing F1 scores for both classes (0=absence, 1 = presence) as well as the number of photographs compromising each class for colony level photographs from our October 2022 site visit

F1 scores for colony photographs collected in October ranged from 0.31-0.42 for absence, averaging .0.37. For presence, F1 scores ranged from 0.04 - 0.64, averaging 0.26 (Table 3). Macro level photos averaged lower scores with F1 scores for absence ranging from 0.04-0.72 and averaging 0.29 and presence scores ranging from 0.06-0.15 and averaging 0.10. (Table 4). Datasets were skewed toward absence with more skewed datasets performing worse. Although F1-scores are low, this was expected as we are testing on a novel dataset using a different protocol than the one for which the machines were trained on. However, they can still be improved, especially for Macro-level photographs.

Macro Level Photos

Category	F1 Score Class 0	F1 Score Class 1	Support Class 0	Support Class 1
Bleached	0.72	0.15	338	439
Tissue Appearance	0.04	0.06	762	15
Predation	0.10	0.10	738	39

Table 4: Summary table showing F1 scores for both classes (0=absence, 1 = presence) as well as the number of photographs compromising each class for macro level photographs from our October 2022 site visit

Thus, we modified the photographic and underwater surveying methodologies to obtain attribute-specific macro photos and *in-situ* annotations for all nine attributes (bleaching, predation, boring organisms, pigmentation, macroalgal contact, CCA, turf algae, tissue appearance, and sediment contact). Each colony was photographed 20 times as opposed to 3 (Figure 7A) (Figure S1) (SI Datasheet 2). The main photographic difference resided at the macro level with colony-level and wide-angle photos still being taken. In May, 536 colonies were photographed with 176 *Acropora* colonies, 170 *Pocillopora* colonies and 190 *Porites* colonies. Table 5 summarizes the number of colonies where a physiological attribute was marked as present per site across all 9 attributes.

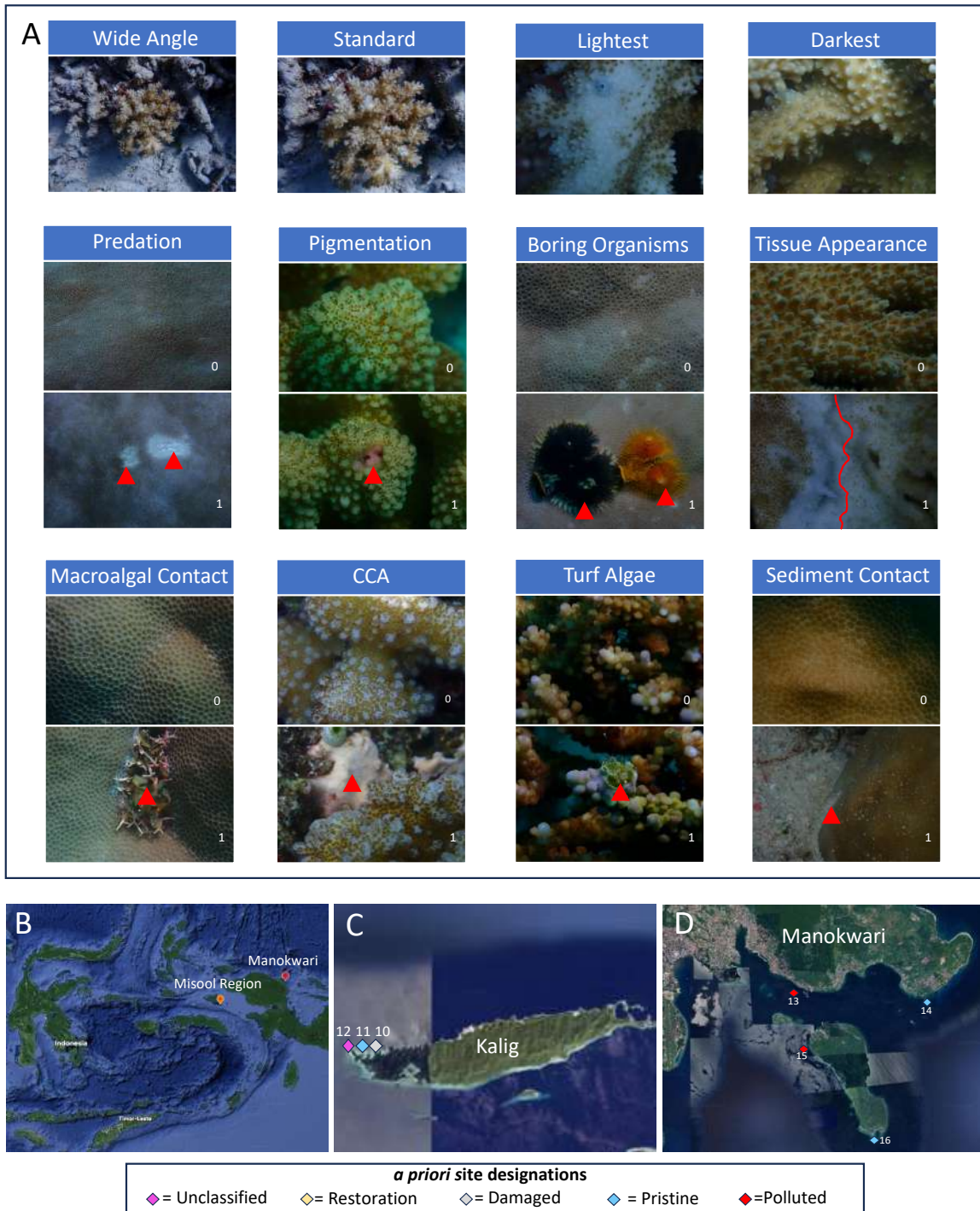


Figure 7: (A) Example of 20 photographs with two per each attribute (one showing absence and one showing presence) using the novel methodology implemented in May 2023 (B) Map showing the Misool Region where sites 1-12 are located and Manokwari where sites 13-16 are located (C) Map showing Kalig island and the location of site 12 which was added in May 2023 (D) Map showing Manokwari where sites 13-16 are located

Site	Tissue Appearance	Predation	Bleached	Boring Organism	Algal Contact	Turf Algae	CCA	Pigmentation	Sediment
01	0	43	44	17	24	13	30	2	34
02	0	5	12	7	7	0	1	0	6
03	1	20	40	16	28	5	7	3	19
04	0	9	21	12	16	3	2	1	11
05	0	13	21	4	15	3	11	0	17
06	0	20	31	10	15	2	5	2	15
07	0	31	31	14	18	10	21	3	24
08	0	35	29	14	23	8	13	1	29
09	0	22	23	7	9	0	18	1	18
10	0	10	14	5	4	5	4	1	9
11	1	11	16	2	5	1	4	0	1
12	0	11	14	5	6	1	8	1	3
13	7	11	13	8	12	8	15	10	19
14	6	12	16	12	17	11	11	8	18
15	1	7	6	4	10	0	7	8	10
16	3	8	6	6	5	3	3	0	9
Total	19	268	337	143	214	73	160	41	242

Table 5: Table depicting "presence" by site as calculated through underwater water annotations. Here bleaching is "present" if the color code value annotated for the lightest portion of the colony was 1 or 2.

Here, F1 scores ranged from 0.3 to 0.7 for absence when utilizing the colony photographs and .06-.65 for presence (Table 6). The lowest score for both classes resulted from the machine for tissue appearance which also had the most skewed dataset (Table 6). When using the macro level photographs, F1 scores ranged from .08 to .82 for absence and .03 to .45 for presence (Table 7). Once again, tissue appearance performed the worst for both categories and once again had the most skewed dataset (Table 7).

We can also compare F1 scores for Tissue Appearance, Bleaching and Predation to the F1 scores of the Moorea dataset presented in Clampitt et al(a), in prep. We will focus on class 1, or presence, which is the class indicating physiological perturbations, *i.e.*, tissue loss. With resize as a pretreatment, the Moorea photographs average 0.27 (Clampitt et al(a), in prep) for all three attributes compared to 0.10, and 0.29 for our October Indonesia dataset and 0.29 and 0.43 for our May Indonesia dataset, for macro and colony level photographs, respectively (Tables 3,4,6, & 7). For May and October datasets colony photographs achieved higher F1 scores than macro photographs, suggesting that the colony level photos are more similar to the training set and the macro photos are less similar. These results were expected, as novel methodology was used that focused on the macro attribute. The average F1 score for colony level photos increased as well, indicating that the new methodology may be more successful.

To improve upon these scores, more training samples are needed as the photographic protocol is now completely different from initial protocol (Clampitt et al(a)., in prep) and underwater annotations are more robust than AI extracted annotations. Additionally, new machines will be trained utilizing the photographs generated in Indonesia which would then be based on the more robust underwater annotations. Thus, improved AI machines are being developed.

Colony Level Photos

Category	F1 Score Class 0	F1 Score Class 1	Support Class 0	Support Class 1
Bleached	0.47	0.65	242	294
Boring Organism	0.61	0.48	295	241
Algal Contact	0.45	0.34	414	122
Turf Overgrowth	0.70	0.40	348	188
Sediment	0.49	0.18	495	41
Tissue Appearance	0.30	0.06	518	18
CCA	0.44	0.22	470	66
Pigmentation	0.49	0.50	383	153
Predation	0.39	0.59	312	224

Table 6: Summary table showing F1 scores for both classes (0=absence, 1 = presence) as well as the number of photographs comprising each class for colony level photos

Macro Level Photos

Category	F1 Score Class 0	F1 Score Class 1	Support Class 0	Support Class 1
Bleached	0.13	0.45	772	300
Boring Organism	0.28	0.35	828	240
Algal Contact	0.17	0.20	950	122
Turf Overgrowth	0.82	0.31	884	188
Sediment	0.09	0.08	1030	41
Tissue Appearance	0.08	0.03	1054	18
CCA	0.14	0.12	1004	66
Pigmentation	0.22	0.26	917	153
Predation	0.10	0.35	848	224

Table 7: Summary table showing F1 scores for both classes (0=absence, 1 = presence) as well as number of photographs comprising each class for macro level photographs

Discussion

Redefining the term “pristine”

Coral coverage has long been utilized as a common metric for defining reef health (Bruno & Selig, 2007) with most evaluations being based on the linear scale as defined by Gomez and Yap (1988) where 0-<25% indicates poor, 25-<50% indicates moderate (fair or average), 50-<75% indicates good and 75-100% indicates excellent (very good) (Wilkinson, 2008) (Tun et al., 2005) (Habibi & Sartin, 2007). Worldwide, living coral cover has decreased dramatically with losses estimated at 1/3 or 2/3 of living coral cover from a few decades ago (Knowlton & Jackson, 2008). Given global decreases in coral coverage and shifting baselines, metrics evaluating the state of coral reefs in the Anthropocene need reevaluation. It is important to note that historical baselines and data on coral coverage do not represent sites unaffected by human impacts, with “pristine” reefs essentially non-existent (Knowlton & Jackson, 2008). Thus, to clarify, our use of the term pristine doesn't correspond to historical baselines but is used in this instance to refer to some of the least disturbed reefs, that exist in a remote location and are thus hopefully experiencing fewer human disturbances than other reef localities. Here, the term “pristine” is loosely used to refer to undisturbed reefs by today's standards when considering the worldwide documented degradation of these ecosystems. Furthermore, when looking at data evaluating reef health, depths are usually grouped together or not specified when presenting region wide or global averages. In this study, surveys were conducted at 5 or 10m depth with pristine site surveys done at 10m. If surveys had been done at shallower depths say 2m, based on personal estimation, these sites would have most likely shown greater than 75% coral coverage following into the excellent category (personal observations).

Coral biodiversity index a good predictor of site type

Overall, the three types of sites as were originally designated and as confirmed by our clustering and LIT data correspond to ranges within the Simpson biodiversity index with restoration sites showing the lowest diversity followed by damage sites and pristine sites having the highest biodiversity (Table 1). Coral coverage as determined by the total length of all hard coral genera divided by the total length of the transect appears to be a less reliable predictor of site types with variations in coral cover seen across the site types. Based on *a priori* site designations, pristine sites averaged 23.2% coral cover, restoration sites averaged 24.6% and damaged sites averaged 9.72%. With the consensus dendrogram these values shifted to 18.05%, 23.8%, and 14.36% respectively. In both designations, all three sites would fall under the poor classification, thus proving identical based on the metric of coral coverage alone. However, they show drastic differences in the Simpson biodiversity indexes which range from 0.0 to 0.91 across all 22 replicates and often do not correlate to percent coral coverage (Table 1). Thus, the biodiversity index, calculated using only the generic diversity of hard corals seems to be a more reliable predictor of site type. To confirm this trend, further studies need to be conducted ideally incorporating a wider range of coral coverage. Our site with the highest coral coverage has 37% coverage, ideally, we would evaluate sites showing all four categories of coral coverage from poor to excellent (0-100%) as defined by Gomez and Yap (1988). As our average coral coverage was considered poor with a few sites showing moderate coverage but none in the good or excellent categories. It is possible that different trends would arise when including sites exhibiting moderate, good, or excellent coverage.

Coral Coverage is important but may not compare sites effectively, consider additional metrics chosen with study goals in mind

Despite this finding, we do not mean to disregard the importance of coral cover especially in terms of 3-D structure and habitat creation which becomes even more apparent when evaluating sites damaged by dynamite fishing. Low coral cover results in a loss of 3D

topography that is vital to these ecosystems both directly and indirectly (Knowlton & Jackson, 2008). When conducting surveys, although in certain instances damaged sites showed higher diversity than restoration sites, the colonies accounting for this diversity were very small (Personal observation) and may not survive to adulthood due to the shifting substratum typical of these sites (Pearson, 1981). Thus, a damaged site with high diversity of very small coral colonies does not have the same value as a restoration site with high coverage but low diversity.

Coral coverage can be a very useful metric in providing general information about a site and can be evaluated faster and with less expertise than a biodiversity survey requiring the researcher to differentiate between genera, but our results indicate that it may not adequately compare different sites. Especially given that *a priori* designations, which proved relatively robust, were based on global topography or site characteristics prior to surveying *i.e.*, an observer can see differences between these sites even though coral coverage classifications alone would group them together. This is lightly supported by our MFA results with rock defining pristine sites, restoration structures defining restoration sites and rubble defining damaged sites. The substrate type could be loosely viewed as one of the main metrics of the *a priori* site designations and was supported by LIT survey data. Madduppa & Zamani (2011) support the idea that coral coverage alone does not quantify reef health and suggest incorporating other metrics such as algal coverage, sand coverage, and mortality index while also accounting for reef location (estuary, coastal, and small island).

Possibility for faster, more efficient methodology by focusing on hard coral genera

These factors were chosen as means to evaluate the current state of the reef while also attempting to predict its future trajectory *i.e.*, if two sites had similar coral coverage but one was showing high mortality or algal coverage and the other was not, the site with increased algae and mortality indicates a trajectory towards further degradation. A methodology using biodiversity indexes, as proposed here, would be more aimed at evaluating the current state of the reef. Furthermore, collecting biodiversity indexes alone (*i.e.*, counting the number of genera present at a site) as opposed to conducting a full LIT survey would be much quicker in terms of data collection and analysis. Even if the LIT was adapted to only collect information about hard corals (so that hard coral coverage could be calculated as well) this would reduce surveying time and allow for further investigation of the pattern between biodiversity and site type as presented in this study. Thus, the LIT survey could be adapted dependent on the goal of the study *i.e.*, algal coverage or mortality index could also be incorporated to tease out the potential future trajectory of the site.

Another possible confounding factor that we are aware of is that certain benthic categories of the LIT survey were only found in certain site types. This mainly applies to restoration structures which was a benthic category that was specific to restoration sites. However, despite this, restoration structure was not the most important category in explaining variance. Although restoration structure did help to explain the variance for the first dimension (Figure 3D), it only accounted for about 12% of overall variance whereas *Acropora* accounted for almost double that (data not shown). Despite this, our cluster corresponding to restoration sites can still be validated by looking at hard coral only data which did not take restoration structures into account. When creating clusters utilizing only the hard corals, the majority of restoration sites were always grouped together strengthening the validity of this cluster (Table S1).

Methodological changes could reveal relationships between the 9 attributes and types of site/state of the reef

When investigating the correlation between the attributes and site type, current analyses didn't provide any concrete results or highlight certain attributes that may carry more weight in

defining type of site and thus evaluating reef condition. This could be explained by the fact that the nine attributes are pertinent metrics for evaluating coral physiology/ health but may not provide enough information for ecological assessments. Alternatively, this could be due to a biased selection of colonies for annotations that did not reflect the general coral health status of the site. Actually, the current methodology was aimed at increasing photographic samples in order to increase training samples and improve our AI models presented in Clampitt et al(a)., in prep. Thus researchers/divers would attempt to select an equal number of colonies for the three target genera that were in the same depth category and in the general vicinity of the area where the LIT survey was conducted. Using this methodology, selected colonies were more so the colonies that the divers chose to observe as compared to colonies that represented the global trends of the site. To illustrate this, we can look at genus; divers attempted to photograph an equal number of colonies for the three target genera when the three genera were not equally represented in terms of coverage as determined by the LIT survey (Figure S3). Therefore, we would like to implement a stricter, less biased methodology regarding individual colony selection for in water annotations and photographs, such as using transects or quadrats and surveying all colonies of target genera within a certain area or along the transect line. Such methodology, similar to the one deployed for the LIT survey, would allow us to better decipher the presence or absence of the nine physiological hallmarks by site while respecting the composition of the site.

Additionally, to further study whether attributes could be affecting site types more targeted studies would be useful. For example, it has already been shown that boring organisms increase in relation to eutrophication (Rice et al., 2020). Thus, using our data to conduct an analysis focused on the difference of boring organisms between polluted vs. non polluted sites may be able to give us more information on whether our photographic sampling method can validate this relationship. Seeing as we have included polluted sites in our May site visits, this is something we plan to further investigate. Since, this is a known relationship a more targeted analysis may be able to confirm this relationship. As a caveat, however, we did not take measures of pollution but know that the sites were polluted based on local descriptions. Thus, to truly inform analyses and in future site visits, maybe nutrient measurements could be taken to see if we can a) validate differences between sites and then b) see if the analyses can confirm these differences. As another example of how studies could be targeted towards a specific attribute, surveys could be conducted during a bleaching event and during a period of non-bleaching thus differences could be compared not across site types but also within individual sites with the temporal aspect being of more relevance to this particular attribute and research questions surrounding bleaching. Studies have already shown that some of the hallmarks may be useful as proxies in determining reef health (as presented in Clampitt et al(a)., in prep) thus, by creating more targeted studies specialized to one physiological attribute we may be able to discern patterns among sites.

Another interpretation of the absence of correlation between the attributes and the type of site could be that analysis provides qualitative information which may not be sensitive enough to provide global information about the site beyond that of the state of the individual colony. Our analysis contains information about presence or absence. This can be demonstrated by using the same examples as above: boring organisms and bleaching. With boring organisms for example, we know that they are present in the colony but presence for one annotation/photo could mean one boring organism in the colony whereas in another annotation/photo the colony could be completely infested. Thus, in reality these two colonies are differentially affected by this hallmark. Studies have shown that density of boring organisms differs between types of sites (Scaps & Denis, 2008) (Rice et al, 2020). Thus, it's possible that to truly investigate this

question, analysis would need to be quantitative. Importantly, we have taken photographic data along with our in-water annotations. Therefore, we have permanent records of each colony and conducting such an analysis would not be outside the scope of possibility. Similar to boring organisms, bleaching is also annotated in a qualitative matter. We may know that the colony is exhibiting bleaching, but we don't know what percentage or to what extent. Thus, a colony experiencing a small patch of bleaching covering maybe 5% of the colony would have the same status as a colony exhibiting 95% bleaching. In defining site type or defining the state of the reef, the information we have collected on presence or absence, although useful for characterizing the individual colony, may not be sensitive enough to explore global trends present at a particular site. Furthermore, qualitative information may lead to a wider array of statistical analyses to explore. Thus, Quantitative information might be necessary to truly examine links between the physiological hallmarks and site type.

AI testing set provides promising results and will help lead to improved machines

Although, at first glance, results from our AI machines may seem subpar they performed quite well due to the use of novel photographs and in-water annotations. Photographs used were from a novel photographic dataset that uses not only different photographic methodology than the one which all the machines were trained on, but we also used in water annotations. When using novel photographs on AI machines, it is expected for a poorer performance than when machines are tested on photos from the same dataset as they were trained on. Especially since our machines were built using only annotations done on photographs and we are testing them exclusively on in-situ annotations. These two styles of annotations (in water vs photographic) could result in differences in annotations with underwater annotations allowing for a more thorough investigation of the colony as well as a full view of information from the surrounding environment. An example of this can be seen in classifying snail predation vs disease or rapid tissue loss. Snail predation can sometimes look like white syndrome or another form of rapid tissue loss (personal observation) (Figure S4). Through *in situ* observation of the colony, this difference is easily discerned mostly by the presence of snails which are often hidden, clustered in-between the branches or at the base of the colony, especially during the day (personal observation). They would most likely not be visible or evident from a photograph, especially for branching colonies like *Pocillopora* where branches are thick and compact. Thus, this could result in a photographic annotation of presence for tissue appearance but an underwater annotation of presence for predation. Similarly, the quality of the photo may have complicated the annotation with zooming resulting in fuzzy resolution or an inability to fully investigate a point of interest whereas in situ this would not be a problem and furthermore all angles of the colony could be examined.

On a similar note, with regards to presence and absence in our photographic dataset, we can see that the datasets are skewed towards absence for both macro and colony level photos. This is an issue we have already seen, and which is discussed in more detail in Clampitt et al(a), in prep. Class one photographs (presence) for the nine attributes surveyed in the May dataset averaged 14% of the total photos per attribute for the macro photos versus 29% for the colony photos with the lowest f1 scores resulting from the more skewed datasets (Table 6, Table 7). The macro photographs also averaged a lower F1 score for class 1 at 0.24 than the colony photographs which averaged .38. These results reflect those presented in Clampitt et al(a), in prep prior to photo augmentation. Photo augmentation allowed for a marked increase in F1 scores (Clampitt et al(a), in prep). Thus, the probability of increasing F1 scores by increasing the size of the dataset is heavily supported not only by the data we show here but also by data presented in Clampitt et al(a), in prep. Thus, to build upon the work in the paper, we hope to incorporate this dataset into our previous dataset (photographic datasets presented in Clampitt

et al(a), in prep) to increase examples of presence and add underwater annotations to our training set. Additionally, the novel methodology incorporating macro photos might prove interesting to develop machines capable of identifying and categorizing specific attributes either as a research tool or an educational tool for curious divers or local coral gardeners who may want to gain more knowledge about what they are seeing. It could provide an easy-to-use monitoring schema for those who might not have a background in coral physiology.

References

- Bruno, J. F., & Selig, E. R. (2007). Regional Decline of Coral Cover in the Indo-Pacific: Timing, Extent, and Subregional Comparisons. *PLOS ONE*, 2(8), e711. <https://doi.org/10.1371/journal.pone.0000711>
- Bruno, J., & Bertness, M. (2001). Habitat modification and facilitation in benthic marine communities. In *Marine Community Ecology* (Vol. 413, pp. 201–218).
- Burke, L., Reytar, K., Spalding, M., & Perry, A. (2012). *Reefs At Risk Revisited in the Coral Triangle*.
- Done, T. J. (1992). Phase shifts in coral reef communities and their ecological significance. *Hydrobiologia*, 247(1), 121–132. <https://doi.org/10.1007/BF00008211>
- Edinger, E. N., & Risk, M. J. (1998). Reef Degradation and Coral Bleaching in Indonesia: Effects of Land-based Pollution, Destructive Fishing Practices and Changes Over Time. *Marine Pollution Bulletin*, 36(8).
- English, S. A., Wilkinson, C., Baker, V., & Australian Institute of Marine Science (Eds.). (1997). *Survey manual for tropical marine resources* (2. ed). Australian Institute of Marine Science
- Fadli, N., Campbell, S. J., Ferguson, K., Keyse, J., Rudi, E., Riedel, A., & Baird, A. H. (2012). The role of habitat creation in coral reef conservation: A case study from Aceh, Indonesia. *Oryx*, 46(4), 501–507. <https://doi.org/10.1017/S0030605312000142>
- Fox, H. E., & Caldwell, R. L. (2006). Recovery from blast fishing on coral reefs: A tale of two scales. *Ecological Applications: A Publication of the Ecological Society of America*, 16(5), 1631–1635. [https://doi.org/10.1890/1051-0761\(2006\)016\[1631:rbbfoc\]2.0.co;2](https://doi.org/10.1890/1051-0761(2006)016[1631:rbbfoc]2.0.co;2)
- Graham, N. A. J., Wilson, S. K., Jennings, S., Polunin, N. V. C., Bijoux, J. P., & Robinson, J. (2006). Dynamic fragility of oceanic coral reef ecosystems. *Proceedings of the National Academy of Sciences*, 103(22), 8425–8429. <https://doi.org/10.1073/pnas.0600693103>
- Gomez E D, Yap H T. 1988. Monitoring reef condition. In Kenchington RA, Hudson BET (eds). *Coral Reef Management Handbook*. Jakarta: UNESCO regional office for science and technology for southeast Asia (ROSTSEA). 171-178.
- Hadi, T., Muhammad, A., Giyanto, G., Prayudha, B., Johan, O., Budiyo, A., Rezza, A., Alifatri, L. O., Sulha, S., & Shar, S. (2020). *The Status of Indonesian Coral Reefs 2019*. Research Center for Oceanography. https://www.researchgate.net/publication/342663285_The_Status_of_Indonesian_Coral_Reefs_2019
- Habibi, A., & Sartin, J. (2007). *A Decade of Reef Check Monitoring: Indonesian Coral Reefs, Condition and Trends*. <https://www.semanticscholar.org/paper/A-Decade-of-Reef-Check-Monitoring%3A-Indonesian-Coral-Habibi-Sartin/6380350c12d3602d2c897bb5b24482746ef82a89>

Husson, F., Josse, J., Le, S. & Mazet, J. FactoMineR: Multivariate Exploratory Data Analysis and Data Mining. 2023. <https://cran.r-project.org/web/packages/FactoMineR/index.html>

Huston, M. A. (1985). Patterns of Species Diversity on Coral Reefs. *Annual Review of Ecology and Systematics*, 16(1), 149–177.
<https://doi.org/10.1146/annurev.es.16.110185.001053>

Kassambara, A. & Mundt, F. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. 2020. <https://cran.r-project.org/web/packages/factoextra/index.html>

Knowlton, N., & Jackson, J. B. C. (2008). Shifting Baselines, Local Impacts, and Global Change on Coral Reefs. *PLOS Biology*, 6(2), e54.
<https://doi.org/10.1371/journal.pbio.0060054>

Madduppa, H., & Zamani, N. (2011). A Standard Criteria for Assessing the Health of Coral Reefs: Implication for Management and Conservation. *Journal of Indonesia Coral Reefs*, 1, 137–146.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. cluster: ‘Finding Groups in Data’: Cluster Analysis Extended. 2022. <https://cran.r-project.org/web/packages/cluster/index.html>

Mcfield, M., Kramer, P., Gorrez, M., & McPherson, M. (2007). *Healthy Reefs for Healthy People: A Guide to Indicators of Reef Health and Social Well-being in the Mesoamerican Reef Region*.

Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, et al. vegan: Community Ecology Package. 2022. <https://cran.r-project.org/web/packages/vegan/index.html>

Pearson, R. (1981). Recovery and Recolonization of Coral Reefs. *Marine Ecology Progress Series*, 4, 105–122. <https://doi.org/10.3354/meps004105>

Permana, R., Akbarsyah, N., Putra, P. K., & Andhikawati, A. (2020). Analysis Condition of Coral Reef Covering in Pramuka Island Waters, Seribu Islands using Line Intercept Transect (LIT) Method. *Jurnal Riset Biologi Dan Aplikasinya*, 2(2), Article 2.
<https://doi.org/10.26740/jrba.v2n2.p77-81>

Rice, M. M., Maher, R. L., Correa, A. M. S., Moeller, H. V., Lemoine, N. P., Shantz, A. A., Burkepile, D. E., & Silbiger, N. J. (2020). Macroborer presence on corals increases with nutrient input and promotes parrotfish bioerosion. *Coral Reefs*, 39(2), 409–418.
<https://doi.org/10.1007/s00338-020-01904-y>

Roberts, C., Mcclean, C., Veron, J., Hawkins, J., Allen, G., McAllister, D., Mittermeier, C., Schueler, F., Spalding, M., Wells, F., Vynne, C., & Werner, T. (2002). Marine Biodiversity Hotspots and Conservation Priorities for Tropical Reefs. *Science (New York, N.Y.)*, 295, 1280–1284. <https://doi.org/10.1126/science.1067728>

Rotjan, R., & Lewis, S. (2008). Impact of coral predators on tropical reefs. *Marine Ecology Progress Series*, 367, 73–91. <https://doi.org/10.3354/meps07531>

Scaps, P., & Denis, V. (2008). Can organisms associated with live scleractinian corals be used as indicators of coral reef status? *Atoll Research Bulletin*, 566, 1–18. <https://doi.org/10.5479/si.00775630.566.1>

Siebeck, U. E., Marshall, N. J., Klüter, A., & Hoegh-Guldberg, O. (2006). Monitoring coral bleaching using a colour reference card. *Coral Reefs*, 25(3), 453–460. <https://doi.org/10.1007/s00338-006-0123-8>

R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2020. <https://www.R-project.org/>.

Tun, K., Chou, L. M., Cabanban, A., Tuan, V. S., Suharsono, T. Y., Sour, K., & Lane, D. (2005). *Status of coral reefs, coral reef monitoring and management in Southeast Asia 2004*. <https://digitalarchive.worldfishcenter.org/handle/20.500.12348/1894>

Vercammen, A., McGowan, J., Knight, A. T., Pardede, S., Muttaqin, E., Harris, J., Ahmadi, G., Estradivari, Dallison, T., Selig, E., & Beger, M. (2019). Evaluating the impact of accounting for coral cover in large-scale marine conservation prioritizations. *Diversity and Distributions*, 25(10), 1564–1574. <https://doi.org/10.1111/ddi.12957>

Williams, S. L., Sur, C., Janetski, N., Hollarsmith, J. A., Rapi, S., Barron, L., Heatwole, S. J., Yusuf, A. M., Yusuf, S., Jompa, J., & Mars, F. (2019). Large-scale coral reef rehabilitation after blast fishing in Indonesia. *Restoration Ecology*, 27(2), 447–456. <https://doi.org/10.1111/rec.12866>

Wilkinson, C. (2004). Status of Coral Reefs of the World. Global Coral Reef Monitoring Network and Australian. *Institute of Marine Science, Townsville, 1*.

Supplementary Information:

The following information may be found below or attached to this article:

Attached:

Table S1. Cluster membership by site across the four analytical approaches and for a priori and consensus groupings

Figure S1. Figure depicting the three types of photographs as presented in Clampitt et al(a), in prep and utilized for the October 2022 site visit.

Figure S2. Scree plots for MCA analysis of underwater annotations by site

Figure S3. (A) Bar plot showing target genera annotated in situ per site (B, C) specific coral genera coverage by site

Figure S4. Photographs showing a wide angle and a macro photo for one colony experiencing snail predation and another that is most likely diseased.

Data Sheet S1. Example datasheet and codes for LIT Surveys

Data Sheet S2. Example datasheet for underwater annotations

4.3 Supplementary Information

Site Number	Site Groupings					
	All Count	All Length	Hard Coral Count	Hard Coral Length	Consensus	A priori
01_R1	3	3	3	3	3	3
01_R2	3	3	3	3	3	3
08_R1	3	3	3	3	3	3
08_R2	3	3	3	3	3	3
09_R1	3	3	3	3	3	3
09_R2	3	3	3	1	3	3
05_R1	2	3	3	3	3	2
05_R2	2	3	3	3	3	2
07_R1	2	3	3	3	3	0
07_R2	2	3	3	3	3	0
03_R1	1	1	1	1	1	1
06_R2	1	1	1	1	1	1
11_02	1	1	1	1	1	1
11_01	1	1	1	1	1	1
06_R1	1	3	1	1	1	1
10_R2	2	1	1	1	1	2
03_R2	2	2	1	1	1	1
10_R1	2	2	2	2	2	2
02_R1	2	2	1	2	2	2
02_R2	2	2	1	2	2	2
04_R1	2	2	2	1	2	0
04_R2	2	2	2	1	2	0

Table S 1: Cluster membership by site across the four different analytical approaches (all benthic data length, all benthic data count, hard coral only length, hard coral only count) as well as a priori and consensus site membership. Cluster one is yellow (Restoration), cluster two is gray (Damaged) and cluster 3 is blue (Pristine). 0 refers to unclassified and displays as white

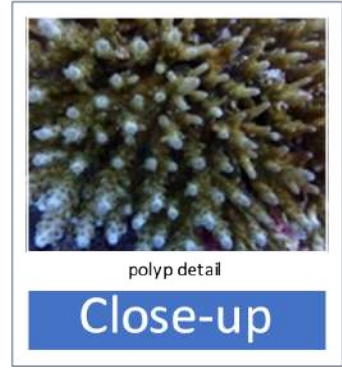


Figure S 1: Example photographs of a *Pocillopora* Colony using the three photos only methodology as presented in Clampitt et al., in prep

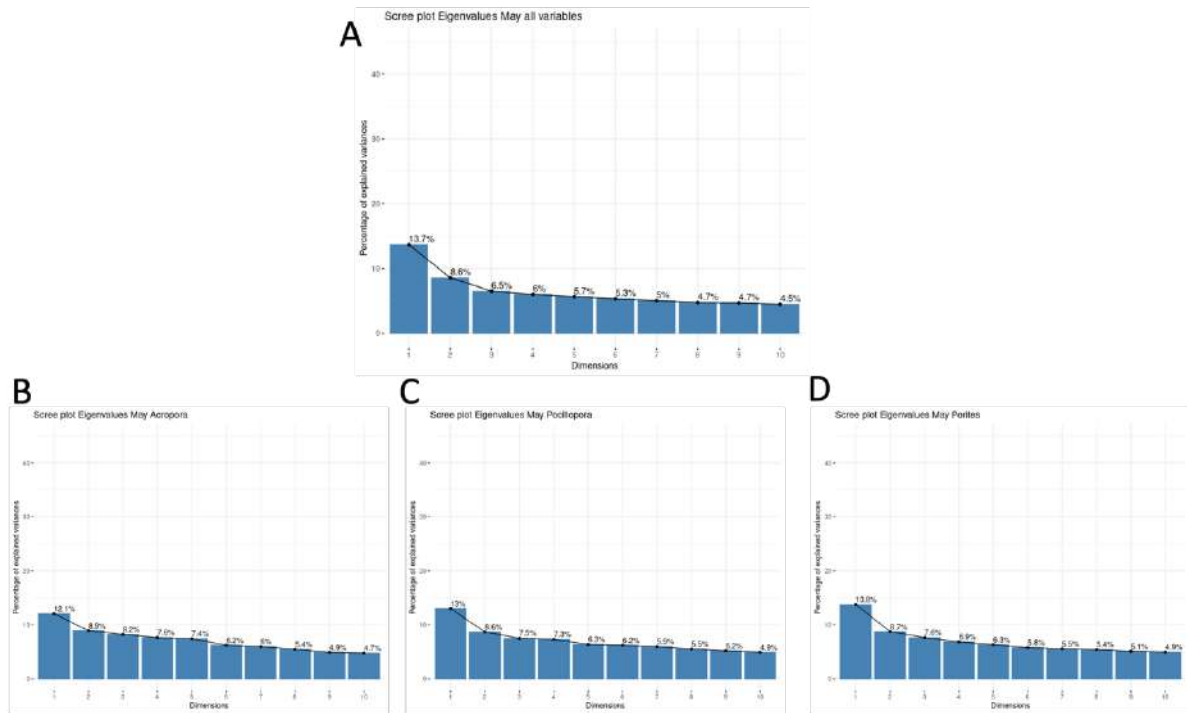


Figure S2: Scree plots for MCA analysis on physiological attributes and type of site for May 2023 visit (A) shows all variables (B-D) are per individual genera with (A) for Acropora (B) for Pocillopora and (D) for Porites

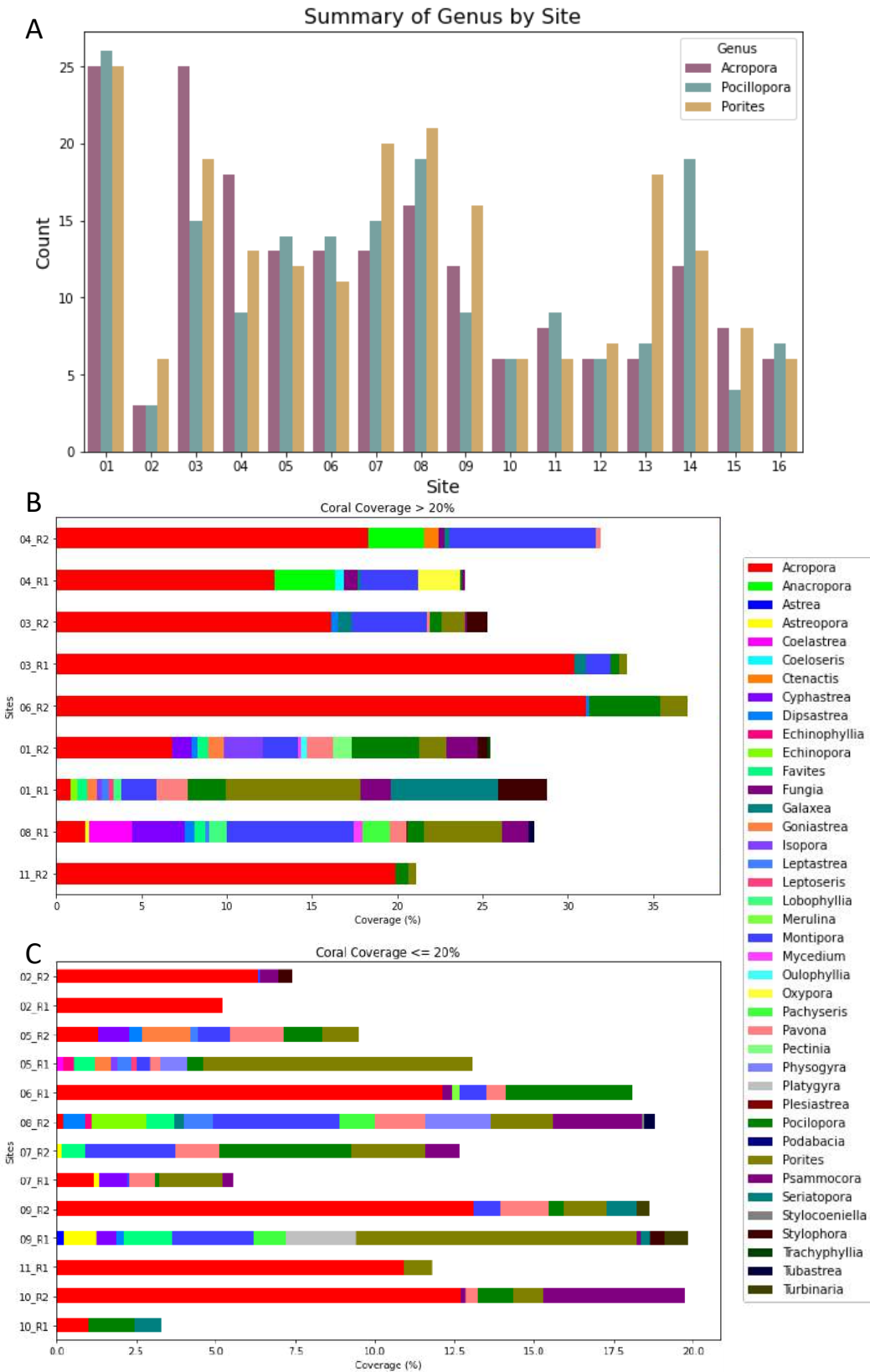


Figure S3: (A) Bar plot showing the number of the three target genera (*Acropora*, *Pocillopora*, *Porites*) which were annotated in-situ per site (B) Coral genera by site for coverage greater than 20% (C) Coral genera by site for coverage less than or equal to 20% - graphs were separated for ease of lecture.

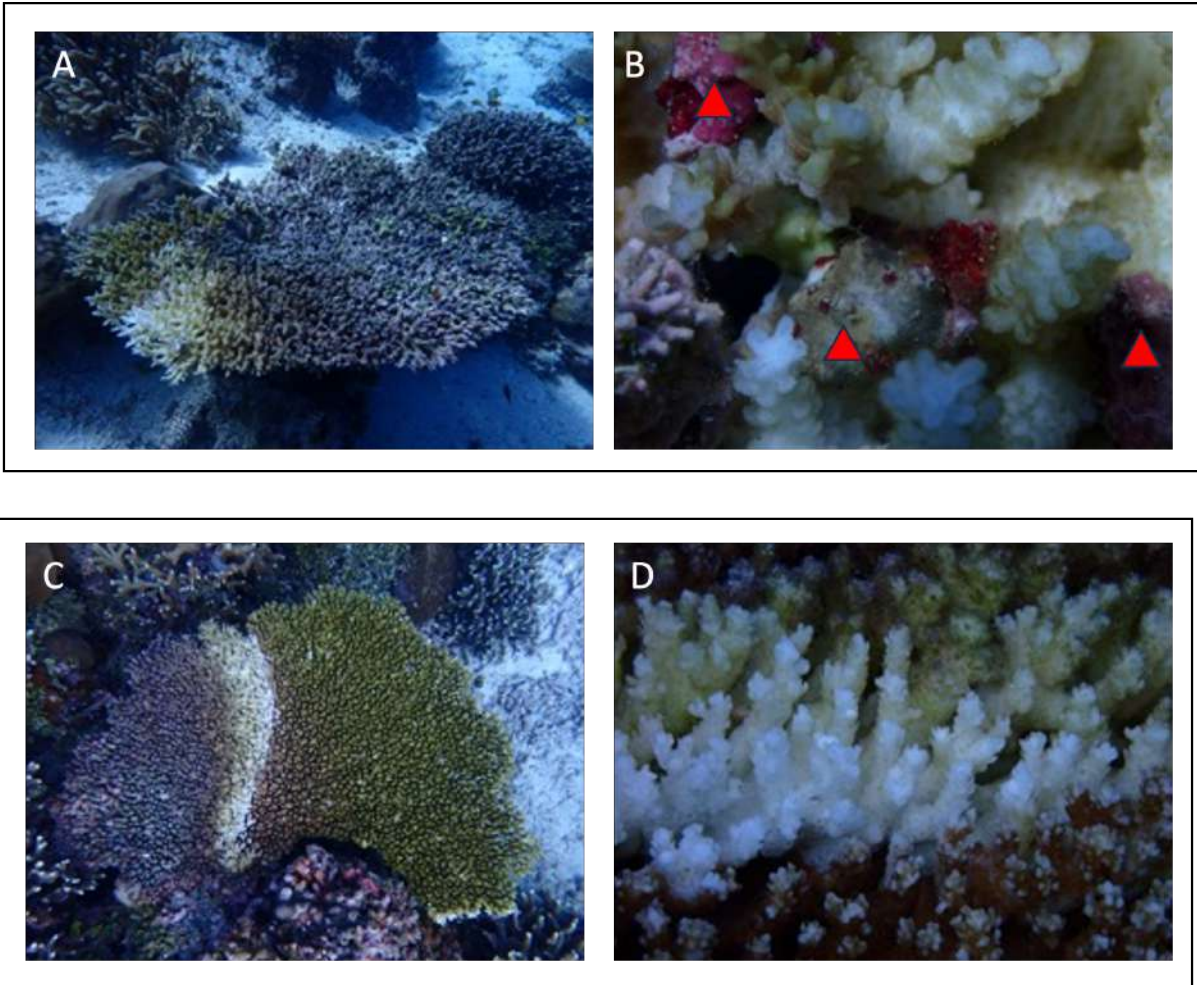


Figure S4: Example photographs illustrating how photographic annotations could be misinterpreted depending on angle, photo quality, etc. Photos A & B show a colony that is experiencing snail predation as was evidenced by examining the colony and discovering snails nestled within the branches. In photo B, visible snails are indicated with red arrowheads. Photos C&D show a colony that is most likely diseased and close-up observation revealed no snails or predators present. Photographs were taken to reflect this, but macro level photos taken of a different section of colony A could have had snails hidden and thus classified as disease or recent tissue loss.

Hard Coral					
Dead Coral	DC	Recently dead- white to dirty white			
Dead coral + algae	DCA	Favites	FAV	Oulophyllia	OPA
Acanthastrea	ACA	Fimbriaphyllia	FIM	Oxypora	OXY
Acanthophyllia	ATP	Fungia	FUN	Pachyseris	PAC
Acropora	ACR	Galaxea	GAL	Palaustrea	PAL
Alveopora	ALV	Gardineroseris	GAR	Paragoniastrea	PGA
Anacropora	ANA	Goniasatrea	GOS	Paramontastrea	PMA
Astrea	AST	Goniopora	GON	Pavona	PAV
Astreopora	ASP	Halomitra	HAL	Pectinia	PEC
Australogyra	AUG	Heliofungia	HEL	Physogyra	PHS
Australophyllia	AUP	Heliopora	HEP	Platygyra	PLA
Blastomussa	BLA	Herpolitha	HER	Plerogyra	PLE
Cantharellus	CAN	Heterocyathus	HTC	Plesiastrea	PLS
Catalaphyllia	CAT	Heteropsammia	HTP	Pleuractis	PCS
Caulastrea	CAU	Homophyllia	HOM	Podillopora	POC
Coelastrea	COA	Hydnophora	HYD	Podabacia	POD
Coeloseris	COS	Ispopora	ISO	Polyphyllia	POL
Ctenactis	CTE	Leptastrea	LER	Porites	POR
Cydoseris	CYC	Leptoria	LEP	Psammocora	PSA
Cynarina	CYN	Leptoseris	LES	Rhizopsammia	RHZ
Cyphastrea	CYP	Lithophyllon	LIT	Pseudosiderastrea	PSD
Danafungia	DAN	Lobactis	LBC	Sandalothia	SAN
Dendrophyllia	DEN	Lobophyllia	LOB	Seriatopora	SER
Dipsastrea	DPS	Merulina	MER	Stylocoeniella	SCL
Diploastrea	DIP	Micromussa	MIC	Stylophora	STP
Duncanopsammia	DUN	Montipora	MON	Trachyphyllia	TRA
Echinomorpha	EMA	Moseleya	MOS	Tubastrea	TUB
Echinopora	ENP	Mycedium	MCM	Turbinaria	TUR
Echinophyllia	EPL	Oulastrea	OLA	Zoopilus	ZOO
Euphyllia	EUP				
Other Fauna					
Soft Coral	SC			Coralline Algae	CA
Sponges	SP			Halimeda	HA
Zoanthids	ZO			Macroalgae	MA
Others	OT	(tunicates, anenomes, gorgonians, giant dams.etc)			
Algae	AA	Algal Assemblages (> 1 sp)		Turf Algae	TA
Substrate					
Sand	S	Rubble	R	Silt	SI
Rock	RCK	Water	W	(Fissures deeper than 50cm)	
Lifeforms					
Branching	CB	Massive	CM	Submassive	CS
Encrusting	CE	Foliose	CF	Mushroom	CMR
Acropora Tabular	ACT	Heliopora	CHL	Millepora	CML
Missing Data	DDD			Unidentifiable	UND

Supplementary Datasheet 1: Example Datasheet for LIT Survey Codes

Date: _____ Time: _____ Site: _____ Surveyor: _____ Sunny/Cloudy/Rainy
 Cloud: _____ Visibility: _____ Sea State: _____ Monitoring Interval: _____
Health: 1 = healthy 2: 1-25% dead 3: 26-50% dead 4: 51-75% dead 5: 76-99% dead 6: 100% dead

Genus	Tissue %	Bleaching		Predation	Disease	Boring Org.	Algal Con.	Turf Algae	CCA	Pigmentation	Sediment
		Lightest	Darkest								

Supplementary Datasheet 2: Example datasheet for in-situ annotations of all 9 attributes

4.4 Chapter 4 Annex

In this section, I will present the preliminary results of the field visit to Indonesia conducted in May 2023. These results are directly related to the fourth chapter which investigates differences in benthic coverage and coral diversity in site determination. Photographic data was presented briefly in Chapter 4. Here, I will focus on the data collected from the LIT benthic survey. Benthic surveys and methods for statistical analyses were presented or identical to those presented in Chapter 4 and details can be found there.

During the site visit in May 2023, additional sites were added within the bird's head seascape, Indonesia. Most notably around Manokwari, Indonesia which is geographically distinct from the sites around the Misool Island Region (South Raja Ampat) which were visited in October and the focus of Chapter 4 (Figure 1). With the addition of sites around Manokwari, a fourth site designation, pollution, was also introduced. Polluted sites were defined by the presence of trash or anthropogenic waste. In addition to the sites around Manokwari, additional depths were added to sampled sites in the Misool Island Region and is discussed in more detail after presentation of the results. In this subsection, I will present the preliminary results from the May 2023 visit along with a brief discussion. Due to time constraints, this analysis was not completed in time for the submission of the manuscript but will be covered in more detail in the final version of the paper in prep (Chapter 4).

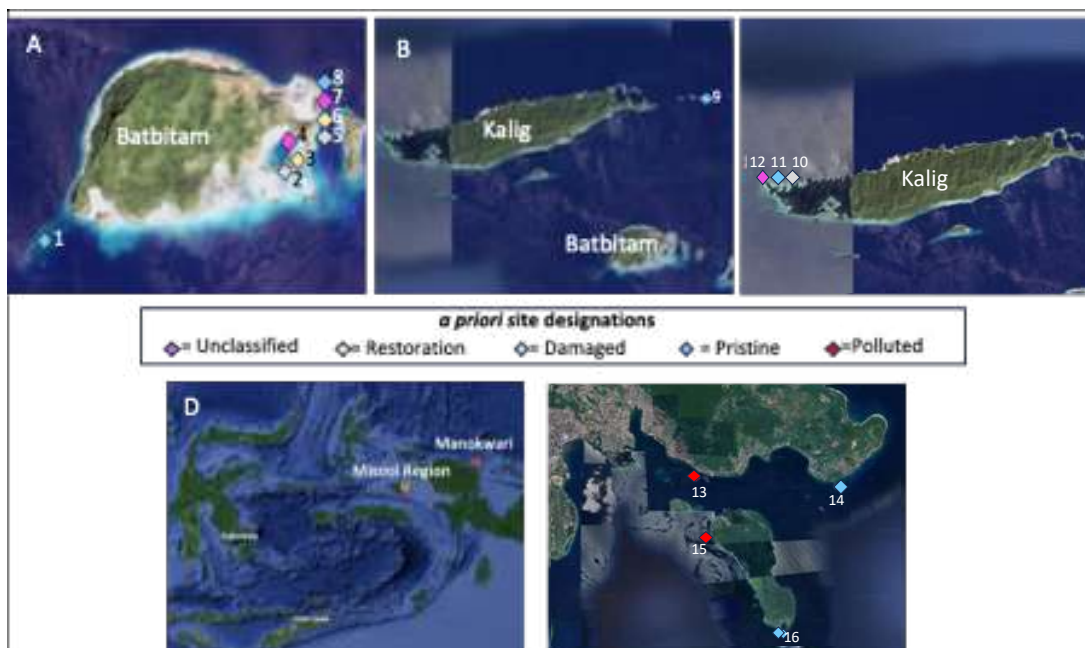


Figure 1: Maps showing site locations for May 2023 Visit (A) Sites around Misool (B) Site 9 (C) Sites around Kalig (D) Location of Manokwari and Misool (E) Sites around Manokwari

Preliminary Results:

Lit Surveys were conducted for 16 different sites with two replicates per site and certain sites having replicates at both 5m and 10m depth resulting in 42 completed LIT Surveys and 840m of reef surveyed (Figure 1).

To further flush out the differences in evaluating sites, depth needs to be considered. During October, all pristine sites were at 10m depth with restoration and damaged sites being at 5m. It has long been known that coral diversity and coverage are affected by depth with increases in depth showing increasing diversity and decreasing coverage (Huston, 1985). With certain sites, we couldn't mitigate the effects of depth, as for example, all the restoration sites were shallow and didn't have structures or active restoration at 10m depth. Another example being one pristine site which had a minimum depth of 10m (Site 09). In October, we wanted to survey all sites at 5m and 10m depth but were not able to due to time constraints budget constraints. But this was a factor that we were aware of and was considered in the planning for the field visits which occurred in May. Here, more depth points were added with the hopes of better investigating the effects of depth on how benthic categories and diversity affect site type.

Here, preliminary analyses proved to be much more complicated than the initial visit in October 2022 due to the dual depth points at certain sites, and the addition of the sites at Manokwari which although part of the Birds Head Seascape region were geographically distinct from the sites in the Misool Island Region (Raja Ampat). Thus, attached you can find data broken down similarly to how it was presented in Chapter four (by count and length) but also by various site groupings. Please note that as analysis are currently underway, the ideal placement for the unclassified sites wasn't determined so these sites have been included on MFA plots and are labeled as "intermediate."

1. Misool 22 – This represents the same 11 sites (22 replicates) that were visited in October 2022
2. Misool 23 – This represents all the sites sampled from the Misool island region in 2023 (Sites from October 23(11) and additional sites). Additions included Site 12 – which was a site at 10m depth on Kalig island, and replicates at 5m depth for site 01 and site 08. 12 Sites were sampled with 28 replicates.
3. Manokwari – This includes the four sites sampled at Manokwari (12 replicates total with two of the four sites sampled at both 5m and 10m depth)
4. All – All sites, (Manokwari and Misool) 16 sites total with 40 replicates.

Attached, you will find the dendrograms for count and length for all of the above groupings as well as MFA plots and barplots explaining what variables contributed to what dimension for each grouping for both count and length data. First, we analyzed all the sites together, with the algorithm on its own producing more than five clusters. To better understand the data, we selected five as the number of clusters as we didn't want to move the "unclassified" sites into other clusters at this point in the analysis. Thus, the five clusters were ideally for Pristine, Polluted, Restoration, Damaged, and Intermediate ("unclassified"). In order to try to better understand how sites were cluttered together, we broke down the analysis into the above groupings. In the other groupings, clustering was done based on what sites were part of that grouping *i.e.*, Manokwari had two types of sites and thus two clusters (pristine and polluted). Quick preliminary analysis seems to reveal that the length data for the Misool 22 was more similar to the groupings we found in October and presented in chapter four than the count data for the Misool 22 groupings. As a reminder, this grouping (Misool 22) was the same sites that

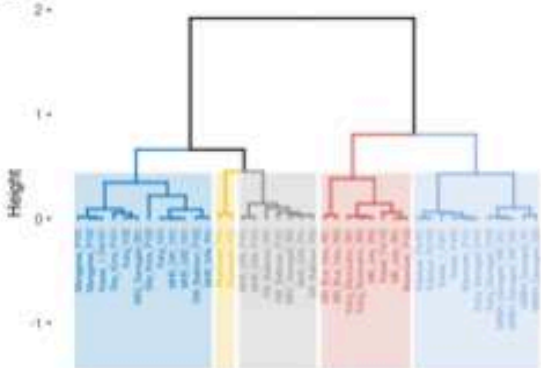
were surveyed in October. As the pristine sites typically tended to be the most robust and consistent in grouping together, they apart from one were all clustered together in the May 23 length data.

Misool 22 length data: Here, substrate, mainly rubble and fauna, soft coral and sponge contributed the most to the first dimension with hard coral (*Acropora*) and algae (turf algae) contributing the most to the second dimension. Again, with the MFA we see that pristine and restoration sites are easily distinguished with two very distinct groupings. Again, rock and Porites seem to be important factors in defining pristine sites.

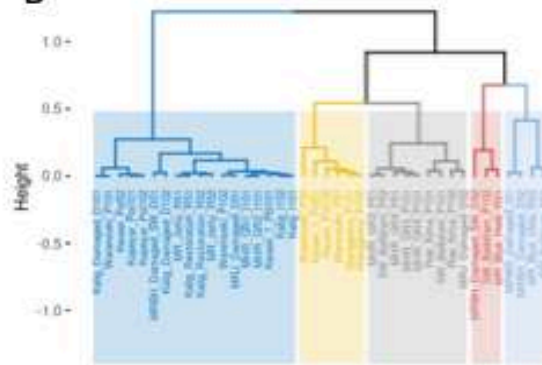
Misool 22 count data: Substrate (*Acropora* and rock) and fauna (sponge, other and soft coral) were the most important predictors for the first dimension with *Acropora* and CCA also playing an important role. Substrate, distinguished by rubble but also sand and rock was the biggest contributor for the second dimension with algae (turf algae, and CCA) also playing a role. Here once again we see two very distinct clusters between restoration and pristine. Here the clustering seems to respect the *a priori* site designations with all the pristine sites grouped together plus one intermediate site and one damaged site (Blue), all the damaged sites grouped together plus one restoration replicate (Yellow) and all the other restoration sites grouped together (gray). This is also a similar pattern to what we saw in October with the Count data reflecting the *a priori* site designations.

Cluster 1 = Restoration
 Cluster 2 = Damaged
 Cluster 3 = Pristine

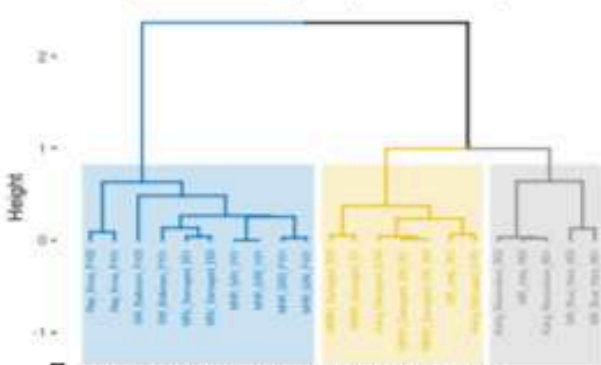
A Dendrogram Count May 23 (All sites)



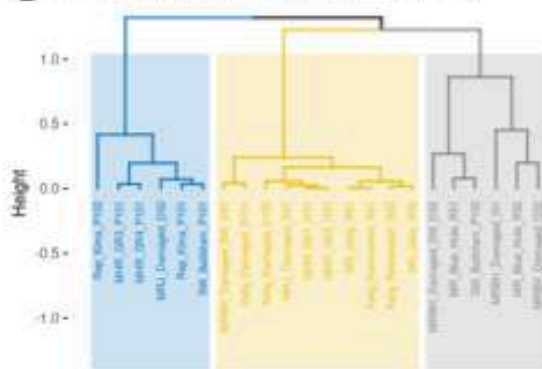
B Dendrogram Length May 23 (All sites)



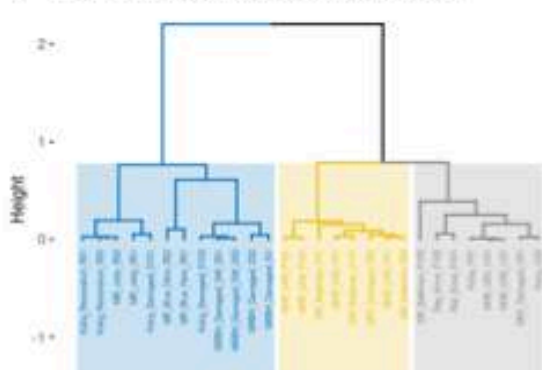
C Dendrogram Count May 23 (Misool site 22)



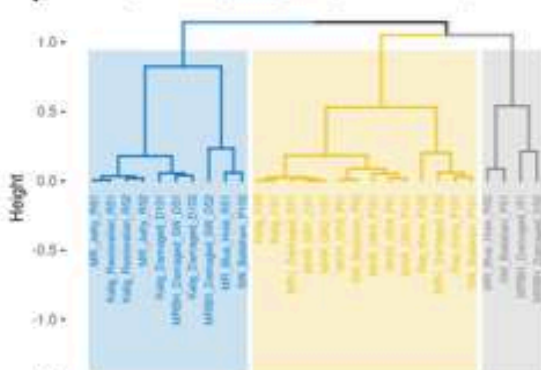
D Dendrogram Length May 23 (Misool site 22)



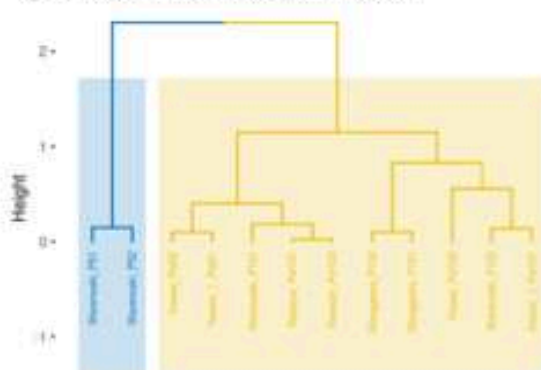
E Dendrogram Count May 23 (Misool site 23)



F Dendrogram Length May 23 (Misool site 23)



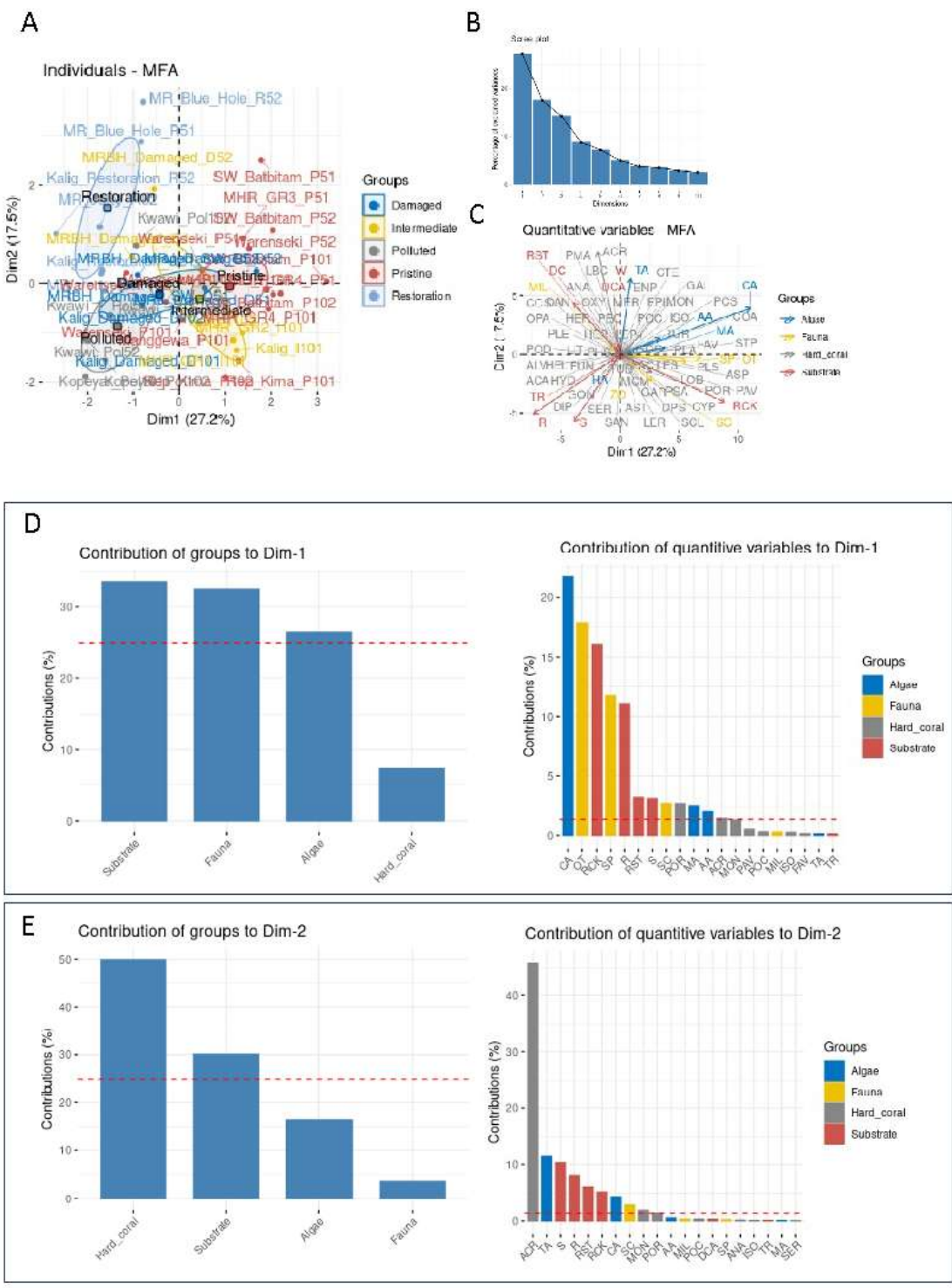
G Dendrogram Count Manakwari May 23



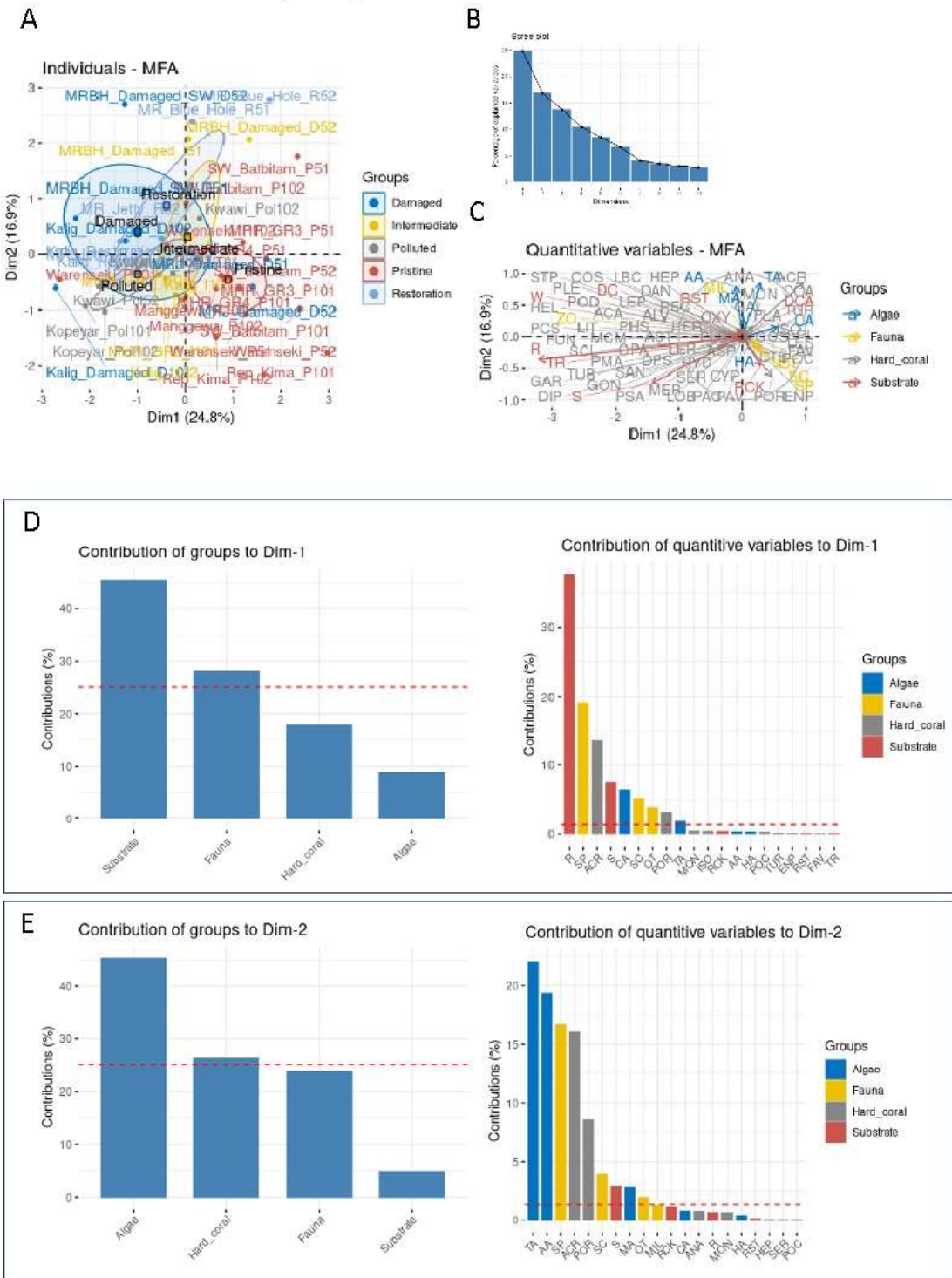
H Dendrogram Length Manakwari May 23



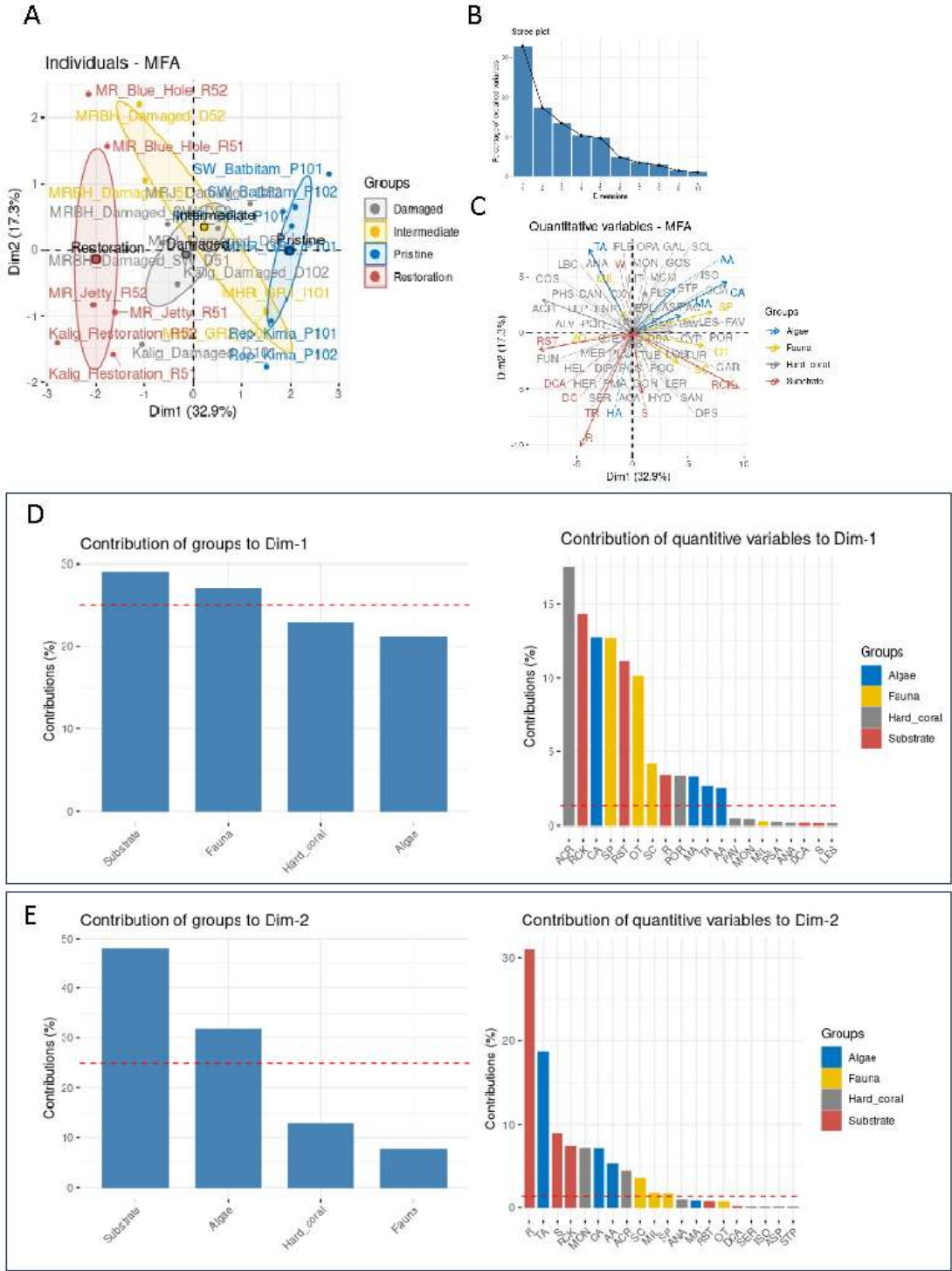
May Count Data – All Sites



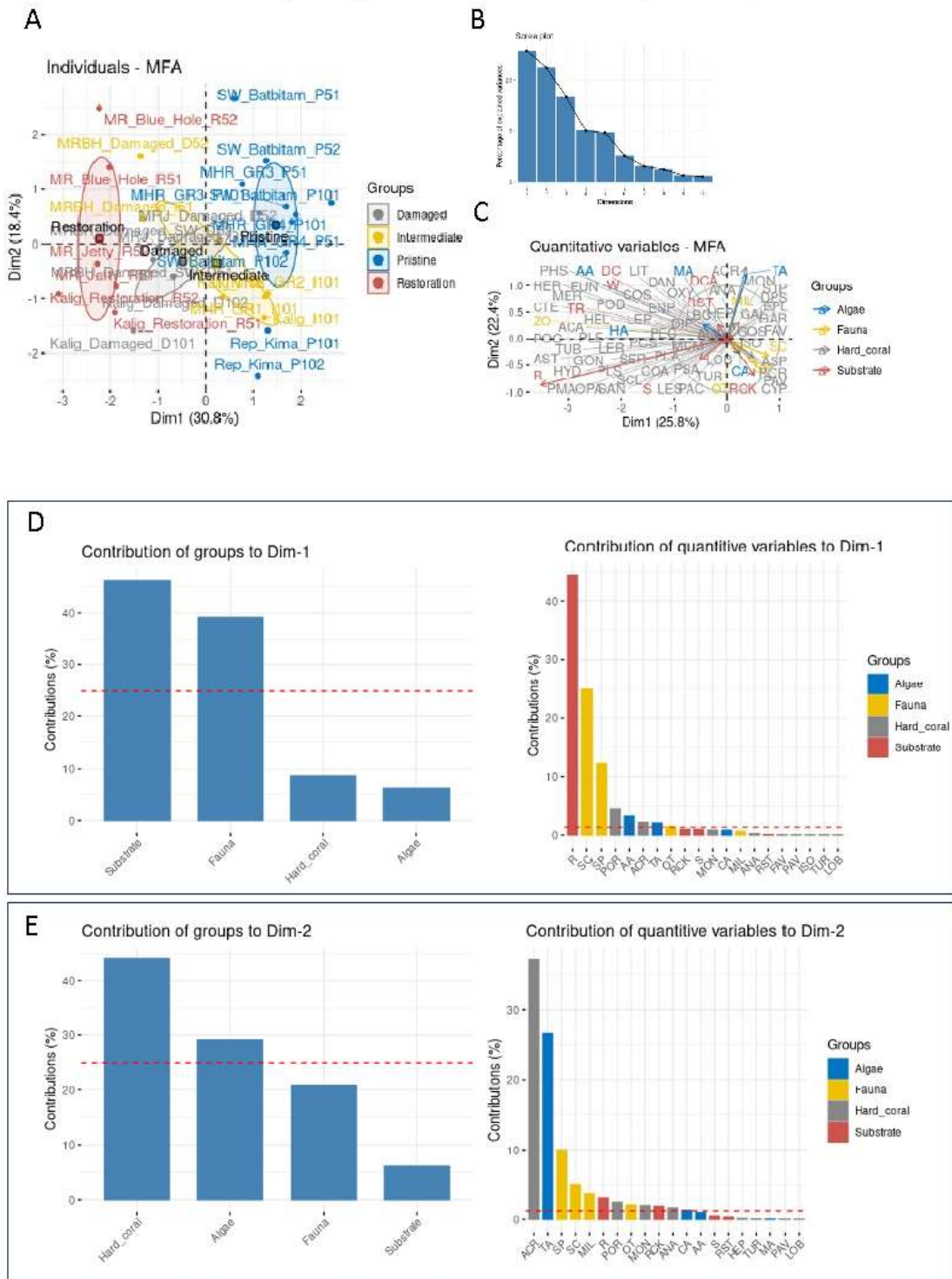
May Length Data – All Sites



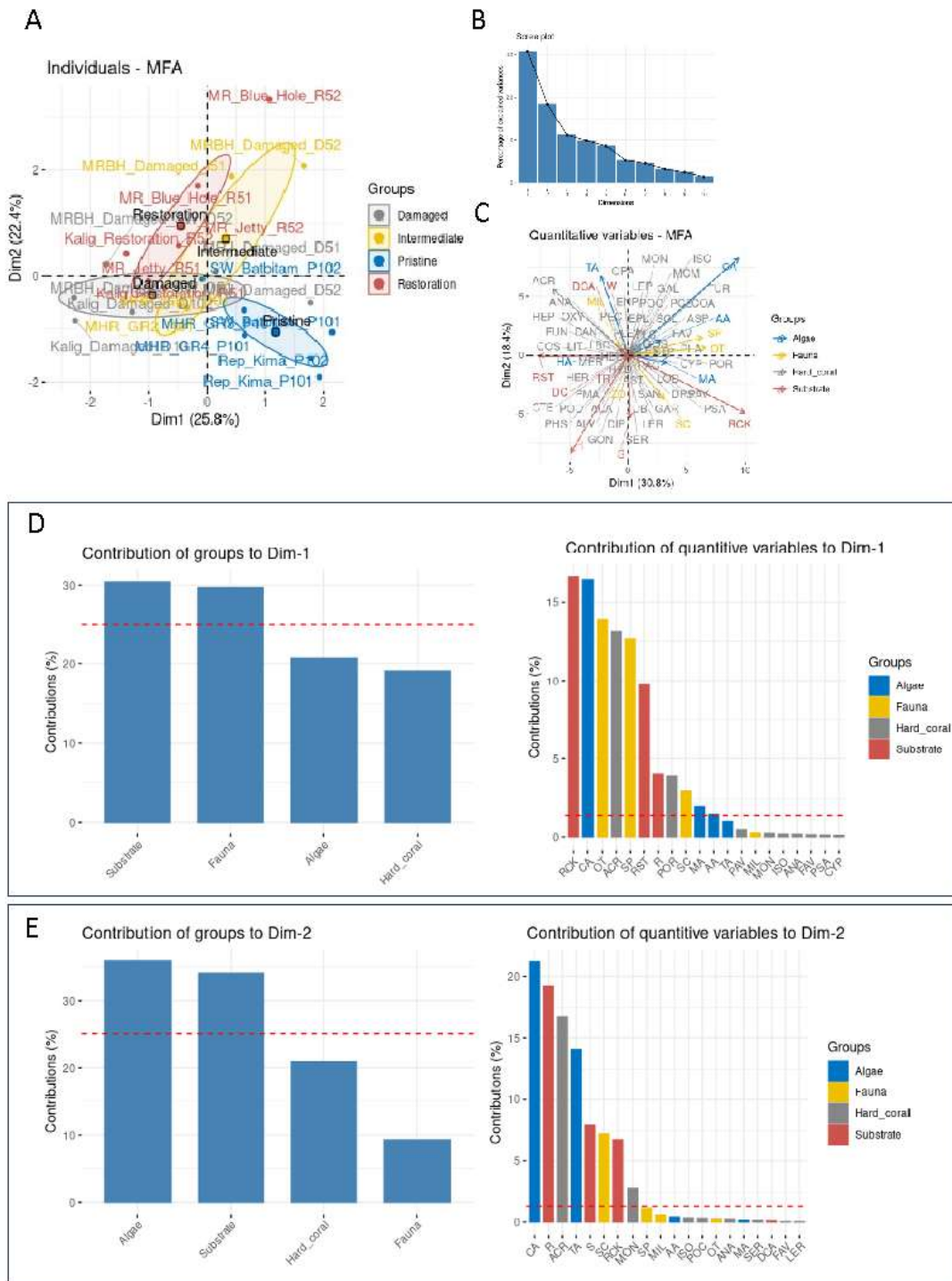
May Count data for Misool 22 (22 sites)



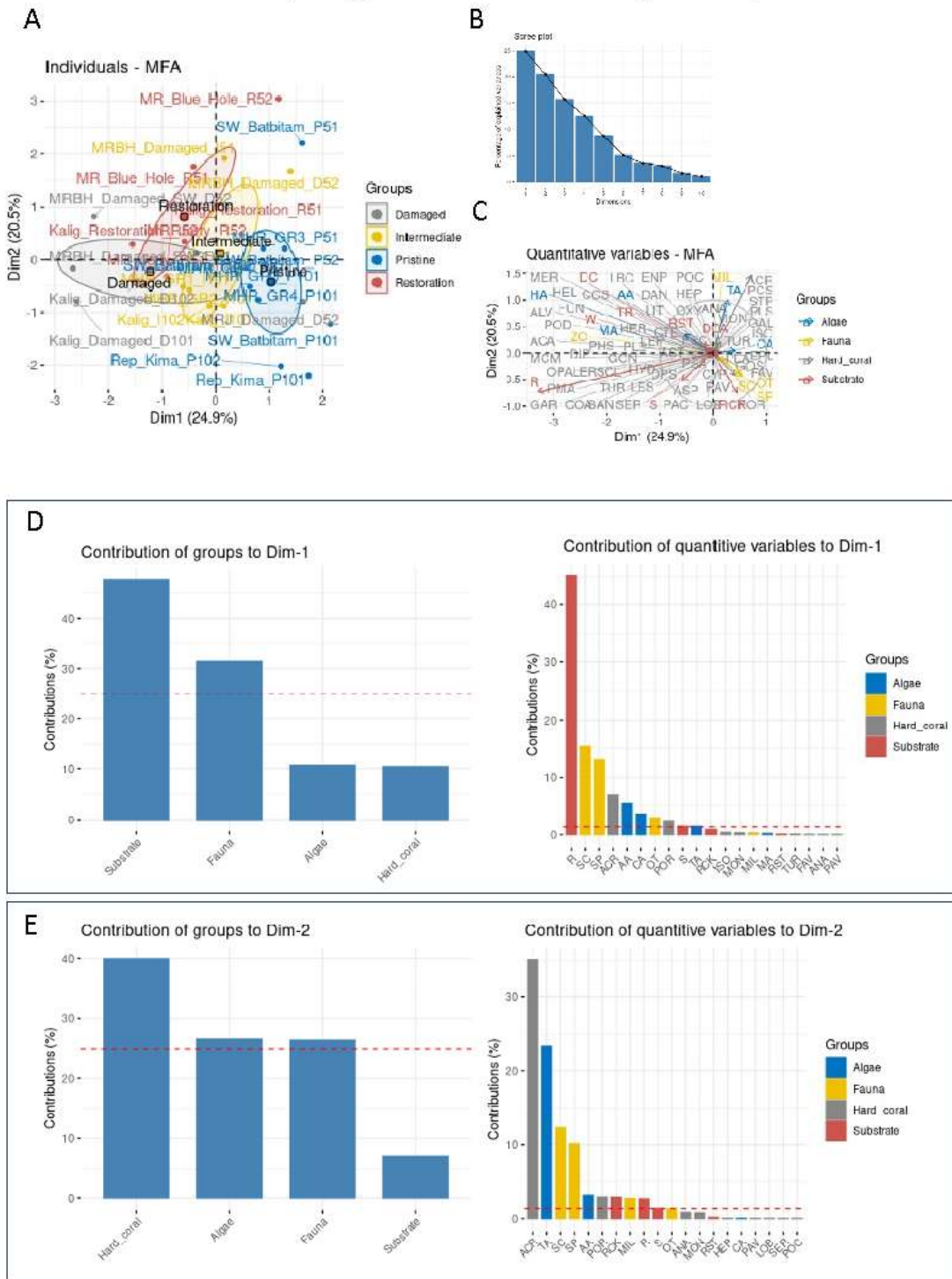
May Length data for Misool 22 (22 sites)



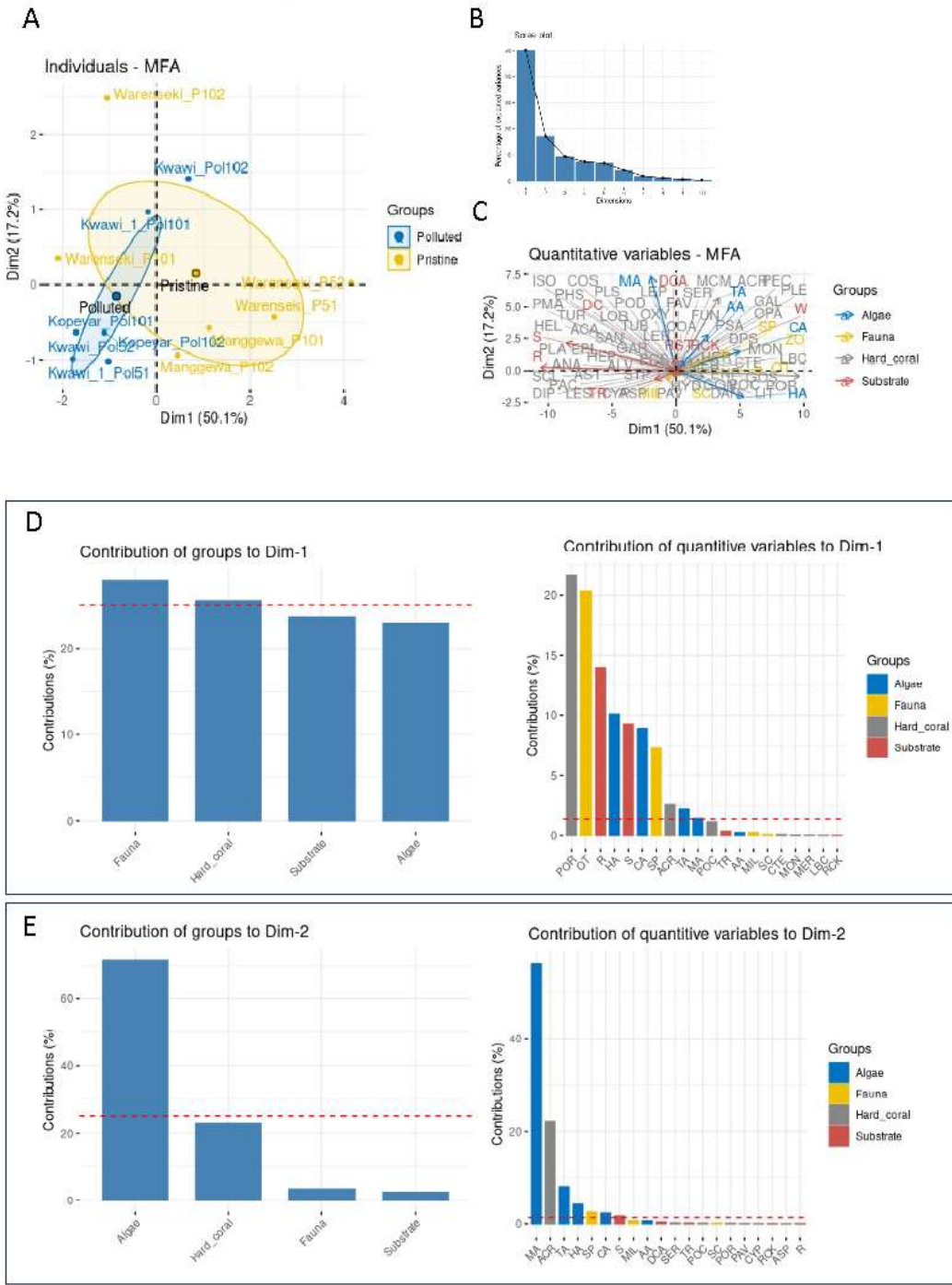
May Count data for Misool 23 (28 sites)



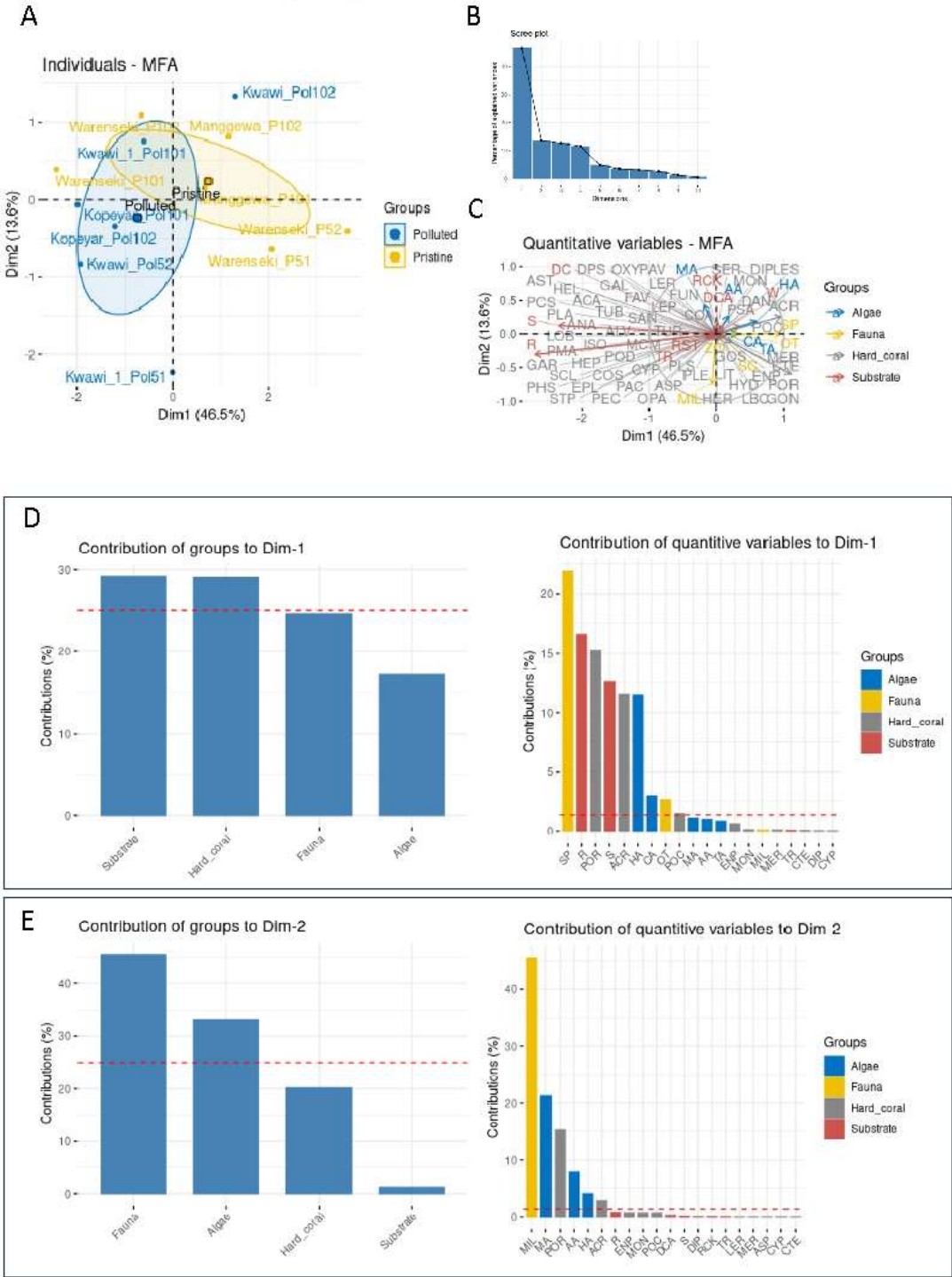
May Length data for Misool 23 (28 sites)



May Count data for Manokwari



May Length data for Manokwari



5 Conclusions & Perspectives

5.1 Conclusions

As presented in this manuscript, coral reefs are incredible marine ecosystems that harbor extraordinary biodiversity and provide essential ecosystem services while also having pronounced cultural, scientific and educational value (Spalding et al., 2001) (Gonzalez-Rivero et al., 2014). (Gómez-Rios et al., 2018) (Neal et al., 2015) (Haas et al., 2015) (Spalding et al., 2001). Furthermore, they remain one of the most beautiful and aesthetically pleasing ecosystems on the planet (personal opinion). However, despite their incredible value, they are severely threatened by anthropogenic activities and climate change with the IPCC forecasting their eventual collapse if greenhouse gas emissions are not mitigated (Hoegh-Guldberg et al., 2018) (Souter et al., 2020).

Hard corals, are the ecosystem engineers behind coral reefs, and are responsible for building the complex 3-D topography necessary for other marine species to thrive with coral reefs supporting more than 25% of marine life (Graham & Nash, 2013) (Souter et al., 2020). Various factors such as predation, nutrient levels, light availability, depth, temperature, salinity, water flow, and substrate type affect the survival of these animals. A comprehensive approach that investigates the physiological state of the animal, environmental conditions, and interactions with other organisms over time is needed to assess their health.

Through technological advancements such as scuba diving and photography, our understanding and knowledge about marine life has greatly increased. Scuba diving has enabled direct access to previously unexplored underwater habitats and plays a significant role in scientific data collection whilst underwater photography has become more accessible, allowing for faster, non-invasive methodology, creating a permanent record (Witman et al., 2013) (Todd, Sanderson & Chou 2001). However, this has resulted in a large number of generated images which require a huge time commitment for manual annotation often subject to bias and inconsistencies. AI, particularly machine learning and deep learning, has the potential to automate image annotation and analysis, improving the speed and accuracy of data processing while reducing costs (Gonzales-Rivero et al., 2020). By combining scuba diving, photography, and AI, researchers can provide useful tools for coral reef conservation, enabling faster and more efficient monitoring, analysis, and assessment of coral health. Such technology

is crucial due to the current degradation of these ecosystems and the need for standardized monitoring protocols that are rapid and efficient. However, more work is needed to address challenges specific to coral reef datasets for further advancements in AI based coral reef research.

The principal aim of my PhD was to investigate **how artificial intelligence tools can be used to assess coral health states from colony photographs**. This novel exploration of individual colony health via an artificial intelligence approach used photographic data from the Tara Pacific Expedition as well as novel data generated from two field studies (one in Moorea, French Polynesia, and another in The Bird's Head Seascape, Indonesia). Results were presented in the preceding chapters.

In my second chapter (Lombard et al., 2023), I presented the Tara Pacific Expedition, with a focus on the photographic dataset and manual annotations that were generated during this expedition. The manual annotations consisted of per photograph annotations of individual coral colonies across 9 physiological hallmarks: Predation, Bleaching, Boring Organisms, Algal Contact, Sediment Contact, Tissue Appearance, CCA, Turf Algae, and Pigmentation. These annotations created a dataset that provides additional biological information about the coral colonies that was not collected during the expedition, and is also the foundation of our AI machines. However, the completion of these annotations was very time consuming.

Thus, to speed up future annotations, an AI approach was implemented and presented in the third chapter (Clampitt et al(a), in prep). Here, I presented the successful creation of AI machines capable of categorizing the nine physiological hallmarks as absent or present in photographs of coral colonies. The Tara Pacific photographic dataset was not taken with an AI purpose in mind and therefore the developed machines were highly catered to these photographs through the use of transfer learning and data augmentation. As a result, the machines performed extremely well on datasets generated from the Tara Pacific dataset with F1 scores averaging 83% across all nine attributes. However, they were less performant on novel photographs highlighting the need for strict photographic methodology and increased training samples.

Hence, in the fourth chapter (Clampitt et al(b), in prep) not only do I present a novel photographic methodology aimed at improving machine performance, but I also explored the

relationship between the physiological hallmarks and the type of site / current state of the reef. Ecological data about benthic composition and coral diversity was collected by using LIT surveys which revealed that hard coral biodiversity may be a good predictor of site type, and could lead to simplified monitoring efforts if this trend proves to be robust. Novel methodology which focused on macro level photography was also introduced which resulted in twenty photos per colony and underwater annotations for all 9 physiological hallmarks. When testing macro and colony level photos on our machines (Clampitt et al(a), in prep) we saw higher F1 scores when using the colony level photo indicating the macro level photos may be more different from the tara pacific photographic training set than the colony photographs. Finally, we analyzed how visual cues relate to site type and although visual cues / physiological attributes were not found to be predictors of site type, methodological changes could reveal relationships.

Thus, through these various studies, I have achieved my PhD objective and laid down the groundwork to further investigate how artificial intelligence tools can be used to assess coral health states from colony photographs. I successfully created AI machines capable of discerning 9 visual attributes that are relevant to coral colony health. This non-invasive tool is the first step in developing machines capable of categorizing the physiological state of the colony or even overall reef health. Additionally, I have set-up studies to further respond to the question of defining coral health while also collecting and annotating a large photographic dataset to improve upon and build further machines.

5.2 Discussion & Perspectives

In coral-related research using AI approaches, identification and categorizing the benthos has been one of the primary areas of study (González-Rivero et al., 2014)(Chirayath & Instrella, 2019) (Hopkinson et al., 2020) (Purser et al., 2009)(Bejibom et al., 2012) (Zhang, 2015)(González-Rivero et al., 2020). AI-based automatic annotations have shown high accuracy in determining benthic composition, with comparable results to manual annotations and long-term monitoring programs (González-Rivero et al., 2014)(González-Rivero et al., 2020). Machine learning has been successfully utilized to calculate percent coral cover and categorize benthos into functional groups, such as branching coral, mounding coral, rock, and sand (Chirayath & Instrella, 2019). Furthermore, AI has been combined with remote sensing for habitat mapping (Hamylton et al., 2020)More recently, studies have become more

diversified, further pushing the boundaries of possibilities. For example, A 2022 study (Scucchia et al.) used deep learning to study coral biomineralization and was the first study to use AI to analyze the internal skeletal network of reef-building corals.

However, to my knowledge, very few studies have investigated health-related assessments of individual coral colonies using an artificial intelligence approach. Here again, we find the literature to be biased towards bleaching. Borban et al., (2021) used AI to categorize photographs of colonies into three classes: bleached, dead, and healthy, achieving an accuracy of 85% but with limited sample sizes. Similarly, AI was used to successfully predict bleaching susceptibility of the massive colony, *Orbicella faveolata*, based on its protein signatures although the author stresses that achieving this result was both expensive and time-consuming and would not be a practical approach across all species (Mayfield & Lin, 2023). In 2021, in the context of a restoration project, AI machines were developed to identify coral fragments of *Pocillopora* sp. and *Acropora* sp. transplanted onto restoration structures and to measure their growth and survival rates significantly reducing monitoring time (Morand, Dixon, & Le Berre, 2021). The only article using a multifactorial approach to analyze health was Mayfield et al., (2022) where machine learning models analyzed environmental, biological, and benthic survey data to successfully identify (~85%) climate resilient colonies of *Pocillopora Acuta*. Overall, studies focused on health-related assessments of coral colonies, apart from bleaching, are limited.

Thus, the works presented in this manuscript represent the first approach which attempts to investigate multiple visual cues that are relevant to individual colony health. As far as I'm aware, bleaching and disease are the only attributes of the nine visual cues described in Chapter 3 (Clampitt et al(a), in prep) that have been investigated using an AI approach. No studies have used AI to determine predation, algal contact, sedimentation, pigmentation, CCA, turf overgrowth or boring organisms. As briefly presented in the introduction and further developed in chapter 3, these attributes can provide information about the current state of individual colonies and certain can be used as proxies providing information about the state of the reef. Here, we have the unique opportunity to combine these attributes with the entire TP dataset (Microbiome, biomarkers, telomere, environmental...etc) to gain additional information on a) the physiological status of the corals during the time of sampling and b) investigate how the nine attributes are correlated to the physiological state of an individual colony. Using this

information, we can provide a baseline for longitudinal studies, such as the one implemented in Moorea to determine coral colony health.

In terms of AI, the next steps will be to continue data collection while also integrating the various datasets we have collected throughout the duration of this PhD (Moorea & Indonesia) into training sets. As we now have three datasets with: Tara Pacific, Moorea Longitudinal, and Indonesia - we plan on building machines that have been trained on different combinations of these datasets to create a machine with high performance power on novel images. As, Moorea and Indonesia both consist of in-situ annotations, it would be interesting to see how a machine trained with all three photographic datasets compares to a machine trained on Moorea and Indonesia alone as well as to the original machines trained on the tara dataset. This is due to the difference between inwater and photographic annotations, discussed in more detail in perspectives below.

Furthermore, we have collected data that has yet to be exploited. For instance, for all of our photographs (N=20,049) we have information about the percentage of living tissue on a six point scale. This could be used to create a machine that can classify colonies by percent of living tissue which would facilitate monitoring of individual colonies. Additionally, all underwater annotations for bleaching have been conducted using the coralwatch color card which also has a 6-point scale. As to date, we have been transforming this into a binary scale. But, plan on investigating a machine capable of capturing more subtle differences between colonies. We also have data on the type of predation (snail vs fish vs multiple) that has yet to be explored. Finally, ideally we would like to explore developing a machine that could classify reef state based on coral colony photographs. This was the idea behind chapter four.

Furthermore, ideally, our AI machines would incorporate more genera or make machines based on lifeforms so it would be more general and applicable to any colony. As of right now, we have two branching corals (Acropora and Pocillopora) and one massive coral (Porites). Overall, in the scope of this project and our dataset alone, there are many future research routes that would add not only to this work but also to the field of coral science.

Furthermore, a huge feat of this PhD Project is the photographic datasets and annotations that were generated throughout the life of the project as detailed below:

Photo Datasets Generated:

1. **Tara Pacific Dataset (Lomard et al, 2023):** 1249 Millepora, 2207 Porites, 2099 Acropora
 1. 5555 photographs
 2. Photographic annotations comprising all 9 visual cues
2. **Moorea Longitudinal (Clampitt et al(a), in prep) :** 30 Acropora, 30 Pocillopora, 30 Porites colonies
 1. 327 photos : 3 photos per colony
 2. In-situ annotations: Tissue Appearance, Bleaching, Predation, & Percentage of living tissue
3. **Indonesia (Clampitt et al(b), in prep)**
 1. May 2022 : 100 Acropora, 100 Pocillopora, 100 Porites, 2 Millepora
 1. 890 photos : 3 photos per colony
 2. October 2022 : 435 Acropora, 236 Pocillopora, 224 Porites
 1. 2557 photographs: 3 photos per colony
 3. May 2023: 176 Acropora, 170 Pocillopora, 190 Porites
 1. 10,720 photographs: 20 photos per colony

Finally, once the final machines are developed and the images analyzed, we will transform these machines into accessible tools available for the whole coral community. Such a tool would allow for a non-invasive, unbiased, replicable monitoring system that could reduce sampling time, analysis and costs. Such a machine would eliminate the need for a coral biologist which is important as often the places with the most coral reefs have the fewest resources. Sampling would be simple and efficient, only requiring a photograph with the added benefit of producing a permanent record that could be referred to in the future and would allow for easy comparison between projects and reef localities. This technology could be used by scientists, conservationists, restoration practitioners Restoration projects, curious divers, tropical resorts or even for educational purposes.

Additionally, I would like to highlight the success of these machines. As someone who is intimately familiar with the Tara Pacific Photographic dataset, I was highly doubtful that we would be able to build functional machines from it. The dataset, on paper, is an excellent idea but creating a photographic dataset without a clear, scientific protocol, with data collection spanning two-years, constantly changing teams of researchers/divers, and generation of 1000's of samples is bound to result in a non cohesive dataset. Annotating the dataset alone was time consuming and tedious. Often, I had to zoom in on photographs, spending several minutes on one attribute to try to determine what something was. Identification was made extra challenging with ever changing angles, shadows, quadrat placement, distance between the cameras and the colonies, and lighting. Furthermore, sometimes colonies were out of focus, water was not clear, or photo quality was not ideal. I say all this not to depreciate this dataset but rather to marvel

at the ability of the Artificial Intelligence Machines we developed and to highlight that with continued research, aimed at increasing photographic samples with proper photographic methodology and annotation procedures could result in truly remarkable machines.

In terms of the Tara dataset, I would also like to comment on the photographic cleaning process. The cleaning process was a huge task that gets overlooked in my opinion, in the scope of this research project. The majority of the photographic cleaning was done by using the annotations which prompted me not only to determine if a photograph was useful and contained the target colony but also had a section for notes where I left detailed notes about photo quality and other potentially interesting environmental variables that might be visible in the photo. Additionally, we used a specifically developed database to further whittle down the dataset. Thus, through cleaning we were able to transform this muddled dataset into a much more concise one suited for our purposes (aka using AI). I think it's important to note that had this cleaning process not been implemented, we would have seen much worse performance on the machines. However, I believe the dataset could still be further refined. For example, some of the wide-angle photos that were kept may be too far away or some of the closeups may not be sharp enough. etc. It would be interesting to see if further cleaning could raise performance. However, since cleaning vastly reduced the samples (11,470 -> 5555), that meant that we needed to use transfer learning and data augmentation. We spent lots of time fine-tuning this dataset which was not generated for AI purposes and uses AI tools to make it work. This helps explain why the machines are highly specific to the dataset and again makes future possibility so exciting as the data that we have started collecting was generated with an AI purpose in mind.

Another important overarching remark of this project concerns the photographic methodology and the *in-situ* annotation. Firstly, the photographic methodology we started using for our field studies (Moorea & Indonesia October 2022) was based more so on the theoretical procedure than the actual practical application of the methodology. For example, colonies typically had multiple photos taken (as discussed in chapter 3), but these photos didn't necessarily follow the one close up and one colony protocol. The actuality of the photographic dataset was thus, very different from the theoretical procedure of 2 different photos per colony (*i.e.*, a colony and a macro) (as shown in Chapter 3 with photo numbers per site and average photos per colony). Very few colonies had a macro, a colony and a wide angle photo taken of good quality. The three types of photo were added to the annotations as a means to further refine and distinguish between the photos during the annotations (the annotation prompt we added was

this a wide angle, standard (colony) or close-up). This annotation was decided after doing a first test round of annotations to discern variables that we could extract, as this was not part of the naming protocol or in any type of log. Thus, our photographic methodology implemented during our field studies although based on the Tara Pacific Expedition, was essentially a completely different methodology.

Similarly, it is important to discuss the difference between annotations that were done *a posteriori* on the photographs and annotations that were done *in-situ* (this is mentioned briefly in Chapter 4). This is a very important distinction that I would like to highlight in more detail here. The machines were all trained on *a posteriori* photographic annotations whereas any novel test sets we ran, were on photographs with in-water annotations. The best example of this would be the annotations for the attribute bleaching annotations. For me, the bleaching attribute as annotated via the photographs is extremely biased because everyone sees color differently, light availability and color processing changes with depths (which were not noted or analyzed), and effects of white balance. This means that a photo at depth with more muted colors might be annotated as normal when in fact it was dark, or a colony that is from a blown-out image where the white balance is off would have to be annotated as bleached when in reality that is just due to the image quality. Thus, I made sure that bleaching was included in the in-situ annotations as a means to improve these annotations and make less biased, more accurate machines. Interestingly, we did see a correlation between the “dark” annotation and symbiont biomass for *Pocillopra* which is because the dark category is probably the least likely to be affected by white balance and depth as it should still appear relatively dark. Again, when you take the difference in annotation into account along with the fact that the methodology was essentially completely different, it further highlights how astonishing these machines are.

As a further detail on in-water annotations, the reason only three of the attributes were included at first for underwater annotations was essentially due to trying to balance timing constraints and maximizing the photographic samples. For AI, the more photographs the better. Thus, we knew that the more categories we had, the longer it would take to annotate a single colony and the less colonies we would be able to annotate, meaning less photographs. Furthermore, when using scuba, bottom time is obviously limited. Thus, to start off – we kept only three underwater annotation categories: Bleaching, Predation and Tissue Appearance. Bleaching was chosen to combat bias introduced by *a posteriori* annotations, but also because it is one of the most widely used markers of coral health. Tissue Appearance annotated as “disease or recently losing tissue”

was also chosen as these are hot topics in the literature that are relevant to current health state. Finally, predation was chosen as it is a relatively easy category to annotate visually (specifically for Porites colonies) this was further specialized by annotating the type of predation (snail, fish or multiple) which should be developed in future studies. We also included percent of living tissue on a six-point scale which has not yet been trained or tested as a machine but is also a future avenue of this research. That information would be especially pertinent for the Moorea longitudinal studies where colonies will be tracked over time. With this attribute, we would be able to track changes in living tissue from one year to the next. This would also be of particular interest for restoration monitoring projects. Often, these projects use photographic analysis to keep track of mortality or number of structures/lengths of reefs restored. Having an easy way to evaluate how individual colonies are surviving over time could help guide these projects and be an indicator of stressors. This aspect was not sufficiently explored in the function of this PhD and will require continued data collection.

In Chapter four, I also touched on the difference between qualitative vs quantitative data collection. Seeing as these were foundational studies starting to investigate how artificial intelligence can be used to monitor the current state of the colony, it only seems logical that the first point of investigation was simple, i.e. absence or presence. Now that we have seen that this method can be successful, we hope to improve upon these machines to make them more general and less specific to the Tara Pacific dataset by incorporating the new images from our field studies into training. Once those machines have been created, the next step would be to explore ways to exploit each of the individual attributes in order to not only optimize the machine but to optimize the information that can be extracted from a photograph. In my opinion, this would involve the inclusion of quantitative information such as the number of boring organisms in a colony or the percent of the colony that is bleached. Such machines would require different AI processes than what we are currently using, such as segmentation.

In regards to the AI machines, there is no end of possibilities to how they could be tweaked/played with to see how that affects performance. For example, a different CNN could be used for transfer learning, the number of fully connected layers could be changed, tuning of the hyperparameters, changing the percentage of dropout, etc etc. One thing that could be envisioned using an underwater dataset as well for transfer learning instead of imagenet as this would affect the initial weights and perhaps be more pertinent for our purposes. There is so

much possibility for exploration in regards to this factor, without even taking into account all the various versions of the machines we could create with the different datasets that we have.

I would also like to touch on the topic of reef health vs coral health. I present coral health in the introduction as complex and multifactorial. Reef health is much the same with location, depth, environmental factors, anthropogenic influences and other factors all directly effecting the overall state of the reef. Furthermore, like coral health, reef health is also rarely properly defined and often based on coral cover alone. With shifting baselines, and the fact that we have not been studying marine ecosystems as long as terrestrial ecosystems, it has been stated that we actually have no baseline for reefs unaffected by human influence (Knowlton & Jackson, 2008). Which brings up an interesting question of how we can really define and measure a healthy reef. We still truly don't understand all the complex, interwoven interactions that occur on coral reef. Thus, like coral health, it is essential that we use precise objectives with specified metrics along with careful vocabulary when studying or discussing reef health

6 References

- Abigail, B., Patil, P., Kleisner, K., Radar, D., Viridin, J., & Basurto, X. (2018). *Contribution of Fisheries to Food and Nutrition Security: Current Knowledge, Policy, and Research*. <http://nicholasinstitute.duke.edu/publication>
- Aeby, G. (2003). Corals in the genus *Porites* are susceptible to infection by a larval trematode. *Coral Reefs*, 22, 216–216. <https://doi.org/10.1007/s00338-003-0310-9>
- Al-Horani, F. A., Hamdi, M., & Al-Rousan, S. A. (2011). *Prey Selection and Feeding Rates of Drupella cornus (Gastropoda: Muricidae) on Corals from the Jordanian Coast of the Gulf of Aqaba, Red Sea*. 4(4).
- Apprill, A., Girdhar, Y., Mooney, T. A., Hansel, C. M., Long, M. H., Liu, Y., Zhang, W. G., Kapit, J., Huguen, K., Coogan, J., & Greene, A. (2023). Toward a New Era of Coral Reef Monitoring. *Environmental Science & Technology*. <https://doi.org/10.1021/acs.est.2c05369>
- Baker, A. C., Starger, C. J., McClanahan, T. R., & Glynn, P. W. (2004). Corals' adaptive response to climate change. *Nature*, 430(7001), Article 7001. <https://doi.org/10.1038/430741a>
- Becken, S., Connolly, R., Stantic, B., Scott, N., Mandal, R., & Le, D. (n.d.). *Monitoring aesthetic value of the Great Barrier Reef by using innovative technologies and artificial intelligence*.
- Beenaerts, N., & Berghe, E. V. (2005). Comparative Study of Three Transect Methods to Assess Coral Cover, Richness and Diversity. *Western Indian Ocean Journal of Marine Science*, 4(1), Article 1. <https://doi.org/10.4314/wiojms.v4i1.28471>
- Beijbom, O., Edmunds, P. J., Roelfsema, C., Smith, J., Kline, D. I., Neal, B. P., Dunlap, M. J., Moriarty, V., Fan, T.-Y., Tan, C.-J., Chan, S., Treibitz, T., Gamst, A., Mitchell, B. G., & Kriegman, D. (2015). Towards Automated Annotation of Benthic Survey Images: Variability of Human Experts and Operational Modes of Automation. *PLOS ONE*, 10(7), e0130312. <https://doi.org/10.1371/journal.pone.0130312>
- Beijbom, O., Edmunds, P., Kline, D., Mitchell, B., & Kriegman, D. (2012). *Automated Annotation of Coral Reef Survey Images*. 1170–1177. <https://doi.org/10.1109/CVPR.2012.6247798>
- Bell, J. D., Kronen, M., Vunisea, A., Nash, W. J., Keeble, G., Demmke, A., Pontifex, S., & Andréfouët, S. (2009). Planning the use of fish for food security in the Pacific. *Marine Policy*, 33(1), 64–76. <https://doi.org/10.1016/j.marpol.2008.04.002>
- Bellwood, D. R., Hughes, T. P., Folke, C., & Nyström, M. (2004). Confronting the coral reef crisis. *Nature*, 429(6994), Article 6994. <https://doi.org/10.1038/nature02691>
- Bennett, A., Patil, P., Kleisner, K., Rader, D., Viridin, J., & Basurto, X. (n.d.). *Contribution of Fisheries to Food and Nutrition Security*.
- Bharathi, M. J. (2021). Importance of Marine Algae and Marine Derived Commercial Products. *Journal of Molecular Microbiology*, 5(3). <https://www.imedpub.com/abstract/importance-of-marine-algae-and-marine-derived-commercial-products-41132.html>

- Bircher, J. (2005). Towards a Dynamic Definition of Health and Disease. *Medicine, Health Care and Philosophy*, 8(3), 335–341. <https://doi.org/10.1007/s11019-005-0538-y>
- Blackall, L. L., Wilson, B., & van Oppen, M. J. H. (2015). Coral—The world’s most diverse symbiotic ecosystem. *Molecular Ecology*, 24(21), 5330–5347. <https://doi.org/10.1111/mec.13400>
- Borbon, J., Javier, J., Llamado, J., Dadios, E., & Billones, R. K. (2021). *Coral Health Identification using Image Classification and Convolutional Neural Networks* (p. 6). <https://doi.org/10.1109/HNICEM54116.2021.9731905>
- Bourne, D. G., Morrow, K. M., & Webster, N. S. (2016). Insights into the Coral Microbiome: Underpinning the Health and Resilience of Reef Ecosystems. *Annual Review of Microbiology*, 70(1), 317–340. <https://doi.org/10.1146/annurev-micro-102215-095440>
- Briggs, J. C., & Bowen, B. W. (2013). Marine shelf habitat: Biogeography and evolution. *Journal of Biogeography*, 40(6), 1023–1035. <https://doi.org/10.1111/jbi.12082>
- Brownlee, J. (2020). *Imbalanced Classification with Python: Better Metrics, Balance Skewed Classes, Cost-Sensitive Learning*. Machine Learning Mastery.
- Bruckner, A. W. (2002). Life-Saving Products from Coral Reefs. *Issues in Science and Technology*, 18(3), 39–44.
- Bruno, J., & Bertness, M. (2001). Habitat modification and facilitation in benthic marine communities. In *Marine Community Ecology* (Vol. 413, pp. 201–218).
- Bruno, J. F., & Selig, E. R. (2007). Regional Decline of Coral Cover in the Indo-Pacific: Timing, Extent, and Subregional Comparisons. *PLOS ONE*, 2(8), e711. <https://doi.org/10.1371/journal.pone.0000711>
- Buchanan, B. G. (2005). A (Very) Brief History of Artificial Intelligence. *AI Magazine*, 26(4), Article 4. <https://doi.org/10.1609/aimag.v26i4.1848>
- Budd, A. F., Romano, S. L., Smith, N. D., & Barbeitos, M. S. (2010). Rethinking the Phylogeny of Scleractinian Corals: A Review of Morphological and Molecular Data. *Integrative and Comparative Biology*, 50(3), 411–427. <https://doi.org/10.1093/icb/icq062>
- Burke, L. M., Selig, E., & Spalding, M. (2002). *Reefs at risk in southeast Asia*. World Resources Institute.
- Burke, L., Reyntar, K., Spalding, M., & Perry, A. (2011). *Reefs at Risk Revisited*. <https://www.wri.org/research/reefs-risk-revisited>
- Burke, L., Reyntar, K., Spalding, M., & Perry, A. (2012). *Reefs At Risk Revisited in the Coral Triangle*.
- Campoy, A. N., Addamo, A. M., Machordom, A., Meade, A., Rivadeneira, M. M., Hernández, C. E., & Venditti, C. (2020). The Origin and Correlated Evolution of Symbiosis and Coloniality in Scleractinian Corals. *Frontiers in Marine Science*, 7. <https://www.frontiersin.org/articles/10.3389/fmars.2020.00461>

- Carilli, J., Donner, S. D., & Hartmann, A. C. (2012). Historical Temperature Variability Affects Coral Response to Heat Stress. *PLOS ONE*, 7(3), e34418. <https://doi.org/10.1371/journal.pone.0034418>
- Cesar, H., Lundin, C., Bettencourt, S., & Dixon, J. (1997). Indonesian coral reefs-An economic analysis of a precious but threatened resource. *Ambio*, 26.
- Chateaufreynaud, Y., & Lapierre, A. (Eds.). (2019). La plongée subaquatique. In *Aspects du sport à la Réunion* (pp. 37–43). Maison des Sciences de l’Homme d’Aquitaine. <http://books.openedition.org/msha/16907>
- Cheung, R. C. F., Ng, T. B., & Wong, J. H. (2015). Marine Peptides: Bioactivities and Applications. *Marine Drugs*, 13(7), Article 7. <https://doi.org/10.3390/md13074006>
- Chirayath, V., & Instrella, R. (2019). Fluid lensing and machine learning for centimeter-resolution airborne assessment of coral reefs in American Samoa. *Remote Sensing of Environment*, 235, 111475. <https://doi.org/10.1016/j.rse.2019.111475>
- Cinner, J. (2014). Coral reef livelihoods. *Current Opinion in Environmental Sustainability*, 7, 65–71. <https://doi.org/10.1016/j.cosust.2013.11.025>
- Clements, C. S., & Hay, M. E. (2018). Overlooked coral predators suppress foundation species as reefs degrade. *Ecological Applications*, 28(7), 1673–1682. <https://doi.org/10.1002/eap.1765>
- Cole, A. J., Lawton, R. J., Pratchett, M. S., & Wilson, S. K. (2011). Chronic coral consumption by butterflyfishes. *Coral Reefs*, 30(1), 85–93. <https://doi.org/10.1007/s00338-010-0674-6>
- Cole, A. J., Pratchett, M. S., & Jones, G. P. (2008). Diversity and functional importance of coral-feeding fishes on tropical coral reefs. *Fish and Fisheries*, 9(3), 286–307. <https://doi.org/10.1111/j.1467-2979.2008.00290.x>
- Coles, S. L., & Riegl, B. M. (2013). Thermal tolerances of reef corals in the Gulf: A review of the potential for increasing coral survival and adaptation to climate change through assisted translocation. *Marine Pollution Bulletin*, 72(2), 323–332. <https://doi.org/10.1016/j.marpolbul.2012.09.006>
- Cooper, E. L. (2010). Evolution of immune systems from self/not self to danger to artificial immune systems (AIS). *Physics of Life Reviews*, 7(1), 55–78. <https://doi.org/10.1016/j.plrev.2009.12.001>
- Coral Triangle Initiative. (n.d.). *The Coral Triangle*. Coral Triangle Atlas. Retrieved March 22, 2023, from <http://ctatlas.coraltriangleinitiative.org/About>
- Costanza, R., d’Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O’Neill, R. V., Paruelo, J., Raskin, R. G., Sutton, P., & van den Belt, M. (1997). The value of the world’s ecosystem services and natural capital. *Nature*, 387(6630), Article 6630. <https://doi.org/10.1038/387253a0>
- Counsell, C. W. W., Johnston, E. C., & Sale, T. L. (2019). Colony size and depth affect wound repair in a branching coral. *Marine Biology*, 166(11), 148. <https://doi.org/10.1007/s00227-019-3601-6>

- Crabbe, J. (2006). Challenges for sustainability in cultures where regard for the future may not be present. *Sustainability: Science, Practice, & Policy*, 2(2). <https://www.tandfonline.com/doi/epdf/10.1080/15487733.2006.11907985?needAccess=true&role=button>
- Cróquer, A., Villamizar, E., & Noriega, N. (2002). Environmental factors affecting tissue regeneration of the reef—Building coral *Montastraea annularis* (Faviidae) at Los Roques National Park, Venezuela. *Revista de Biología Tropical*, 50(3–4), 1055–1065.
- Darwin, C. (1842). *The structure and distribution of coral reefs*. Smith Elder and Co. <http://darwin-online.org.uk/content/frameset?pageseq=1&itemID=F271&viewtype=text>
- Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, 6(2), 94–98. <https://doi.org/10.7861/futurehosp.6-2-94>
- Day, J. (2008). The need and practice of monitoring, evaluating and adapting marine planning and management—Lessons from the Great Barrier Reef. *Marine Policy*, 32(5), 823–831. <https://doi.org/10.1016/j.marpol.2008.03.023>
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (n.d.). *ImageNet: A Large-Scale Hierarchical Image Database*.
- Di Castri, F., & Balaji, V. (2002). *Tourism, Biodiversity and Information*. Backhuys.
- Done, T. J. (1992). Phase shifts in coral reef communities and their ecological significance. *Hydrobiologia*, 247(1), 121–132. <https://doi.org/10.1007/BF00008211>
- Dove, S. G., Takabayashi, M., & Hoegh-Guldberg, O. (1995). Isolation and Partial Characterization of the Pink and Blue Pigments of Pocilloporid and Acroporid Corals. *Biological Bulletin*, 189(3), 288–297. <https://doi.org/10.2307/1542146>
- Downs, C. A., Woodley, C. M., Richmond, R. H., Lanning, L. L., & Owen, R. (2005). Shifting the paradigm of coral-reef ‘health’ assessment. *Marine Pollution Bulletin*, 51(5), 486–494. <https://doi.org/10.1016/j.marpolbul.2005.06.028>
- Drap, P., Merad, D., Mahiddine, A., Seinturier, J., Gerenton, P., Peloso, D., Boi, P.-M., Bianchimani, O., & Garrabou, J. (2013). AUTOMATING THE MEASUREMENT OF RED CORAL IN SITU USING UNDERWATER PHOTOGRAMMETRY AND CODED TARGETS. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-5/W2, 231.
- Dubinsky, Z., & Falkowski, P. (2011). Light as a Source of Information and Energy in Zooxanthellate Corals. In Z. Dubinsky & N. Stambler (Eds.), *Coral Reefs: An Ecosystem in Transition* (pp. 107–118). Springer Netherlands. https://doi.org/10.1007/978-94-007-0114-4_8
- Edinger, E. N., & Risk, M. J. (1998). Reef Degradation and Coral BO IO/0alversllo;~y in Indonesi•a. • Effects of Land-based Pollution, Destructive Fishing Practices and Changes Over Time. *Marine Pollution Bulletin*, 36(8).
- English, S. A., Wilkinson, C., Baker, V., & Australian Institute of Marine Science (Eds.). (1997). *Survey manual for tropical marine resources* (2. ed). Australian Institute of Marine Science.

- Ereshefsky, M. (2009). Defining ‘health’ and ‘disease.’ *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 40(3), 221–227. <https://doi.org/10.1016/j.shpsc.2009.06.005>
- Fadli, N., Campbell, S. J., Ferguson, K., Keyse, J., Rudi, E., Riedel, A., & Baird, A. H. (2012). The role of habitat creation in coral reef conservation: A case study from Aceh, Indonesia. *Oryx*, 46(4), 501–507. <https://doi.org/10.1017/S0030605312000142>
- FAO. (2018). *French Polynesia GLOBEFISH Market Profile—2018*. Food and Agricultural Organization of the United Nations. <https://www.fao.org/3/cb5614en/cb5614en.pdf>
- Finelli, C. M., Helmuth, B. S. T., Pentcheff, N. D., & Wethey, D. S. (2006). Water flow influences oxygen transport and photosynthetic efficiency in corals. *Coral Reefs*, 25(1), 47–57. <https://doi.org/10.1007/s00338-005-0055-8>
- Foale, S., Adhuri, D., Aliño, P., Allison, E., Andrew, N., Cohen, P., Evans, L., Fabinyi, M., Fidelman, P., Gregory, C., Stacey, N., Tanzer, J., & Weeratunge, N. (2013). Food security and the Coral Triangle Initiative. *Marine Policy*, 38, 174–183. <https://doi.org/10.1016/j.marpol.2012.05.033>
- Fox, H. (2004). Coral recruitment in blasted and unblasted sites in Indonesia: Assessing rehabilitation potential. *Marine Ecology-Progress Series - MAR ECOL-PROGR SER*, 269, 131–139. <https://doi.org/10.3354/meps269131>
- Fox, H. E., & Caldwell, R. L. (2006). Recovery from blast fishing on coral reefs: A tale of two scales. *Ecological Applications: A Publication of the Ecological Society of America*, 16(5), 1631–1635. [https://doi.org/10.1890/1051-0761\(2006\)016\[1631:rbbfoc\]2.0.co;2](https://doi.org/10.1890/1051-0761(2006)016[1631:rbbfoc]2.0.co;2)
- Fox, H. E., Pet, J. S., Dahuri, R., & Caldwell, R. L. (2003). Recovery in rubble fields: Long-term impacts of blast fishing. *Marine Pollution Bulletin*, 46(8), 1024–1031. [https://doi.org/10.1016/S0025-326X\(03\)00246-7](https://doi.org/10.1016/S0025-326X(03)00246-7)
- Glynn, P. W. (1974). The Impact of “Acanthaster” on Corals and Coral Reefs in the Eastern Pacific. *Environmental Conservation*, 1(4), 295–304.
- Goldberg, W. M. (2018). Coral Food, Feeding, Nutrition, and Secretion: A Review. In M. Kloc & J. Z. Kubiak (Eds.), *Marine Organisms as Model Systems in Biology and Medicine* (Vol. 65, pp. 377–421). Springer International Publishing. https://doi.org/10.1007/978-3-319-92486-1_18
- Gomez, E. D., Cabaitan, P. C., Yap, H. T., & Dizon, R. M. (2014). Can Coral Cover be Restored in the Absence of Natural Recruitment and Reef Recovery? *Restoration Ecology*, 22(2), 142–150. <https://doi.org/10.1111/rec.12041>
- Gómez-Ríos, A., Tabik, S., Luengo, J., Shihavuddin, A., Krawczyk, B., & Herrera, F. (2019). Towards highly accurate coral texture images classification using deep convolutional neural networks and data augmentation. *Expert Systems with Applications*, 118, 315–328. <https://doi.org/10.1016/j.eswa.2018.10.010>
- Gonsard, H. (2016). *French Polynesia at a Glance* (C. Periou, Ed.). Institut d’Émission d’Outre-Mer. https://www.ieom.fr/IMG/pdf/ne206_portrait_panorama_2015_polynesie_va.pdf

- González-Rivero, M., Beijbom, O., Rodríguez-Ramírez, A., Bryant, D. E. P., Ganase, A., González-Marrero, Y., Herrera-Reveles, A., Kennedy, E. V., Kim, C. J. S., Lopez-Marcano, S., Markey, K., Neal, B. P., Osborne, K., Reyes-Nivia, C., Sampayo, E. M., Stolberg, K., Taylor, A., Vercelloni, J., Wyatt, M., & Hoegh-Guldberg, O. (2020). Monitoring of Coral Reefs Using Artificial Intelligence: A Feasible and Cost-Effective Approach. *Remote Sensing*, 12(3), Article 3. <https://doi.org/10.3390/rs12030489>
- González-Rivero, M., Beijbom, O., Rodríguez-Ramírez, A., Holtrop, T., González-Marrero, Y., Ganase, A., Roelfsema, C., Phinn, S., & Hoegh-Guldberg, O. (2016). Scaling up Ecological Measurements of Coral Reefs Using Semi-Automated Field Image Collection and Analysis. *Remote Sensing*, 8(1), Article 1. <https://doi.org/10.3390/rs8010030>
- Gonzalez-Rivero, M., Bongaerts, P., Beijbom, O., Pizarro, O., Friedman, A., Rodríguez-Ramírez, A., Upcroft, B., Laffoley, D., Kline, D., Bailhache, C., Vevers, R., & Hoegh-Guldberg, O. (2014). The Catlin Seaview Survey—Kilometre-scale seascape assessment, and monitoring of coral reef ecosystems. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 24, 184–198. <https://doi.org/10.1002/aqc.2505>
- Goreau, T. F., Goreau, N. I., & Goreau, T. J. (1979). Corals and Coral Reefs. *Scientific American*, 241(2), 124–137.
- Graham, N. A. J., & Nash, K. L. (2013). The importance of structural complexity in coral reef ecosystems. *Coral Reefs*, 32(2), 315–326. <https://doi.org/10.1007/s00338-012-0984-y>
- Graham, N. A. J., Wilson, S. K., Jennings, S., Polunin, N. V. C., Bijoux, J. P., & Robinson, J. (2006). Dynamic fragility of oceanic coral reef ecosystems. *Proceedings of the National Academy of Sciences*, 103(22), 8425–8429. <https://doi.org/10.1073/pnas.0600693103>
- Gratwicke, B., & Speight, M. (2005). Effects of habitat complexity on Caribbean marine fish assemblages. *Marine Ecology Progress Series*, 292, 301–310. <https://doi.org/10.3354/meps292301>
- Grawunder, D., Hambleton, E. A., Bucher, M., Wolfowicz, I., Bechtoldt, N., & Guse, A. (2015). Induction of Gametogenesis in the Cnidarian Endosymbiosis Model *Aiptasia* sp. *Scientific Reports*, 5(1), Article 1. <https://doi.org/10.1038/srep15677>
- Greer, L., & Benson, M. (2012). *Texture and Color Distribution-based Classification for Live Coral Detection*. <https://www.semanticscholar.org/paper/Texture-and-Color-Distribution-based-Classification-Greer-Benson/e2ddde221c5e6f315ae394c7d2b549b853eeb6ac>
- Guan, Y., Hohn, S., & Merico, A. (2015). Suitable Environmental Ranges for Potential Coral Reef Habitats in the Tropical Ocean. *PLOS ONE*, 10(6), e0128831. <https://doi.org/10.1371/journal.pone.0128831>
- Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J. L., & Webster, D. R. (2016). Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*, 316(22), 2402–2410. <https://doi.org/10.1001/jama.2016.17216>
- Haapkylä, J., Seymour, A. S., Trebilco, J., & Smith, D. (2007). Coral disease prevalence and coral health in the Wakatobi Marine Park, south-east Sulawesi, Indonesia. *Journal of the Marine*

Biological Association of the United Kingdom, 87(2), 403–414.
<https://doi.org/10.1017/S0025315407055828>

- Haas, A. F., Guibert, M., Foerschner, A., Co, T., Calhoun, S., George, E., Hatay, M., Dinsdale, E., Sandin, S. A., Smith, J. E., Vermeij, M. J. A., Felts, B., Dustan, P., Salamon, P., & Rohwer, F. (2015). Can we measure beauty? Computational evaluation of coral reef aesthetics. *PeerJ*, 3, e1390. <https://doi.org/10.7717/peerj.1390>
- Habibi, A., & Sartin, J. (2007). *A Decade of Reef Check Monitoring: Indonesian Coral Reefs, Condition and Trends*. <https://www.semanticscholar.org/paper/A-Decade-of-Reef-Check-Monitoring%3A-Indonesian-Coral-Habibi-Sartin/6380350c12d3602d2c897bb5b24482746ef82a89>
- Hadi, T., Muhammad, A., Giyanto, G., Prayudha, B., Johan, O., Budiyanto, A., Rezza, A., Alifatri, L. O., Sulha, S., & Shar, S. (2020). *The Status of Indonesian Coral Reefs 2019*. Research Center for Oceanography. https://www.researchgate.net/publication/342663285_The_Status_of_Indonesian_Coral_Reefs_2019
- Haenlein, M., & Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*, 61(4), 5–14. <https://doi.org/10.1177/0008125619864925>
- Hallock, P. (2001). Coral Reefs, Carbonate Sediments, Nutrients, and Global Change. In G. D. Stanley (Ed.), *The History and Sedimentology of Ancient Reef Systems* (Vol. 17, pp. 387–427). Springer US. https://doi.org/10.1007/978-1-4615-1219-6_11
- Hampton-Smith, M., Bower, D. S., & Mika, S. (2021). A review of the current global status of blast fishing: Causes, implications and solutions. *Biological Conservation*, 262, 109307. <https://doi.org/10.1016/j.biocon.2021.109307>
- Hamylton, S. M., Zhou, Z., & Wang, L. (2020). What Can Artificial Intelligence Offer Coral Reef Managers? *Frontiers in Marine Science*, 7. <https://www.frontiersin.org/articles/10.3389/fmars.2020.603829>
- Harriott, V., Goggin, L., & Sweatman, H. (2006). *Crown of Thorns Starfish on the Great Barrier Reef*. CRC Reef Research Center. https://www.reefresilience.org/pdf/COTS_Nov2003.pdf
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep Residual Learning for Image Recognition* (arXiv:1512.03385). arXiv. <https://doi.org/10.48550/arXiv.1512.03385>
- Hedley, J. D., Roelfsema, C. M., Chollett, I., Harborne, A. R., Heron, S. F., Weeks, S., Skirving, W. J., Strong, A. E., Eakin, C. M., Christensen, T. R. L., Ticzon, V., Bejarano, S., & Mumby, P. J. (2016). Remote Sensing of Coral Reefs for Monitoring and Management: A Review. *Remote Sensing*, 8(2), Article 2. <https://doi.org/10.3390/rs8020118>
- Heyward, A. J., & Negri, A. P. (1999). Natural inducers for coral larval metamorphosis. *Coral Reefs*, 18(3), 273–279. <https://doi.org/10.1007/s003380050193>
- Hoegh-Guldberg, O., Jacob, D., Taylor, M., Bindi, M., Brown, S., Camilloni, I., Diedhiou, A., Djalante, R., Ebi, K. L., Engelbrecht, F., Hijioka, Y., Mehrotra, S., Payne, A., Seneviratne, S.

- I., Thomas, A., Warren, R., Zhou, G., Halim, S. A., Achlatis, M., ... Sherstyukov, B. (2018). *Impacts of 1.5°C of Global Warming on Natural and Human Systems*.
- Hopkinson, B. M., King, A. C., Owen, D. P., Johnson-Roberson, M., Long, M. H., & Bhandarkar, S. M. (2020). Automated classification of three-dimensional reconstructions of coral reefs using convolutional neural networks. *PLOS ONE*, *15*(3), e0230671. <https://doi.org/10.1371/journal.pone.0230671>
- Houlbreque, F., & Ferrier-Pagès, C. (2008). Heterotrophy in Tropical Scleractinian Corals. *Biological Reviews of the Cambridge Philosophical Society*, *84*, 1–17. <https://doi.org/10.1111/j.1469-185X.2008.00058.x>
- Huggett, M. J., & Apprill, A. (2019). Coral microbiome database: Integration of sequences reveals high diversity and relatedness of coral-associated microbes. *Environmental Microbiology Reports*, *11*(3), 372–385. <https://doi.org/10.1111/1758-2229.12686>
- Huston, M. A. (1985). Patterns of Species Diversity on Coral Reefs. *Annual Review of Ecology and Systematics*, *16*(1), 149–177. <https://doi.org/10.1146/annurev.es.16.110185.001053>
- IPCC. (2018). *Global Warming of 1.5°C: IPCC Special Report on Impacts of Global Warming of 1.5°C above Pre-industrial Levels in Context of Strengthening Response to Climate Change, Sustainable Development, and Efforts to Eradicate Poverty* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/9781009157940>
- Ismail, A., & Wediawati, D. B. (2023). *Understanding the Fundamentals of Machine Learning and AI for Digital Business*. Asadel Publisher.
- Jennings, S., & Polunin, N. V. C. (1996a). Impacts of Fishing on Tropical Reef Ecosystems. *Ambio*, *25*(1), 44–49.
- Jennings, S., & Polunin, N. V. C. (1996b). Impacts of Fishing on Tropical Reef Ecosystems. *Ambio*, *25*(1), 44–49.
- Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology*, *2*(4). <https://doi.org/10.1136/svn-2017-000101>
- Johnson-Roberson, M., Kumar, S., & Williams, S. (2006). Segmentation and Classification of Coral for Oceanographic Surveys: A Semi-Supervised Machine Learning Approach. *OCEANS 2006 - Asia Pacific*, 1–6. <https://doi.org/10.1109/OCEANSAP.2006.4393835>
- Johnston, W., Hine, D., & Southgate, P. (2019). Overview of the Development and Modern Landscape of Marine Pearl Culture in the South Pacific. *Journal of Shellfish Research*, *38*, 499. <https://doi.org/10.2983/035.038.0301>
- Jokiel, P., & Guinther, E. (1978). Effects of Temperature on Reproduction in the Hermatypic Coral *Pocillopora damicornis*. *Bulletin of Marine Science*, *28*, 786–789.
- Jordan, I. E., & Samways, M. J. (2001). Recent changes in coral assemblages of a South African coral reef, with recommendations for long-term monitoring. *Biodiversity & Conservation*, *10*(7), 1027–1037. <https://doi.org/10.1023/A:1016622030597>

- Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), 15–25. <https://doi.org/10.1016/j.bushor.2018.08.004>
- Knowlton, N., & Jackson, J. B. C. (2008). Shifting Baselines, Local Impacts, and Global Change on Coral Reefs. *PLOS Biology*, 6(2), e54. <https://doi.org/10.1371/journal.pbio.0060054>
- Kocak, D. M., & Caimi, F. M. (2005). The Current Art of Underwater Imaging – With a Glimpse of the Past and Vision of the Future. *Marine Technology Society Journal*, 39(3), 5–26. <https://doi.org/10.4031/002533205787442576>
- Komyakova, V., Munday, P. L., & Jones, G. P. (2013). Relative Importance of Coral Cover, Habitat Complexity and Diversity in Determining the Structure of Reef Fish Communities. *PLOS ONE*, 8(12), e83178. <https://doi.org/10.1371/journal.pone.0083178>
- Kramer, P. A. (2003). *Synthesis of coral reef health indicators for the Western Atlantic: Results of the AGRRA program (1997-2000)*. <http://repository.si.edu/xmlui/handle/10088/7759>
- Kubomura, T., Wee, H. B., & Reimer, J. D. (2021). Investigating incidence and possible causes of pink and purple pigmentation response in hard coral genus *Porites* around Okinawajima Island, Japan. *Regional Studies in Marine Science*, 41, 101569. <https://doi.org/10.1016/j.rsma.2020.101569>
- Kuo, C.-Y., Tsai, C.-H., Huang, Y.-Y., Heng, W. K., Hsiao, A.-T., Hsieh, H. J., & Chen, C. A. (2022). Fine intervals are required when using point intercept transects to assess coral reef status. *Frontiers in Marine Science*, 9. <https://www.frontiersin.org/articles/10.3389/fmars.2022.795512>
- Lang, M. A., Marinelli, R. L., Roberts, S. J., & Taylor, P. R. (2019). *Research and Discoveries: The Revolution of Science through Scuba*. Smithsonian Contributions to the Marine Sciences. <https://doi.org/10.5479/si.1943667X.39>
- Larned, S. T. (1998). Nitrogen- versus phosphorus-limited growth and sources of nutrients for coral reef macroalgae. *Marine Biology*, 132(3), 409–421. <https://doi.org/10.1007/s002270050407>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Lesser, M. P. (1997). Oxidative stress causes coral bleaching during exposure to elevated temperatures. *Coral Reefs*, 16(3), 187–192. <https://doi.org/10.1007/s003380050073>
- Lombard, F., Bourdin, G., Pesant, S., Agostini, S., Baudena, A., Boissin, E., Cassar, N., Clampitt, M., Conan, P., Da Silva, O., Dimier, C., Douville, E., Elineau, A., Fin, J., Flores, J. M., Ghiglione, J.-F., Hume, B. C. C., Jalabert, L., John, S. G., ... Gorsky, G. (2023). Open science resources from the Tara Pacific expedition across coral reef and surface ocean ecosystems. *Scientific Data*, 10(1), Article 1. <https://doi.org/10.1038/s41597-022-01757-w>
- Madduppa, H., & Zamani, N. (2011). A Standard Criteria for Assessing the Health of Coral Reefs: Implication for Management and Conservation. *Journal of Indonesia Coral Reefs*, 1, 137–146.

- Mahmood, A., Bennamoun, M., An, S., Sohel, F., Boussaid, F., Hovey, R., Kendrick, G., & Fisher, R. B. (2016). Automatic annotation of coral reefs using deep learning. *OCEANS 2016 MTS/IEEE Monterey*, 1–5. <https://doi.org/10.1109/OCEANS.2016.7761105>
- Mason, B., Beard, M., & Miller, M. W. (2011). Coral larvae settle at a higher frequency on red surfaces. *Coral Reefs*, 30(3), 667–676. <https://doi.org/10.1007/s00338-011-0739-1>
- Mayfield, A. B., Dempsey, A. C., Chen, C.-S., & Lin, C. (2022). Expediting the Search for Climate-Resilient Reef Corals in the Coral Triangle with Artificial Intelligence. *Applied Sciences*, 12(24), Article 24. <https://doi.org/10.3390/app122412955>
- Mayfield, A. B., & Lin, C. (2023). Field-Testing a Proteomics-Derived Machine-Learning Model for Predicting Coral Bleaching Susceptibility. *Applied Sciences*, 13(3), Article 3. <https://doi.org/10.3390/app13031718>
- McClanahan, T., Darling, E., Oddenyo, R., Surya, G., Beger, M., Fox, H., Jupiter, S., McLeod, L., McManus, L., van Woeseik, R., Grantham, H., Logan, C., Maina, J., Patankar, V., Wenger, A., & Zinke, J. (2022). *FORECASTING CLIMATE SANCTUARIES FOR SECURING THE FUTURE OF CORAL REEFS*.
- McClanahan, T. R. (2004). The relationship between bleaching and mortality of common corals. *Marine Biology*, 144(6), 1239–1245. <https://doi.org/10.1007/s00227-003-1271-9>
- McCorduck, P., Minsky, M., Selfridge, O., & Simon, H. (1977). HISTORY OF ARTIFICIAL INTELLIGENCE. *IJCAI*, 951–954.
- McCormick, M. (1994). Comparison of field methods for measuring surface topography and their associations with a tropical reef fish assemblage. *Marine Ecology Progress Series*, 112, 87–96. <https://doi.org/10.3354/meps112087>
- Mcfield, M., Kramer, P., Gorrez, M., & McPherson, M. (2007). *Healthy Reefs for Healthy People: A Guide to Indicators of Reef Health and Social Well-being in the Mesoamerican Reef Region*.
- McWilliams, J. P., Côté, I. M., Gill, J. A., Sutherland, W. J., & Watkinson, A. R. (2005). Accelerating Impacts of Temperature-Induced Coral Bleaching in the Caribbean. *Ecology*, 86(8), 2055–2060. <https://doi.org/10.1890/04-1657>
- Meesters, E., Noordeloos, M., & Bak, R. (1994). Damage and regeneration: Links to growth in the reef-building coral *Montastrea annularis*. *Marine Ecology Progress Series*, 112, 119–128. <https://doi.org/10.3354/meps112119>
- Miller, A. W., & Richardson, L. L. (2015). Emerging coral diseases: A temperature-driven process? *Marine Ecology*, 36(3), 278–291. <https://doi.org/10.1111/maec.12142>
- Mironidou-Tzouveleki, M., & Dokos, C. (2007). Nature’s drug store attacks again: The cone snail’s pain relief toxin. *Aristotle University Medical Journal*, 34(3), Article 3.
- Miththapala, S. (2008). Coral Reefs. Coastal Ecosystems Series (Vol 1). *Ecosystems and Livelihoods Group Asia*, IUCN, 1, 1–35.
- Moberg, F., & Folke, C. (1999). Ecological goods and services of coral reef ecosystems. *Ecological Economics*, 29(2), 215–233. [https://doi.org/10.1016/S0921-8009\(99\)00009-9](https://doi.org/10.1016/S0921-8009(99)00009-9)

- Mohamed, A. R., Ochsenkühn, M. A., Kazlak, A. M., Moustafa, A., & Amin, S. A. (2023). The coral microbiome: Towards an understanding of the molecular mechanisms of coral–microbiota interactions. *FEMS Microbiology Reviews*, 47(2), fuad005. <https://doi.org/10.1093/femsre/fuad005>
- Molle, G., Wadrawane, J.-M., Lagarde, L., & Wright, D. (2023). The sacred stone from the sea. Archaeological and ethnographic perspectives on the ritual value of coral across the Pacific. *Archaeology in Oceania*, 0, 1–23. <https://doi.org/10.1002/arco.5284>
- Moniruzzaman, M., Islam, S., Bennamoun, M., & Lavery, P. (2017). *Deep Learning on Underwater Marine Object Detection: A Survey*. 150–160. https://doi.org/10.1007/978-3-319-70353-4_13
- Morand, G., Dixon, S., & Le Berre, T. (2022). Identifying key factors for coral survival in reef restoration projects using deep learning. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 32(11), 1758–1773. <https://doi.org/10.1002/aqc.3878>
- Morse, D. E., Hooker, N., Morse, A. N. C., & Jensen, R. A. (1988). Control of larval metamorphosis and recruitment in sympatric agariciid corals. *Journal of Experimental Marine Biology and Ecology*, 116(3), 193–217. [https://doi.org/10.1016/0022-0981\(88\)90027-5](https://doi.org/10.1016/0022-0981(88)90027-5)
- Muir, P. R., & Pichon, M. (2019). Biodiversity of Reef-Building, Scleractinian Corals. In Y. Loya, K. A. Puglise, & T. C. L. Bridge (Eds.), *Mesophotic Coral Ecosystems* (Vol. 12, pp. 589–620). Springer International Publishing. https://doi.org/10.1007/978-3-319-92735-0_33
- Mumuni, A., & Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. *Array*, 16, 100258. <https://doi.org/10.1016/j.array.2022.100258>
- Musa, G., & Dimmock, K. (2012). Scuba Diving Tourism Introduction to Special Issue. *Tourism in Marine Environments*, 8(1), 1–5. <https://doi.org/10.3727/154427312X13262430523947>
- Neal, B. P., Lin, T.-H., Winter, R. N., Treibitz, T., Beijbom, O., Kriegman, D., Kline, D. I., & Greg Mitchell, B. (2015). Methods and measurement variance for field estimations of coral colony planar area using underwater photographs and semi-automated image segmentation. *Environmental Monitoring and Assessment*, 187(8), 496. <https://doi.org/10.1007/s10661-015-4690-4>
- Neumann, B., Vafeidis, A. T., Zimmermann, J., & Nicholls, R. J. (2015). Future Coastal Population Growth and Exposure to Sea-Level Rise and Coastal Flooding—A Global Assessment. *PLOS ONE*, 10(3), e0118571. <https://doi.org/10.1371/journal.pone.0118571>
- Nicolet, K. J., Chong-Seng, K. M., Pratchett, M. S., Willis, B. L., & Hoogenboom, M. O. (2018). Predation scars may influence host susceptibility to pathogens: Evaluating the role of corallivores as vectors of coral disease. *Scientific Reports*, 8(1), Article 1. <https://doi.org/10.1038/s41598-018-23361-y>
- Nunes, J. A. C. C., Cruz, I. C. S., Nunes, A., & Pinheiro, H. T. (2020). Speeding up coral reef conservation with AI-aided automated image analysis. *Nature Machine Intelligence*, 2(6), Article 6. <https://doi.org/10.1038/s42256-020-0192-3>
- Obura, D. O., Aeby, G., Amornthammarong, N., Appeltans, W., Bax, N., Bishop, J., Brainard, R. E., Chan, S., Fletcher, P., Gordon, T. A. C., Gramer, L., Gudka, M., Halas, J., Hendee, J., Hodgson, G., Huang, D., Jankulak, M., Jones, A., Kimura, T., ... Wongbusarakum, S. (2019).

Coral Reef Monitoring, Reef Assessment Technologies, and Ecosystem-Based Management. *Frontiers in Marine Science*, 6. <https://www.frontiersin.org/articles/10.3389/fmars.2019.00580>

- Ong, T. F., & Musa, G. (2011). An examination of recreational divers' underwater behaviour by attitude-behaviour theories. *Current Issues in Tourism*, 14(8), 779–795. <https://doi.org/10.1080/13683500.2010.545370>
- O'Shea, K., & Nash, R. (2015). *An Introduction to Convolutional Neural Networks* (arXiv:1511.08458). arXiv. <https://doi.org/10.48550/arXiv.1511.08458>
- PADI. (2021). *2021 PADI Worldwide Statistics*.
- Page, C., Field, S., Pollock, F., Lamb, J., Shedrawi, G., & Wilson, S. (2016). Assessing coral health and disease from digital photographs and in situ surveys. *Assessing Coral Health and Disease from Digital Photographs and in Situ Surveys*, 189. <https://doi.org/10.1007/s10661-016-5743-z>
- Palardy, J., Grottoli, A., & Matthews, K. (2005). Effects of upwelling, depth, morphology and polyp size on feeding in three species of Panamanian corals. *Marine Ecology-Progress Series - MAR ECOL-PROGR SER*, 300, 79–89. <https://doi.org/10.3354/meps300079>
- Palmer, C., & Traylor-Knowles, N. (2018). Cnidaria: Anthozoans in the Hot Seat. In *Advances in Comparative Immunology* (pp. 51–93). https://doi.org/10.1007/978-3-319-76768-0_3
- Palmer, C. V., Mydlarz, L. D., & Willis, B. L. (2008). Evidence of an inflammatory-like response in non-normally pigmented tissues of two scleractinian corals. *Proceedings of the Royal Society B: Biological Sciences*, 275(1652), 2687–2693. <https://doi.org/10.1098/rspb.2008.0335>
- Palmer, C. V., & Traylor-Knowles, N. (2012). Towards an integrated network of coral immune mechanisms. *Proceedings of the Royal Society B: Biological Sciences*, 279(1745), 4106–4114. <https://doi.org/10.1098/rspb.2012.1477>
- Palmer, C. V., Traylor-Knowles, N. G., Willis, B. L., & Bythell, J. C. (2011). Corals Use Similar Immune Cells and Wound-Healing Processes as Those of Higher Organisms. *PLOS ONE*, 6(8), e23992. <https://doi.org/10.1371/journal.pone.0023992>
- Pateman, V. (2009). *Color Correction for Underwater Photography*.
- Paulay, G. (1997). *Diversity and Distribution of Reef Organisms* (pp. 298–353). https://doi.org/10.1007/978-1-4615-5995-5_14
- Pearson, R. (1981). Recovery and Recolonization of Coral Reefs. *Marine Ecology Progress Series*, 4, 105–122. <https://doi.org/10.3354/meps004105>
- Permana, R., Akbarsyah, N., Putra, P. K., & Andhikawati, A. (2020). Analysis Condition of Coral Reef Covering in Pramuka Island Waters, Seribu Islands using Line Intercept Transect (LIT) Method. *Jurnal Riset Biologi Dan Aplikasinya*, 2(2), Article 2. <https://doi.org/10.26740/jrba.v2n2.p77-81>

- Pert, P. L., Thiault, L., Curnock, M. I., Becken, S., & Claudet, J. (2020). Beauty and the reef: Evaluating the use of non-expert ratings for monitoring aesthetic values of coral reefs. *Science of The Total Environment*, 730, 139156. <https://doi.org/10.1016/j.scitotenv.2020.139156>
- Planes, S., Allemand, D., Agostini, S., Banaigs, B., Boissin, E., Boss, E., Bourdin, G., Bowler, C., Douville, E., Flores, J. M., Forcioli, D., Furla, P., Galand, P. E., Ghiglione, J.-F., Gilson, E., Lombard, F., Moulin, C., Pesant, S., Poulain, J., ... Consortium, the T. P. (2019). The Tara Pacific expedition—A pan-ecosystemic approach of the “-omics” complexity of coral reef holobionts across the Pacific Ocean. *PLOS Biology*, 17(9), e3000483. <https://doi.org/10.1371/journal.pbio.3000483>
- Pogoreutz, C., Voolstra, C. R., Rådecker, N., Weis, V., Cardenas, A., & Raina, J.-B. (2020). *The coral holobiont highlights the dependence of cnidarian animal hosts on their associated microbes* (T. C. G. Bosch & M. G. Hadfield, Eds.; 1st ed., pp. 91–118). CRC Press. <https://doi.org/10.1201/9780429277375-7>
- Poplin, R., Varadarajan, A. V., Blumer, K., Liu, Y., McConnell, M. V., Corrado, G. S., Peng, L., & Webster, D. R. (2018). Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering*, 2(3), Article 3. <https://doi.org/10.1038/s41551-018-0195-0>
- Praeger, C., Poole, A., Sexton, B. A., Glenn, F. L., Vucko, M., Williams, M. R., Whalan, S., & de Nys, R. (2012). Enhancing the settlement and attachment strength of pediveligers of *Mytilus galloprovincialis* by changing surface wettability and microtopography. *Biofouling*, 28, 175–186. <https://doi.org/10.1080/08927014.2012.662676>
- Pratchett, M. S., Wilson, S. K., Berumen, M. L., & McCormick, M. I. (2004). “Sublethal effects of coral bleaching on an obligate coral feeding butterflyfish.” *Coral Reefs*, 23(3), 352–356. <https://doi.org/10.1007/s00338-004-0394-x>
- Purser, A., Bergmann, M., Lundälv, T., & Nattkemper, T. (2009). Use of machine-learning algorithms for the automated detection of cold-water coral habitats: A pilot study. *Marine Ecology Progress Series*, 397, 241. <https://doi.org/10.3354/meps08154>
- Raphael, A., Dubinsky, Z., Iluz, D., & Netanyahu, N. S. (2020). Neural Network Recognition of Marine Benthos and Corals. *Diversity*, 12(1), Article 1. <https://doi.org/10.3390/d12010029>
- Ravindran, J., & Raghukumar, C. (2006). Histological observations on the scleractinian coral *Porites lutea* affected by pink-line syndrome. *CURRENT SCIENCE*, 90(5).
- Razak, T. B., Boström-Einarsson, L., Alisa, C. A. G., Vida, R. T., & Lamont, T. A. C. (2022). Coral reef restoration in Indonesia: A review of policies and projects. *Marine Policy*, 137, 104940. <https://doi.org/10.1016/j.marpol.2021.104940>
- Reaka-Kudla, M. L., Wilson, D. E., & Wilson, E. O. (1996). *Biodiversity II: Understanding and Protecting Our Biological Resources*. Joseph Henry Press.
- Rebikoff, D. (1967). History of underwater photography. *Photogrammetric Engineering*, 33(8), 897–904.

- Reshef, L., Koren, O., Loya, Y., Zilber-Rosenberg, I., & Rosenberg, E. (2006). The Coral Probiotic Hypothesis. *Environmental Microbiology*, 8(12), 2068–2073. <https://doi.org/10.1111/j.1462-2920.2006.01148.x>
- Rice, M. M., Ezzat, L., & Burkepile, D. E. (2019). Corallivory in the Anthropocene: Interactive Effects of Anthropogenic Stressors and Corallivory on Coral Reefs. *Frontiers in Marine Science*, 5. <https://www.frontiersin.org/articles/10.3389/fmars.2018.00525>
- Rice, M. M., Maher, R. L., Correa, A. M. S., Moeller, H. V., Lemoine, N. P., Shantz, A. A., Burkepile, D. E., & Silbiger, N. J. (2020). Macroborer presence on corals increases with nutrient input and promotes parrotfish bioerosion. *Coral Reefs*, 39(2), 409–418. <https://doi.org/10.1007/s00338-020-01904-y>
- Richardson, D. (1999). *A BRIEF HISTORY OF SCUBA DIVING IN THE UNITED STATES*. 29.
- Roberts, C., Mcclean, C., Veron, J., Hawkins, J., Allen, G., McAllister, D., Mittermeier, C., Schueler, F., Spalding, M., Wells, F., Vynne, C., & Werner, T. (2002). Marine Biodiversity Hotspots and Conservation Priorities for Tropical Reefs. *Science (New York, N.Y.)*, 295, 1280–1284. <https://doi.org/10.1126/science.1067728>
- Rogers, A., Blanchard, J. L., & Mumby, P. J. (2014). Vulnerability of Coral Reef Fisheries to a Loss of Structural Complexity. *Current Biology*, 24(9), 1000–1005. <https://doi.org/10.1016/j.cub.2014.03.026>
- Rosebrock, A. (2017). *Deep Learning for Computer Vision with Python: Starter Bundle*. PyImageSearch.
- Rosenberg, E., Koren, O., Reshef, L., Efrony, R., & Zilber-Rosenberg, I. (2007). The role of microorganisms in coral health, disease and evolution. *Nature Reviews Microbiology*, 5(5), Article 5. <https://doi.org/10.1038/nrmicro1635>
- Röthig, T., Ochsenkühn, M. A., Roik, A., van der Merwe, R., & Voolstra, C. R. (2016). Long-term salinity tolerance is accompanied by major restructuring of the coral bacterial microbiome. *Molecular Ecology*, 25(6), 1308–1323. <https://doi.org/10.1111/mec.13567>
- Rotjan, R., & Lewis, S. (2008). Impact of coral predators on tropical reefs. *Marine Ecology Progress Series*, 367, 73–91. <https://doi.org/10.3354/meps07531>
- Rouphael, A. B., & Inglis, G. J. (2001). “Take only photographs and leave only footprints”?: An experimental study of the impacts of underwater photographers on coral reef dive sites. *Biological Conservation*, 100(3), 281–287. [https://doi.org/10.1016/S0006-3207\(01\)00032-5](https://doi.org/10.1016/S0006-3207(01)00032-5)
- Ruiz-Moreno, D., Willis, B. L., Page, A. C., Weil, E., Cróquer, A., Vargas-Angel, B., Jordan-Garza, A. G., Jordán-Dahlgren, E., Raymundo, L., & Harvell, C. D. (2012). Global coral disease prevalence associated with sea temperature anomalies and local factors. *Diseases of Aquatic Organisms*, 100(3), 249–261. <https://doi.org/10.3354/dao02488>
- Sabine, A. M., Smith, T. B., Williams, D. E., & Brandt, M. E. (2015). Environmental conditions influence tissue regeneration rates in scleractinian corals. *Marine Pollution Bulletin*, 95(1), 253–264. <https://doi.org/10.1016/j.marpolbul.2015.04.006>

- Salvat, B., Aubanel, A., Adjeroud, M., Bouisset, P., Dominique, C., Chancerelle, Y., Cochenec, N., Davies, N., Fougerousse, A., Galzin, R., Lagouy, E., LO, C., Monier, C., Ponsonnet, C., Remoissenet, G., Schneider, D., Stein, A., Tatarata, M., & Villiers, L. (2008). Monitoring of French Polynesia coral reefs and their recent development. *Revue D Ecologie-La Terre Et La Vie*, 63, 145–177.
- Salvat, B., & Pailhe, C. (2002). Islands and coral reefs, population and culture, economy and tourism: World view and a case study of French Polynesia. In *Tourism, Biodiversity, and Information* (pp. 213–231). Backhuys Publishers. <http://ifrecor-doc.fr/items/show/1521>
- Samways, M. J., & Hatton, M. J. (2001). An appraisal of two coral reef rapid monitoring manuals for gathering baseline data. *Bulletin of Marine Science*, 69(2), 471–485.
- Sanciango, J. C., Carpenter, K. E., Etnoyer, P. J., & Moretzsohn, F. (2013). Habitat Availability and Heterogeneity and the Indo-Pacific Warm Pool as Predictors of Marine Species Richness in the Tropical Indo-Pacific. *PLOS ONE*, 8(2), e56245. <https://doi.org/10.1371/journal.pone.0056245>
- Sayer, M. (2007). Scientific diving: A bibliographic analysis of underwater research supported by SCUBA diving, 1995-2006. *Underwater Technology*, 27(3), 75–94. <https://doi.org/10.3723/175605407783360035>
- Sayer, M., & Barrington, J. (2005). Trends in scientific diving: An analysis of scientific diving operation records, 1970-2004. *Underwater Technology*, 26(2), 51–55. <https://doi.org/10.3723/175605405783101458>
- Scaps, P., & Denis, V. (2008). Can organisms associated with live scleractinian corals be used as indicators of coral reef status? *Atoll Research Bulletin*, 566, 1–18. <https://doi.org/10.5479/si.00775630.566.1>
- Scucchia, F., Sauer, K., Zaslansky, P., & Mass, T. (2022). Artificial Intelligence as a Tool to Study the 3D Skeletal Architecture in Newly Settled Coral Recruits: Insights into the Effects of Ocean Acidification on Coral Biomineralization. *Journal of Marine Science and Engineering*, 10(3), Article 3. <https://doi.org/10.3390/jmse10030391>
- Sebens, K., Bernardi, G., Patterson, M., & Burkepile, D. (2012, October 1). *Saturation Diving and Underwater Laboratories: How Underwater Technology Has Aided Research on Coral Biology and Reef Ecology*.
- Sebens, K. P., Grace, S. P., Helmuth, B., Maney Jr., E. J., & Miles, J. S. (1998). Water flow and prey capture by three scleractinian corals, *Madracis mirabilis*, *Montastrea cavernosa* and *Porites porites*, in a field enclosure. *Marine Biology*, 131(2), 347–360. <https://doi.org/10.1007/s002270050328>
- Sebens, K. P., Vandersall, K. S., Savina, L. A., & Graham, K. R. (1996). Zooplankton capture by two scleractinian corals, *Madracis mirabilis* and *Montastrea cavernosa*, in a field enclosure. *Marine Biology*, 127(2), 303–317. <https://doi.org/10.1007/BF00942116>
- Sheppard, C., Davy, S., Pilling, G., & Graham, N. (2017). *The Biology of Coral Reefs*. Oxford University Press.

- Siebeck, U. E., Marshall, N. J., Klüter, A., & Hoegh-Guldberg, O. (2006). Monitoring coral bleaching using a colour reference card. *Coral Reefs*, 25(3), 453–460. <https://doi.org/10.1007/s00338-006-0123-8>
- Sorauf, J. E. (1988). *Ancient and Modern Reefs: The Evolution of Reef Communities*. J. A. Fagerstrom. Wiley-Interscience, New York, 1987. xviii, 600 pp., illus. \$74.95. *Science*, 240(4852), 667–668. <https://doi.org/10.1126/science.240.4852.667>
- Sorokin, Y. I. (2013). *Coral Reef Ecology*. Springer Science & Business Media.
- Souter, D., Planes, S., Wicquart, J., Logan, M., Obura, D., & Staub, F. (2020). *Status of Coral Reefs of the World: 2020*.
- Spalding, M., Ravilious, C., & Green, E. (2001). *World Atlas of Coral Reefs*. UNEP, WCMC.
- 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(1), 23–43. <https://doi.org/10.1016/j.ppees.2006.04.001>
- Stokes, M. D., & Deane, G. B. (2009). Automated processing of coral reef benthic images. *Limnology and Oceanography: Methods*, 7(2), 157.
- Sweet, M. J., & Bulling, M. T. (2017). On the Importance of the Microbiome and Pathobiome in Coral Health and Disease. *Frontiers in Marine Science*, 4. <https://doi.org/10.3389/fmars.2017.00009>
- Teague, J., Allen, M. J., Day, J. C. C., & Scott, T. B. (2020). A Review of Current Coral Monitoring Tools/ Diagnostic Methods & Introducing a New Tool to the Coral Health Toolkit. *A Review of Current Coral Monitoring Tools/ Diagnostic Methods & Introducing a New Tool to the Coral Health Toolkit*. <https://doi.org/10.20944/preprints202008.0045.v1>
- Tebben, J., Motti, C. A., Siboni, N., Tapiolas, D. M., Negri, A. P., Schupp, P. J., Kitamura, M., Hatta, M., Steinberg, P. D., & Harder, T. (2015). Chemical mediation of coral larval settlement by crustose coralline algae. *Scientific Reports*, 5(1), Article 1. <https://doi.org/10.1038/srep10803>
- Thurber, R. V., Willner-Hall, D., Rodriguez-Mueller, B., Desnues, C., Edwards, R. A., Angly, F., Dinsdale, E., Kelly, L., & Rohwer, F. (2009). Metagenomic analysis of stressed coral holobionts. *Environmental Microbiology*, 11(8), 2148–2163. <https://doi.org/10.1111/j.1462-2920.2009.01935.x>
- Todd, P. A., Sanderson, P. G., & Chou, L. M. (2001). Morphological variation in the polyps of the scleractinian coral *Favia speciosa* (Dana) around Singapore. *Hydrobiologia*, 444(1), 227–235. <https://doi.org/10.1023/A:1017570100029>
- Toledo-Hernández, C., & Ruiz-Díaz, C. P. (2014). The immune responses of the coral. *Invertebrate Survival Journal*, 11(1), Article 1.
- Traçon, M. L., Pratchett, M. S., & Penin, L. (2010). Comparative Effects of Different Disturbances in Coral Reef Habitats in Moorea, French Polynesia. *Journal of Marine Sciences*, 2011, e807625. <https://doi.org/10.1155/2011/807625>

- Traylor-Knowles, N. (2016). Distinctive wound-healing characteristics in the corals *Pocillopora damicornis* and *Acropora hyacinthus* found in two different temperature regimes. *Marine Biology*, 163(11), 231. <https://doi.org/10.1007/s00227-016-3011-y>
- Triwibowo, A. (2023). STRATEGI PENGELOLAAN EKOSISTEM TERUMBU KARANG DI WILAYAH PESISIR. *Jurnal Kelautan dan Perikanan Terapan (JKPT)*, 1(0), Article 0. <https://doi.org/10.15578/jkpt.v1i0.12048>
- True, J. D. (2012). *Salinity as a structuring force for near shore coral communities*.
- Tun, K., Chou, L. M., Cabanban, A., Tuan, V. S., Suharsono, T. Y., Sour, K., & Lane, D. (2005). *Status of coral reefs, coral reef monitoring and management in Southeast Asia 2004*. <https://digitalarchive.worldfishcenter.org/handle/20.500.12348/1894>
- Turner, S. J. (1994). The biology and population outbreaks of the corallivorous gastropod *Drupella* on Indo-Pacific reefs. *Oceanography and Marine Biology: An Annual Review*. <https://www.vliz.be/en/imis?refid=79593&basketaction=add>
- Tusa, E., Reynolds, A., Lane, D. M., Robertson, N. M., Villegas, H., & Bosnjak, A. (2014). Implementation of a fast coral detector using a supervised machine learning and Gabor Wavelet feature descriptors. *2014 IEEE Sensor Systems for a Changing Ocean (SSCO)*, 1–6. <https://doi.org/10.1109/SSCO.2014.7000371>
- UN Environment, ISU, ICRI, & Trucost. (2018). *The Coral Reef Economy: The business case for investment in the protection, preservation and enhancement of coral reef health*. <http://www.unep.org/resources/report/coral-reef-economy>
- van de Water, J. A. J. M., Ainsworth, T. D., Leggat, W., Bourne, D. G., Willis, B. L., & van Oppen, M. J. H. (2015). The coral immune response facilitates protection against microbes during tissue regeneration. *Molecular Ecology*, 24(13), 3390–3404. <https://doi.org/10.1111/mec.13257>
- Villon, S., Chaumont, M., Subsol, G., Villéger, S., Claverie, T., & Mouillot, D. (2016). *Coral Reef Fish Detection and Recognition in Underwater Videos by Supervised Machine Learning: Comparison Between Deep Learning and HOG+SVM Methods* (Vol. 10016, p. 171). https://doi.org/10.1007/978-3-319-48680-2_15
- Vaughan, M., & Vitousek, P. (2013). Mahele: Sustaining Communities through Small-Scale Inshore Fishery Catch and Sharing Networks 1. *Pacific Science*, 67, 329–344. <https://doi.org/10.2984/67.3.3>
- Vercammen, A., McGowan, J., Knight, A. T., Pardede, S., Muttaqin, E., Harris, J., Ahmadi, G., Estradivari, Dallison, T., Selig, E., & Beger, M. (2019). Evaluating the impact of accounting for coral cover in large-scale marine conservation prioritizations. *Diversity and Distributions*, 25(10), 1564–1574. <https://doi.org/10.1111/ddi.12957>
- Vermeij, M. J. A., Marhaver, K. L., Huijbers, C. M., Nagelkerken, I., & Simpson, S. D. (2010). Coral Larvae Move toward Reef Sounds. *PLOS ONE*, 5(5), e10660. <https://doi.org/10.1371/journal.pone.0010660>
- Veron, J. (2013). Overview of the taxonomy of zooxanthellate Scleractinia. *Zoological Journal of the Linnean Society*, 169(3), 485–508. <https://doi.org/10.1111/zoj.12076>

- Veron, J. E. N. (1995). *Corals in Space and Time: The Biogeography and Evolution of the Scleractinia*. Cornell University Press.
- Veron, J. E. N., Devantier, L. M., Turak, E., Green, A. L., Kininmonth, S., Stafford-Smith, M., & Peterson, N. (2009). Delineating the Coral Triangle. *Galaxea, Journal of Coral Reef Studies*, 11(2), 91–100. <https://doi.org/10.3755/galaxea.11.91>
- Vial, A., Stirling, D., Field, M., Ros, M., Ritz, C., Carolan, M., Holloway, L., & Miller, A. A. (2018). The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: A review. *Translational Cancer Research*, 7(3). <https://doi.org/10.21037/tcr.2018.05.02>
- Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience*, 2018, 7068349. <https://doi.org/10.1155/2018/7068349>
- Vucko, M. J., Poole, A. J., Carl, C., Sexton, B. A., Glenn, F. L., Whalan, S., & de Nys, R. (2014). Using textured PDMS to prevent settlement and enhance release of marine fouling organisms. *Biofouling*, 30(1), 1–16. <https://doi.org/10.1080/08927014.2013.836507>
- Wabnitz, C., Taylor, M., Green, E., & Razak, T. (2003). *From Ocean to Aquarium: The Global Trade in Marine Ornamental Species*.
- Wells, J. W. (1957). Corals. In *Geological Society of America Memoirs* (Vol. 67V1, pp. 1087–1104). Geological Society of America. <https://doi.org/10.1130/MEM67V1-p1087>
- Whalan, S., Wahab, M. A. A., Sprungala, S., Poole, A. J., & Nys, R. de. (2015). Larval Settlement: The Role of Surface Topography for Sessile Coral Reef Invertebrates. *PLoS ONE*, 10(2). <https://doi.org/10.1371/journal.pone.0117675>
- Wiley, V., & Lucas, T. (2018). Computer Vision and Image Processing: A Paper Review. *International Journal of Artificial Intelligence Research*, 2(1), Article 1. <https://doi.org/10.29099/ijair.v2i1.42>
- Wilkinson, C. (2004). Status of Coral Reefs of the World. Global Coral Reef Monitoring Network and Australian. *Institute of Marine Science, Townsville, 1*.
- Wilkinson, C. R. (1996). Global change and coral reefs: Impacts on reefs, economies and human cultures. *Global Change Biology*, 2(6), 547–558. <https://doi.org/10.1111/j.1365-2486.1996.tb00066.x>
- Williams, S. L., Sur, C., Janetski, N., Hollarsmith, J. A., Rapi, S., Barron, L., Heatwole, S. J., Yusuf, A. M., Yusuf, S., Jompa, J., & Mars, F. (2019). Large-scale coral reef rehabilitation after blast fishing in Indonesia. *Restoration Ecology*, 27(2), 447–456. <https://doi.org/10.1111/rec.12866>
- Wilson, S. K., Graham, N. A. J., & Polunin, N. V. C. (2007). Appraisal of visual assessments of habitat complexity and benthic composition on coral reefs. *Marine Biology*, 151(3), 1069–1076. <https://doi.org/10.1007/s00227-006-0538-3>
- Winters, G., Holzman, R., Blekhman, A., Beer, S., & Loya, Y. (2009). Photographic assessment of coral chlorophyll contents: Implications for ecophysiological studies and coral monitoring.

Journal of Experimental Marine Biology and Ecology, 380, 25–35.
<https://doi.org/10.1016/j.jembe.2009.09.004>

Witman, J. D., Dayton, P. K., Arnold, S. N., Steneck, R. S., & Birkeland, C. (2013) *Scuba Revolutionizes Marine Science*.

Yu, K.-H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering*, 2(10), 719–731. <https://doi.org/10.1038/s41551-018-0305-z>

Zawada, K. J. A., Madin, J. S., Baird, A. H., Bridge, T. C. L., & Dornelas, M. (2019). Morphological traits can track coral reef responses to the Anthropocene. *Functional Ecology*, 33(6), 962–975. <https://doi.org/10.1111/1365-2435.13358>

Zhang, C. (2015). Applying data fusion techniques for benthic habitat mapping and monitoring in a coral reef ecosystem. *ISPRS Journal of Photogrammetry and Remote Sensing*, 104, 213–223. <https://doi.org/10.1016/j.isprsjprs.2014.06.005>

Zhao, W. (2017). Research on the deep learning of the small sample data based on transfer learning. *AIP Conference Proceedings*, 1864(1), 020018. <https://doi.org/10.1063/1.4992835>

Zürcher, M. (2002). *02-HISTOIRE DE LA PLONGEE*.