Transcriptome data analysis of the *Millepora* (Hydrozoa: Milleporidae) species complex in Puerto Rico.

By
Ingrid Corasí Ortiz González

A thesis submitted on partial fulfillment of the requirements for the degree of
MASTER OF SCIENCES
in
Biological Oceanography

UNIVERSITY OF PUERTO RICO
MAYAGÜEZ CAMPUS
2018

Approved by:

Nikolaos V. Schizas, Ph.D. **Chairman, Graduate Committee**	**Date**
Ernesto F. Weil, Ph.D. **Member, Graduate Committee**	**Date**
Juan Carlos Martínez Cruzado, Ph.D. **Member, Graduate Committee**	**Date**
Carlos José Santos Flores, Ph.D. **Member, Graduate Committee**	**Date**
Ernesto Otero, Ph.D. **Director, Department of Marine Sciences**	**Date**
Lilliam Ramírez, Ms. **Representative of Graduate Studies**	**Date**

Abstract

Millepora is a relative rich-species genus of hydrocorals, with 18 species distributed around the globe and no shared species between the Atlantic and Pacific Ocean. It is considered one of the important reef building cnidarians. The current diversity of the Caribbean *Millepora* species consist of *Millepora complanata, M. alcicornis, M. squarrosa* and *M. striata* but there are century-old taxonomic uncertainties. Here, we report the *de-novo* transcriptome assembly and phylotrascriptomic analysis of *M. alcicornis, M. complanata, M. squarrosa* and a new ecomorph (*Millepora* sp.) found in exposed, shallow water, *Thalassia* beds and mangrove areas in southwest Puerto Rico. We obtained over 345 million reads from the transcriptomes of the four taxa (Illumina HiSeq4000; 2x150bp). The analysis pipeline consisted of assembly with Trinity, BUSCO for quality check, RSEM for TPM filtration, and ORF call for each transcriptome, prior to ontology and phylogenetic analysis. The ontology analysis was performed using Blast2GO, and genes were categorized by molecular functions, biological processes, and cellular components. The phylogenetic analysis was performed using distinct custom bash programs to select homologous sequences among the transcriptomes, resulting in 10,797 homologous sequences. The concatenation analysis (with either Maximum Likelihood or Bayesian inference) resulted in a topology supporting a clade of *M. complanata* and *M. alcicornis*; with *Millepora* sp., the ecomorph, outside the clade, and *M. squarrosa* as the outgroup. However, in a coalescence-based tree estimation analysis (using RAxML and ASTRAL-II), a different topology resulted, with *M. alcicornis* forming a clade with *Millepora* sp. rather than with *M. complanata*. ASTRAL-II analysis indicated that there is a very high degree of incomplete lineage sorting, suggesting a very recent time of divergence among these three out of the four Caribbean *Millepora* species. Of the three Caribbean species and the ecomorph considered in this

analysis, *M. squarrosa*, is the only species that can be differentiated readily as a genetically distinct milleporid species.

Resumen

Millepora es un género relativamente rico en especies, con 18 especies de hydrocorales reconocidos y distribuidos alrededor del mundo, pero sin especies compartidas entre el Océano Atlántico y el Océano Indo-Pacífico. Dentro de los cnidarios, el género *Millepora* es considerado un constructor importante de arrecifes. Las especies reconocidas de *Millepora* en el Caribe son *Millepora complanata*, *M. alcicornis*, *M. squarrosa* y *M. striata*, pero aun después de muchos años, todavía existen dudas y problemas con la taxonomía de algunas especies. Este estudio presenta nueva información a partir del ensamblaje *de-novo* del transcriptoma y filo-transcriptoma de tres de las especies comunes en Puerto Rico, *M. alcicornis*, *M. complanata*, *M. squarrosa*, y un nuevo ecomorfo (*Millepora* sp.) encontrado en el área somera, expuesta de *Thalassia* y manglares en el suroeste de Puerto Rico. Se obtuvieron más de 345 millones de secuencias de los trasncriptomas de las tres especies y el nuevo ecomorfo de *Millepora* (Illumina HuSeq4000; 2x150bp). El flujo del análisis consistió en el ensamblaje del transcriptoma en Trinity, se verificó la calidad con BUSCO, RSEM para filtrar por TPM, y buscar los Marcos de Lectura Abiertos en los transcriptomas antes del análisis de ontología y filogenia. Se utilizó Blast2Go para hacer el análisis de ontología y los genes fueron clasificados por función molecular, procesos biológicos y componentes celulares. El análisis filogenético se realizó utilizando distintos programas en “bash” para seleccionar las secuencias homólogas entre los transcriptomas, dando como resultado 10,797 secuencias homólogas. El análisis de concatenación (Máxima Verosimilitud e Inferencia Bayesiana) resultó en una homología respaldando un clado con *M. complanata* y *M. alcicornis*, con *Millepora* sp. fuera del clado y *M. squarrosa* como el grupo externo más diferente. Sin embargo, el análisis de incorporación basado en la evaluación de árboles filogenéticos (utilizando RAxML y ASTRAL-II), resultó en

un esquema distinto, con *M. alcicornis* formando un clado con *Millepora* sp. en lugar de que con *M. complanata*. El análisis de ASTRAL-II indicó que hay un alto grado de clasificación incompleta del linaje, sugiriendo que tres de las cuatro especies de *Millepora* divergieron recientemente. La cuarta especie, *M. squarrosa* es la única que se puede diferenciar como una especie aparte.

Copyright

Dedication

To my Family.

“I never been hurt by a sea creature, except for jellyfish and sea urchins”

-**Peter Benchley**

Acknowledgments

THANK YOU!

Special thanks to my Family and Friends.
Thanks to Juan J. González Lagoa because without you I would never be here. Thank you because you said to me "mira lo que hay en una gota de agua, imagínate todo lo que hay en el océano" and that motivated me to become a Marine Scientist. To Ms. Thiele, who taught me science from the early beginning in my life and gave me the opportunity to meet Dr. González. To Dr. Carlos Santos and Ana Estrella, for being there in my undergraduate steps, teaching me how to do a Graduate Research and for the opportunity to do and present my research. More than "chiefs", you have been leaders and friends. Thanks for always being there.
Thanks to all my friends and Lab partners who contributed something to my Master's. From sharing a beer, a cup of coffee or tea, to demanding work in the lab. **Biology**: Juan Gi, Goldo, Danny, Ana, Wilma, Arleen, Mara, Eneilys, Dr. Alejandro Ortiz, Edgar. **CiMa**: Luis (taught me how to do extractions), Alexandra (Travel Partner), Martha, Héctor, Duane (Capitán ☺), Jaaziel (Collecting), Brenda, Suhey, Pomales, Liajay, Ramón, Estefanía, Alex, Hammerman, Gualberto, Jenniffer (Lela) and, of course, Yizus. As for students visiting the lab, I would like to thank Solène from the IUT Génie Biologique for helping me doing the small *Millepora* DNA project to the SMBE 2014.
I will be always appreciative all the workers and administrative people of Magueyes. Lili and Zulma for giving me a job and being good friends. Bill Rosado and Martha Ricaurte helped me with the RNA extractions (I scream **EUREKA** because both of you helped!). Special thanks to Orlando; without you this would not be possible, thank you for encouraging me to go diving, for helping me collecting my samples and having the tank. Negrito (R.I.P.) and Lalo for their help (Thanks for the Spot Light those days I stayed all night doing my extractions), and to all the Magueyes workers that became my friends.
Iván Fernando, thanks for being my friend and helping me in the Pauley Program (HIMB, Coconut Island); you taught me a lot of things about NGS. Every time that I asked about genomics you were there to answer my concerns.
THANKS to Ramón!!! Yes, my bioinformatician: "The person that knows how to talk with the computer to make what we need". Without your help, I will still be learning how to process the samples. Of course, The Bioinformatic Team (Hammerman, Ramón and I), so many meals and nights we shared in the server.
Thanks to my graduate committee.
Special thanks to those who sponsored my research: Sea Grant with the Seed Money, to Art and Sciences Dean for the Seed Money, and to Dr. Richard Appeldoorn (CCRI) for financial support to sequence my samples. Thanks to the ASLOMP for making me part of the program in the ASLO conference in Hawaii 2017 to present the research that became my first paper: Ortiz-González *et al*. (2017).
Thanks to my Family for being there when I need it and when I can't see you. Thanks for supporting me in my choices and goals. Mom, thanks for being there, always, without you I would never be here. Yari, thanks because you are my love, who encourage me, my example to follow. Mamá and Papá for supporting me, making me food so I did not lose time cooking, and for being there all my life. Papi, thanks for supporting me.

Chelo, you have been here in the bad and goods and sharing the calories. Thank you for supporting me even with my bad mood and my existential crisis. No matter what, you have been here during my Bachelor and my Masters.

Table of Contents

List of Tables

List of Figures

Figure 22: Sampling site of *Millepora alcicornis* and *Millepora complanata*, Coral Key, La Juagla Reef, Enrique Key, Playita Azul Reef.

1. Introduction

Species of the genus *Millepora* are commonly known as fire corals because of the "stinging" feeling upon touching their surface. This reaction is produced by toxins injected with their nematocysts. Nematocysts are large, stinging cells characteristic of the phylum Cnidaria. The genus *Millepora* consists of about 18 species distributed worldwide along the tropical/subtropical seas, inhabiting coral reefs, at depths between 0.5m and 50m (Lewis, 2006; Boshma, 1948). *Millepora* species usually grow upright in thin finger-like branches, leaf-like, or blade-like patterns. *Millepora* spp. may form *Millepora*-dominated reefs but inhabit other shallow tropical habitats, such as rocky shores, seagrass beds, mangrove roots; they can also be found competing or overgrowing living corals, gorgonians, zoanthids and sponges (Banaszak *et al*., 2003; Lewis, 2006; Wahle, 1980; Ruiz-Ramos *et al*., 2014). The three-dimensional structure of *Millepora* provides habitat for species such as small fish, ophiuroids, barnacles, and micro-crustaceans, among others (Martins *et al*., 2008). In the Caribbean, milleporids are commonly found in shallow and turbulent waters forming dense reef rims contributing to the stabilization and complexity of the carbonate structure (Edmunds, 1999; Lewis, 2006). *Millepora* is adapted to many environmental gradients of reefs such as light, temperature, water movements, currents or turbidity (de Weerdt, 1984).

The milleporids have received little attention in coral reef studies in comparison with the scleractinian corals (Edmunds, 1999; Lewis, 2006), even though they have important structural and ecological roles. Lewis (1991) estimated that large colonies have a high probability of fragmentation with higher survival capacity of the fragments, resulting in high abundances. The highest densities of *Millepora* colonies are observed in less than 5m depth (Lewis, 2006).

Millepora complanata is a voracious plankton feeder (Lewis, 1989) and in areas of dense thickets, they play a significant role in regulating copepod aggregations and abundance (Lewis, 1992a).

Millepora colonies are gonochoric, each colony releases a male or female medusa, carrying one sperm sac or three to five eggs, respectively (Lewis, 2006; Soon and Cho, 1998). The medusa release begins in the afternoon and continues during the night; the male medusa is released first followed by the female medusa (Soon and Cho, 1998).

There are 18 recognized species of *Millepora*, with no shared species between the Atlantic and the Pacific Ocean (Lewis, 2006; Amaral *et al*, 2008). *Millepora alcicornis* (Linneaus 1758) is the most conspicuous, abundant, and locally and geographically widely dispersed hydrocoral species in the Atlantic. It has an extensive distribution from the tropical/subtropical western Atlantic Ocean, including Brazil, to the west coast of Africa in Cape Verde Archipelago, and Ascension Island (Amaral *et al*., 2008; Clemente *et al*., 2011; deWeerdt, 1984; Hoeksema *et al*., 2014; Lewis, 2006; Morri *et al.*, 2000). In Brazil, three other milleporid species can be found: *M. brazilensis* (Verrill 1868), *M. nitida* (Verrill 1868), and *M. laboreli* (Amaral 2008) (Amaral *et al*., 2008; Lewis, 2006). The three remaining *Millepora* species are Caribbean-specific; *M. striata* (Duchassaing and Michelotti 1864), which is restricted to the southwest Caribbean, *M. complanata* (Lamarck 1816), and *M. squarrosa* (Lamarck 1816) with low density populations and limited depth distribution (de Weedt, 1984; Lewis, 2006; Ruiz-Ramos *et al*., 2014).

Despite the relatively small number of species, the taxonomy of the genus *Millepora* is unsettled (Boshma, 1948). Most of the taxonomic studies have been conducted solely on morphological characters. The size of gastropores, dactylopore to dactylopore distance, the presence or absence of the ampullae, and the shape, texture, and character type of the corallum were among the

characters used. These characters however, exhibit high degree of intra- and interspecies variability obfuscating the clear separation of species based on morphology (Amaral *et al*., 2002; Boshma, 1948; de Weerdt, 1984; Lewis, 1989; Razak and Hoeksema, 2003; Tepper *et al*., 2012; Ruiz-Ramos *et al*., 2014). Descriptions of new species of *Millepora* are still based exclusively on morphological characters (see the latest new species *M. laboreli*; Amaral *et al*., 2008). Recently, Tepper *et al*. (2012) and Ruiz-Ramos *et al*. (2014) used both morphological and molecular markers to address the species problem of *Millepora* but neither study reached firm conclusions about the *Millepora* phylogeny. Tepper *et al*. (2012) could not distinguish between *M. alcicornis*, *M. complanata* and the intermediate morphotypes and Ruiz-Ramos *et al*. (2014) could delineate only *M. squarrosa* as a distinct genetic specie and proposed a *Millepora* species complex consisted of *M. alcicornis*, *M. complanata* and *M. striata*.

The taxonomic issues of *Millepora* still need to be resolved. Lack of knowledge of the real number of species of a widespread taxon that contributes significantly to the calcium carbonate structure of the reefs could undermine the efforts of conservation agencies to implement new regulations on the management of coral reefs and can also limit the value of scientific surveys recording the changing profile of coral reefs. Next Generation Sequencing (NGS) methods are now capable of delivering thousands of genetic markers with no prior genetic information available (Davey *et al*., 2011). NGS methods are ideal for a non-model taxon (species with no developed research tools including genomic ones) such as *Millepora*. Recently, the transcriptome of *M. alcicornis* was published by Ortiz-González *et al*. (2017), showing the potential of NGS methods to deliver gene expression data in a *Millepora* taxon with no available genomic information. They compared the unique genes of *Millepora* against other cnidarian transcriptomes, especially genes of the different zooids (modules) and toxin-related genes, and

they found that all genes reported to the dactylozooid, gastrozoid and gonozoid polyps of *Hydractinia symbiolongicarpus* and the ones described for the scyphomedusa of *Aurelia aurita*.

One of the NGS applications is phylotranscriptomics, which has revolutionized phylogenetics. Phylotranscriptomics is the comparison of orthologous genes (genes found in two or more species that can be traced to the same common ancestor) obtained through sequencing of the transcriptome. The transcriptome is the complete set of transcripts (type and quantity) that a cell has (Nagalakshmi *et al*., 2010). In other words, the transcriptome is all the genes expressed in the form of RNA at a particular moment, and its sequencing can be accomplished through sequencing of the complementary DNA (cDNA). The use of transcriptomics data of non-model organisms can produce new questions and phylogenetic insights (Oakley *et al*., 2013). A phylotranscriptomics study reconstructed the phylogeny of selected molluscans using 308 genes, 41% of the genes expressed in the transcriptome (Kocot *et al*., 2011). Weigert *et al*, (2014) reconstructed the basal annelid relationships using a transcriptomic approach with 622 genes. More recently, Zapata *et al*. (2015) used a combination of genomic and transcriptomic information (a total of 1262 genes) of different groups of Cnidaria to decipher the phylogenetic relationships of the Phylum.

The goal of this project is to compare phylogenetically homologus genes of *Millepora* obtained through transcriptome sequencing (phylotranscriptomics). Here, we report the transcriptome of three species and one ecomoprh of Caribbean *Millepora,* and the phylogenetic relationships obtained from this data set.

2. Literature Review

2.1 Classification and distribution

Domain Eukarya

Opistokonta

Kingdom Metazoa

Eumetazoa

Phylum Cnidaria

Class Hydrozoa

Sub Class Hydriodolina

Order Anthothecata

Sub Order Capitata

Family Milleporidae

Genus *Millepora*

2.2 Distribution

Millepora consists of 18 species with no shared species between the Pacific and Atlantic Ocean (Lewis, 2006). Seven species can be found in the western Atlantic, including the Caribbean, and 11 species in the Pacific and Indo-Pacific Ocean.

2.2.1 Pacific and Indo-Pacific Ocean

In the Pacific and Indo-Pacific region, there are eleven accepted species: *M. exaesa* (Forskal 1775); *M. dichotoma* (Forskal 1775); *M. platyphylla* (Hemprich and Ehrenberg 1834); *M. intricata* (Milne-Edwards 1857); *M. murrayi* (Quelch 1884); *M. tenera* (Boshma 1948); *M. latifolia* (Boshma 1948); *M. faveolata* (Crossland 1952); *M. tuberosa* (Boshma 1966); *M. xishaensis* (Zou 1978); and *M. boschmai* (de Weerdt and Glynn 1991).

2.2.2 Atlantic Ocean

In the Atlantic Ocean, *M. alcicornis* (Linnaeus 1758) is the most extensively distributed species, found in tropical and sub-tropical areas. Three of the *Millepora* species are distributed solely in Brazil: *M. laboreli* (Amaral 2008), *M. nitida* (Verrill 1868), and *M. brasilensis* (Verrill 1868). Three other recognized species are found in the Caribbean: *M. complanata* (Lamarck 1816); *M. striata* (Duchssaing and Michelotti 1834), and *M. squarrosa* (Lamarck 1816).

These numbers of *Millepora* species presented here have been listed in Lewis (2006) and include the most recent described species in 2008 (*M. laboreli*), however, the true number of species is still controversial.

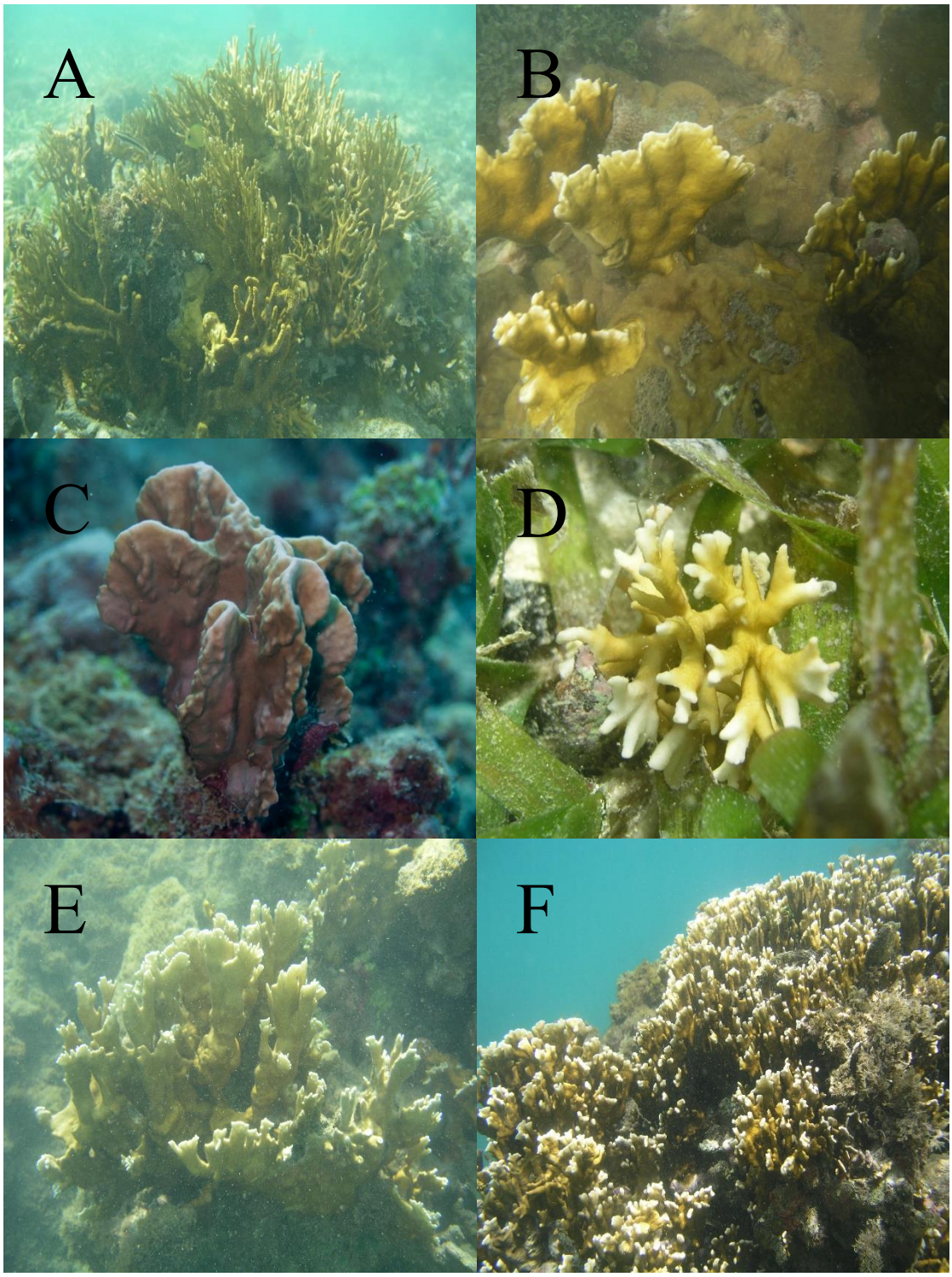


Figure 1: A) *Millepora alcicornis* from La Jungla, Guánica. **B)** *Millepora complanata* from Playita Azul, Cabo Rojo. **C)***Millepora squarrosa* from Turrumote II reef, southwest Puerto Rico. **D)** *Millepora* sp. surrounded by the seagrass *Thalassia testudinum* from Mata Seca Key, southwest Puerto Rico. **E** and **F)** *Millepora* from Puerto Rico Reefs showing different growth forms, a plate-like base and finger-like tips with no clear *M. complanata* or *M. alcicornis* morphology

2.3 Puerto Rico *Millepora* species description

2.3.1 *Millepora alcicornis*

Millepora alcicornis (**Figure 1, A**) was first described by Linnaeus in 1758. Boshma (1948), and de Weerdt (1981; 1984) described *M. alcicornis* as a colony with finger-like branches, oriented perpendicular to the water flow (Wahle, 1980). The colony can be ornamented with fragile branches, bushy branches, or display a more compact growth and even an encrusting form (Boshma, 1948; Lewis, 2006; Ruiz-Ramos *et al*., 2014). *Millepora alcicornis* is commonly seen overgrowing gorgonians, rocks, mangrove roots, or piers (Wahle, 1980; Banaszak *et al*., 2003). It can be found from 0 to 50 meters depth (Lewis, 2006). It has attained an extensive distribution from the tropical/sub-tropical western Atlantic Ocean, including Brazil, to the west coast of Africa in Cape Verde Archipelago, and Ascension Island (Amaral *et al*., 2008; Clemente *et al*., 2011; de Weerdt and Glynn, 1991; Hoeksema *et al*., 2014; Morri *et al*., 2000).

2.3.2 Millepora complanata

Millepora complanata (**Figure 1, B**) was described by Lamarck in 1816. It was reported by de Weerdt (1984) as having a limited geographic distribution in the Atlantic Ocean, but it can be found throughout the Caribbean, except in Bermuda (Mattraw, 1969). Fitt (2012) investigated *M. complanata* bleaching in Florida, Jamaica and Florida Keys. Fenner (1999) reported the species in Cozumel (México) and Belize; Lewis (1991, 1992a) reported it from Barbados; Strömgren (1976) recorded it in Jamaica; Tepper *et al*. (2012) collected *M. complanata* from Bahamas, and

Ruiz-Ramos *et al*. (2014) from Puerto Rico. *Millepora complanata* can be abundantly seen also in Aruba, Bonaire and Curaçao (Schizas and Weil, pers. obs.). *Millepora complanata* forms plate-like and foliose colonies with little or no rugosity. Lewis (1996) described *M. complanata* as an amorphous or not-easily defined shape, with irregular encrusting bases and large, wide, upright blades. It displays linear growth marks and can be found from 0 to 10m deep (Lewis, 2006).

2.3.3 Millepora squarrosa

Millepora squarrosa (**Figure 1, C**) was described by Lamarck 1816. It is reported in the eastern Caribbean and Brazil (de Weerdt, 1984; 1990). *Millepora squarrosa* was classified as an uncommon hydrocoral with a smaller (15 to 25 cm) plate-like colony with lateral expansions and rugose surface (de Weerdt, 1990; Lewis, 1996; Fenner, 1999). Boshma (1948) described the species as a honey-combed complex and Lewis (1996) described *M. squarrosa* as a colony with thick united plates, found between 0 to 20 m deep.

2.3.4 *Millepora* sp.

Several colonies of a different ecomorph of *Millepora* showing different morphology (**Figure 1, D**) were found by Dr. Ernesto Weil (University of Puerto Rico at Mayagüez), in several shallow, exposed *Thalassia testudinum* seagrass beds and near mangrove areas in the Natural Reserve of La Parguera in the southeast coast of Puerto Rico. The first colonies were observed in Mata la Seca (**Appendix, Table 6** Colonies are small, with open or compact branching, looking like miniature *M. alcicornis*, but with thin, short, fragile and pointed or flat-end dichotomous branches. The largest colony observed was about 30 cm in diameter and 20 in height. Colonies are usually found growing within the sea grass, on sandy areas and are not

attached to the substratum. The sting intensity from their nematocyst's toxins is lower than that of *M. alcicornis* and *M. complanata,* similar to *M. squarrosa.* Encrusting barnacles can be seen on the principal branches, and crabs are also living within the structure. No morphological or histological characterization has been made on this morphotype. It will be called *Millepora* sp. for the purposes of this work.

2.4 Feeding and nutrition

All species within the genus *Millepora* have a symbiotic relationship with dinoflagellate algae (*Symbiodinium* and *Gloeodinium* spp.), which contribute significantly to the metabolic needs of the host through carbohydrates resulting from photosynthesis. Banaszak *et al*. (1993) identified the algae in association with *M. dichotoma* as *Gloeodinium viscum*. Schowald *et al*. (1997) estimated that more than 70% of the carbon fixed by photosynthesis was translocated to the hydrocoral host, *M. dichotoma*. However, the studies demonstrated photosynthesis by itself could not meet the carbon demands of *Millepora*. The hydrocoral need additional sources of energy for their metabolic demands to grow, feed, compete and reproduce.

An additional way for hydrocorals to obtain energy is through heterotrophy. *Millepora* is a considerable voracious plankton feeder (Abe, 1938; de Kruijf, 1975; Lewis, 1992a; Lewis, 2006). Lewis (1992a; 2006) showed that *M. complanata* can capture 7.1±4.1 zooplankton prey $\bullet cm^{-2}$ of colony surface, each day. The primary prey of *M. complanata* were copepods, composing as much as 63% of the diet. The digestion rate of *M. complanata* was 36% in six hours and 91% in 24 hours. The lower digestion rate compared to scleractinians is most likely related to the small size of hydrocoral zooids.

2.5 Reproduction

Milleporids have both sexual and asexual reproductive strategies. During the sexual cycle, they develop a sessile (polyp) and free-living swimming stage (medusa); the latter is released during spawning. Fire corals reproduce asexually through budding and by fragmentation. Asexual reproduction through fragmentation in hydrocorals is substantial, especially during storms (Edmunds, 1999; Lewis, 2006), and may contribute more to species propagation than all other reproductive modes.

2.6 Stressors- Bleaching and predation

Milleporids are vulnerable to global warming and are among the first cnidarians to lose their photosynthetic symbionts when water temperatures increase (Williams and Buckley-Williams, 1990). Although the photosynthetic rate in *Millepora* spp. was dramatically reduced during the 1973 bleaching event, these colonies recovered after 14 weeks (Fitt, 2012; Jaap, 1979), but recovery can take as long as one year (Chiappone and Sullivan, 1994). In the south coast of Puerto Rico for example, all *Millepora* populations suffered massive mortalities during and after the massive bleaching and disease outbreak events of 2005-2006, with one species, *M. squarrosa*, disappearing from reefs (locally extinct) off La Parguera Natural Reserve for almost ten years (Weil *et al*., 2009; Weil and Croquer, 2009). Only recently have colonies being appearing in a few reefs (Weil, pers. obs.). Diseases and other health issues have also been observed affecting Caribbean *Millepora* species in many localities sometime producing significant mortalities (Weil and Rogers, 2011).

Few species are known to prey on *Millepora* and most likely the impact is negligible compared to bleaching or habitat destruction. Among them, the corallivore gastropod *Caralliophila abbreviata* is commonly seen feeding on *Millepora* (Lewis, 1992). The fireworm *Hermodice carunculata* is a common predator on *M. alcicornis* and *M. complanata* (Lewis,

2006; Pérez and Gomez, 2012). The damselfish *Microspathodon chrysurus* feeds on *Millepora* in Florida and Panama (Ciardelli, 1967; Glynn, 1973), and Randall (1967) found *M. alcicornis* fragments in the stomach of the fishes *Aluteras cripta* and *Cantherines macrocerus*.

2.7 *Millepora* as habitat for other organisms.

Millepora species form large carbonate structures on reefs and provide habitat for many invertebrate species (Brander *et al*., 1971; Hickson, 1891; Ross, 1999; Lewis, 1992a; 2006). Martins *et al*. (2008; 2009) found 95 non-colonials and 26 colonial species on *Millepora*. Based on Martins *et al.* (2008; 2009), the most commonly occurring organisms associated with *M. alcicornis* were Crustacea (93%), Polychaeta (76%), and Mollusca (69%). Barnacles of the genera *Savignium*, *Balanus* and *Megabalanus* (Roos and Newman, 1973; Lewis, 1992a; Soon and Changlai, 1992; Ross, 1999) are obligatory epizooans of hydrocorals (Pasternak *et al*., 2001). Lewis (1992a) estimated that the abundance of the barnacle *Balanus stultus* in *M. complanata* was 2.5 barnacles per colony. Dead colonies of *Millepora* provide habitat for boring tanaid crustaceans (Lewis, 1998; 2006). Other associates of *Millepora* are polychaetes of the family *Serpulidae (*Hunte *et al*., 1990; Marsden, 1992), which grow in tubes on the surface of the corals (Lewis, 2006), and of the family *Spionidae* which burrow in *Millepora* (Blake, 1996; Lewis, 1998; Lewis, 2006); all potentially reducing the feeding potential of the hydrocoral host.

3. Methods

3.1 Sampling collection and maintenance

Four colonies of each one of the three recognized species, *M. alcicornis, M. complanata, M. squarrosa* and *Millepora* sp. (the new ecomorph), were collected from different localities off the La Parguera Natural Reserve on southwest coast of Puerto Rico. *Millepora alcicornis* and *M.*

complanata were collected from the reefs Corral, Playita Azul, Enrique and La Jungla (**Figure 2, Appendix Table 3 and Appendix Table 4**). *Millepora squarrosa* was collected from Turrumote II (**Figure 2, Figure 20 and Appendix Table 5**) and *Millepora* sp. was collected from Mata Seca reef (**Figure 2, Figure 21, Appendix Table 6**). All samples were collected under a research permit from the Department of Natural and Environmental Resources of Puerto Rico (O-VS-PVS15-MA-00016-26092014) and sampling was conducted under their outlined auspices and regulations. Using a combination of SCUBA diving and snorkeling, colony fragments were collected from between 0-5 m depth. Sixteen samples from the same number of colonies were collected. All samples were placed a closed aquarium system for a period of eight weeks to limit gene expression driven by ambient variability (e.g. location, temperature, salinity), trying to simulate natural conditions. The aquarium was maintained with cycles of 12 hours of broad light spectrum (covering all light covering all light photosynthetic wave lengths) followed by six hours of LED moonlight and six hours of no light. Conditions in the aquarium were maintained at 27°C, salinity of 34 ppt, pH of 8.2, and nitrate levels of 0 mg/L. Water lost by evaporation was replaced when needed and five-gallon water changes were done every week. After the fire coral colonies were placed securely in the aquarium, two batches of copepods were added and two batches of *Artemia salina* were added for nutrient replenishment.

Figure 2: Southwest Puerto Rico. The green balloons indicate the sampling sites of all *Millepora* collected during this study.

3.2 Molecular techniques and sequencing

After the eight-week period in the aquarium, each colony was stored for one hour at -80°C before RNA extraction. Total RNA for *M. complanata* and *M. squarrosa* was extracted using the Qiagen RNA Extraction Kit. For *M. alcicornis* and *Millepora* sp., total RNA was extracted using the TRIZOL RNA Isolation Method (Chomczynski and Mackey, 1995), since initial extractions with the Qiagen RNA Extraction Kit yielded low RNA concentrations. RNA concentrations were measured using a NanoDrop2000. A total of four samples of each species were sent to the Genomics Sequencing and Analysis Facilities (GSAF) at the University of Texas at Austin, where quality assessment was performed using a 2100 Bioanalyzer Instrument (Agilent Inc.). The libraries were prepared combining equimolar quantities of each colony extraction for each species. The library was prepared using the RNA low cost high throughput and Poly-A mRNA capture methods. The quality and quantity of RNA, after the library preparation, were checked

again with the 2100 Bioanalyzer Instrument followed by sequencing on one lane of an Illumina HiSeq 4000 platform with 150bp paired-end reads.

3.3 Transcriptome assembly

The program FastQC v0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used for quality assessment of reads, and possible adapters and barcode contamination. Adapters were removed using the Cutadapt 1.9.1 (Martins, 2011) script followed by another run of FastQC to check the quality change. FastX-Toolkit 0.0.14 was used to trim the last 10bp of each read (http://hannonlab.cshl.edu/fastx_toolkit/index.html), generating high quality reads. *De novo* assembly was performed using Trinity 2.1.1 (Grabherr *et al*., 2011) with a minimum contig length of 200bp (k=200). Reads from the assembled transcriptome were checked for possible contaminant and vector sequences using the NCBI TSA (Transcriptome Shotgun Assembly) check steps for submission of transcriptomes. RSEM 1.2.31 (Li and Dewey, 2011) was used to normalize the transcripts for sequencing depth and gene length and filter the data by a value of 3 TPM (Transcripts Per Million) or higher. The TPM method was selected instead of RPKM (Reads Per Kilobase of transcript per Million mapped reads) and FPKM (Fragments Per Kilobase of cDNA per Million Fragments Mapped) because the mean expressed transcripts may vary between samples as explained by Li *et al*. (2010). BUSCO 1.22 (Simão *et al*., 2015) was also utilized to assess the transcriptome completeness using the BUSCO metazoans single-copy orthologs database.

3.4 Functional annotation

Open reading frames (ORFs) from the filtered transcripts were obtained using TransDecoder 3.0.0 (http://transdecoder.github.io) with the protein sequences of the hydrozoan *Hydra vulgaris*

(UniProt) as reference and a minimum ORF length of 375bp. All the ORFs were used to generate Blast searches, annotation and gene ontology assignments with the program Blast2GO 4.0 (Conesa *et al*., 2005).

3.5 Expressed genes analysis and comparisons with other Cnidaria

To compare the results of the Blast2Go with related organisms, we compared the *Millepora* transcriptomes with the transcriptomic results obtained by Iguchi *et al*., (2007), Brekhman *et al*. (2015), and Sanders *et al*. (2014). Ignuchi *et al*. (2007) described the cytoxine protein (MCTx-1) sequences of the *Millepora* found in nematocysts and Sanders *et al.* (2014) described proteins found in the dactylozooid, gastrozooid and gonozooid polyps of the hydrozoan *Hydractinia symbiolongicarpus*. We also used results from the stage specific transcriptome study of *Aurelia aurita* (Brekhman *et al*., 2015) and compared them to those found during the brief medusoid stage of *Millepora*.

3.6 Phylogenetic Analysis

All the ORFs obtained from TransDecorder were used as input for an All-versus-All-Blast search of the proteins. A custom bash script was created to search the sequences that had >70% similarity, an E-value lower of $1e^{-5}$ and meet the requirement of the reciprocal match. The sequences that met these criteria were clustered in separate files, with a representative of each species of *Millepora,* using a script. All *Hydra* protein sequences, available at the moment in UniProt, were downloaded for further comparison with the *Millepora* proteins. A bash program was used to find the *Hydra* sequences that had more than 25% of similarity with the *Millepora* sequences via Blast. Results of these searches were put it in the same file with the other four *Millepora* sequences. These sequences were then used for the phylogeny reconstruction. For further comparison, we also performed these analyses without *Hydra* as the outgroup, for an in-

depth analysis of the new ecomorph, *Millepora* sp., a potential new species.

3.6.1 Concatenation of DNA sequences

For the concatenation procedure, we aligned each file separately using the alignment program MAFFT (Katoh, 2013). Then, each alignment was concatenated into a separate file per species using a custom bash script. After each file was created, we then concatenated all species in one file for performing the super alignment using MAFFT in CIPRES (Miller *et al*., 2010). We used the program ProtTest (Abascal *et al*., 2005) to select the best model for amino acid substitution for phylogenetic analysis. We ran the program twice, once using all the *Millepora* proteins with *Hydra* as outgroup and another time without *Hydra*. The substitution models obtained were then applied to each program, if available. In cases were the model was not available in the program, we used the model GTR+GAMMA. We performed Maximum likelihood searches using the program RAxML (Stamatakis, 2014) with the model LG+G+F (Le and Gascuel, 2008) for the dataset with *Hydra* as an outgroup, and the model JTT+I+F (Jones *et al*., 1992) for the dataset without *Hydra*. We ran the program in each dataset using the rapid bootstrapping method search option (-f a) with 1,000 bootstraps. The Bayesian analyses were done using the program MrBayes (Ronquist *et al*., 2012), using the same models (if available) as applied in the RAxML program. We run the analyses for 10,000,000 generations with a sample frequency every 10,000. All other parameters were left in default. Convergence was evaluated with the program Tracer (http://tree.bio.ed.ac.uk/software/tracer/). All trees were visualized using the FigTree v1.4.3 program (http://tree.bio.ed.ac.uk/software/figtree/).

3.6.2 Coalescence

For the coalescence analysis, we used only the dataset with *Millepora* sequences. We ran RAxML for every single gene file using the model GTR+GAMMA and the rapid bootstrapping

method (-f a) with 1,000 bootstraps. All the gene trees generated along with their bootstrap replications were used as input for the program Astral-II (Mirarab and Warnow, 2015). Astral was run with the full annotation mode (-t 4) and 100 replicates. All other parameters were left on default. All trees were visualized using FigTree.

4 Results

4.1 Sequencing Data

The sequences were checked for quality and reads content. All four transcriptomes contained reads with inter-quartile (25%-75%) range over 38, with the whiskers (10%-90%) over 27, in the last nucleotide bases. *Millepora* sp. yielded the highest quantity of paired reads (96,215,401). On the contrary *M. alcicornis* yielded the lowest paired reads (76,488,682) (**Table 1**).

4.2 Assembly

After all transcriptomes were assembled in Trinity, *M. alcicornis* yielded the biggest assembly with 479,982 contigs and a N50 of 749, even though its sequencing yielded the lowest number of reads (Ortiz-González *et al*., 2017).

4.3 Transcriptome quality assessment

The quality and completeness of the transcriptomes were assessed with BUSCO. We searched for the Metazoans Orthologs sequences that were present in our assemblies and compared them to the total amount of orthologs shared by metazoans. All the *Millepora* transcriptomes had similar numbers of complete, missing, fragmented and duplicated orthologs (**Table 1**). For example, *M. squarrosa* has 466 completed orthologs, 122 duplicates, 22 fragmented, and another 353 orthologs were missing. The hydrocoral sample with the most complete orthologs, and least missing/fragmented orthologs is *M. squarrosa* (**Table 1**)

4.4 Normalization

After filtration was done with RSEM, we obtained the fewest sequences from *Millepora* sp. and *M. alcicornis*, even though the assemblies yielded the larger contigs (467,089 to 38,200 for *Millepora* sp. and 476,422 to 38,371 for *M. alcicornis*).

Sequences and filtering				
Sequencing Reads (R1, R2)	Number Sequences	Nucleotides (Mbp)	Median length	
M. alcicornis	76,488,682	22,421	150	
M. complanata	81,570,589	23,931	150	
M. squarrosa	90,632,750	26,589	150	
Millepora sp.	96,215,401	28,262	150	
Trinity Assembly				
Contigs (k>200)	Contigs	Nucleotides (Mbp)	Median length	N50
M. alcicornis	479,982	265	327	749
M. complanata	399,669	260	359	1,043
M. squarrosa	392,443	266	376	1,088
Millepora sp.	467,439	271	335	834
BUSCO				
Total BUSCO 843	Complete	Complete and Duplicate	Fragmented	Missing
M. alcicornis	414	88	29	400
M. complanata	417	105	26	400
M. squarrosa	466	122	24	353
Millepora sp.	380	102	33	430
RSEM				
	Before	After (TPM=3)		
M. alcicornis	476,422	38,731		
M. complanata	399,407	45,408		
M. squarrosa	392,082	44,082		
Millepora sp.	467,089	38,200		
TransDecorder				
	Reference organism	Median length (in amino acids)	ORFs	Complete ORFs
M. alcicornis	*Hydra vulgaris*	125	16,024	7,551
M. complanata	*Hydra vulgaris*	125	21,051	9,905
M. squarrosa	*Hydra vulgaris*	125	24,040	11,564
Millepora sp.	*Hydra vulgaris*	125	20,986	9,174

Table 1: Transcriptome information, data, assembly, quality, filtration and ORFs.

4.5 Open Reading Frame (ORF) prediction

Using the RESM output for the TransDecorder program, we concluded that *M. squarrosa* is the taxon with the most ORFs, including complete ORFs (**Table 1**). On the contrary, *Millepora alcicornis* was the taxon with the fewest ORFs (<20,000).

4.6 Gene ontology (GO)

In the Blast2GO steps, a Blast search against other cnidarians was performed as the first step. The best Blast hits (=sequence matches) of all four transcriptomes were against the hydroid *Hydra vulgaris* and the sea anemone *Nematostella vectensis* (**Appendix Figures 10-13**) with more than four thousand hits against each organism. Other cnidarians hits were present but with less than two hundred hits. The number of Blast hits was between 8,000 to 11,000 genes with the four transcriptomes.

Orthology hits against molecular processes, biological processes, and cellular components of the functional annotation classifications were obtained from the Blast2GO program. All transcriptomes had hits in the three categories, with the exception of *M. complanata* in the Molecular Function category (**Figure 3**), where no genes were categorized in protein binding processes.

Using Level 2 in Biological processes, we found that the single-organism processes, metabolic and cellular processes were higher among the seven categories, overall (**Figure 4**). For clarification, the Blast2GOterm “Level” is a semantic feature not a biological feature. For example, the term "response to stress" is Level 2 because it has two terms above it: "response to stimulus" and "biological process". In Cellular Component, Level 2 (**Figure 5**), all seven categories looked very similar in quantity. Cell and cell part categories represent the majority of proteins with assigned functional annotation. The last GO category, Molecular Function, was observed at Level 3. There, from eight categories, organic cyclic compound binding, small

molecule binding transferase activity and protein binding were the ones with most function annotations. However, protein binding proteins were not found in the transcriptome of *M. complanata* (**Figure 3**).

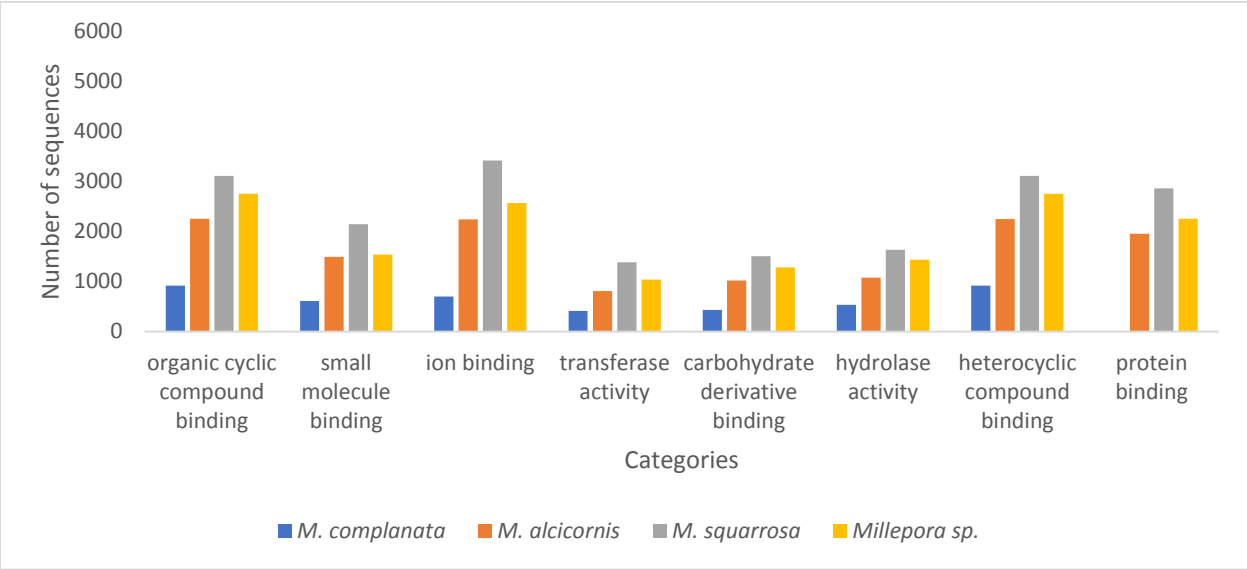


Figure 3: Gene Ontology in Molecular Function (Level 3) of *Millepora* species using Blast2Go.

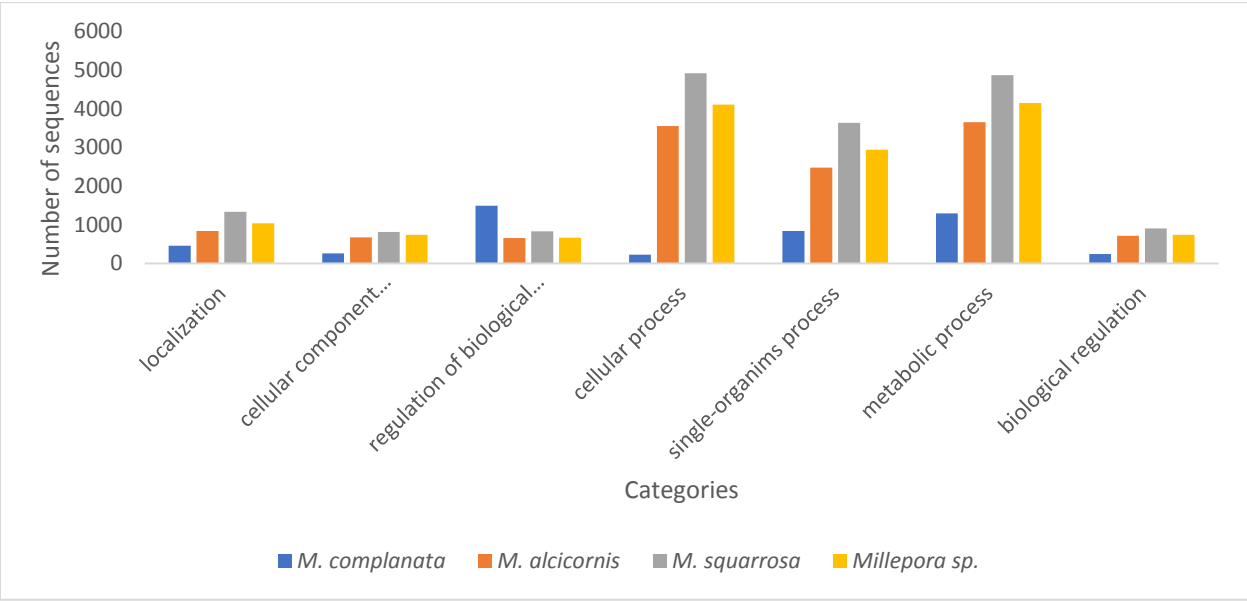


Figure 4: Gene Ontology in Biological Process (Level 2) of *Millepora* species using Blast2Go.

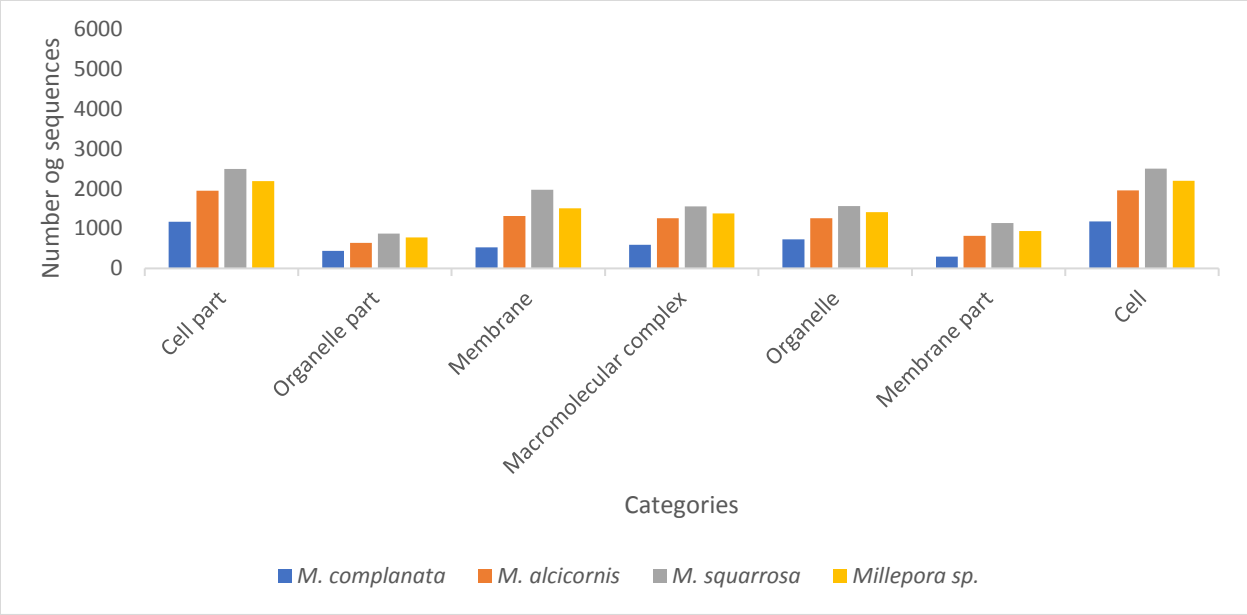


Figure 5: Gene Ontology in Cellular Component (Level 3) of *Millepora* species using Blast2Go.

4.7 Comparison among expressed genes

All four hydrocorals expressed the MCTX1 protein, which is a cytotoxin protein discovered by Iguchi *et al.* (2007). Proteins expressed in gastrozooids and dactylozooids were expected to be found in all *Millepora* species, however, the Transforming Growth Factor – beta (TGF-β) protein was not detected in the *Millepora* sp. colonies. The TGF-β protein is the only one presented in the dactylozooids and not in the gastrozooids. All gastrozooid proteins were found in all four transcriptomes (**Table 2**).

The reproductive biology of *Millepora* spp. is not yet well understood. Sexual reproduction is accomplished during the free living medusoid stage. Researchers still do not know when oogenesis and/or spermatogenesis (reproduction cycle) occurs in each species of *Millepora* and when the medusae are released. *Millepora alcicornis* is the only sample where all sets of genes expressed during the reproductive medusoid stage were detected. Meanwhile, *M. complanata*

expressed all genes for reproduction but not all medusa genes that were described from the scyphozoan jellyfish *Aurelia aurita* (Brekhman *et al*., 2012).

Location	Name	Organims	*M. alcicornis*	*M. squarrosa*	*M. complanata*	*Millepora sp.*
Gastrozooid	pmp 1	*Padycoryna carnea (in Sanders et al. 2014)*	pmp1	pmp1	pmp1	pmp1
	Cnox-2	*H. symbiolongicarpus*	Cnox-3	Cnox-3 Cnox-1	Cnox-3 Cnox-1	Cnox-3
Dactylozooid	Myosine heavy chain	*H. symbiolongicarpus*	myosin heavy chain kinase A	myosin-VI myosin-la myosin-10 myosin-6 myosin x (partial)	myosin X myosin-VI	myosin X myosin-VI
	Cerberus	*H. symbiolongicarpus*	cerberus-like 3 4 (partial)	cerberus cerberus-like 2B	cerberus-like 3 4 (partial) cerberus Family	cerberus Family
	Wnt	*H. symbiolongicarpus*	Wnt5A Wnt4A	Wnt	Wnt	Wnt (partial)
	TGF-β	Metazoans (in Sanders *et al*. 2014)	TGF	TGFBR1	TGFBR1	-
	Cnox-2	*H. symbiolongicarpus*	Cnox-3	Cnox-3 Cnox-1	Cnox-3 Cnox-1	Cnox-3
Medusa	*pmp1*	*Padycoryna carnea (in Sanders et al. 2014)*	pmp1	pmp1	pmp1	pmp1
	collagen 1	*A. aurita*	collagen alpha-1 minicollagen	collagen minicollagen minicollagen 10 minicollagen 12	collagen alpha-5 collagen alpha-1(II) minicollagen-15	Collagen minicollagen collagen alpha-5 collagen alpha-1
	vWFA	*A. aurita*	vWFA	-	-	-
	hemicentin	*A. aurita*	hemicentin-1-like hemicentin-2	-	hemicentin 1-like	-
	myosine light chain	*A. aurita*	myosin light chain kinase A	myosin regulatory light chain	myosin regulatory light chain	Myosin light polypeptide 6 myosin regulatory light chain

	myosine heavy chain	*A. aurita*	myosin heavy chain kinase A	myosin-VI myosin-la myosin-10 myosin-6 myosin x (partial)	myosin X myosin-VI	myosin X myosin-VI
	BHMT	*A. aurita*	betaine-homocysteine S-methyltransferase-1-like	BHMT	BHMT	BMHT
	Wnt	*A. aurita*	Wnt5A Wnt-4	Wnt	Wnt	Wnt (partial)
Reproduction	Capicua	*H. symbiolongicarpus*	capicua homolog	-	capicua	-
	BMP receptor	Metazoans (in Sanders *et al*. 2014)	BMP regulator	BMP 5-8	BMP signaling pathway	BMP signaling pathway
	hedgehog	*H. symbiolongicarpus*	Desert hedgehog A	hedgehog hedgehog 2 (partial)	hedgehog hedgehog 2 (partial)	hedgehog hedgehog 2 (partial)
Nematocysts	MCTX1	*Millepora dichotoma* (in Iguchi *et al*. 2007)	MCTX1	MCTX1	MCTX1	MCTX1

Table 2: Gene ontology analysis with *Hydractinia symbiolongicarpus*, *Aurelia aurita* and MCTX1 of *Millepora dichotoma*.

4.8 Phylogenetic Analysis

4.8.1 Concatenation

4.8.1.1 *Hydra* as the outgroup

The results from the amino acid analysis with the Maximum Likelihood and Bayesian methods are shown in **Figure 6**. The sequence divergence between *Millepora* and *Hydra* is significant, thus the phylogenetic relationships and amount of divergence among the *Millepora* species cannot be clearly observed in the resulting topology. *Millepora complanata* and *M. alcicornis* formed a sister group, which was closely related to *Millepora* sp. *M. squarrosa* was placed as the basal *Millepora* species when using *Hydra* as the outgroup. We used *M. squarrosa* as the outgroup for an additional analysis with the inclusion of only *Millepora* proteins. By excluding *Hydra*, we augmented our data set from 3,888 orthologous protein sequences (i.e.

Hydra + 4 *Millepora* species) to 10,797 orthologous protein sequences (only the 4 *Millepora* taxa).

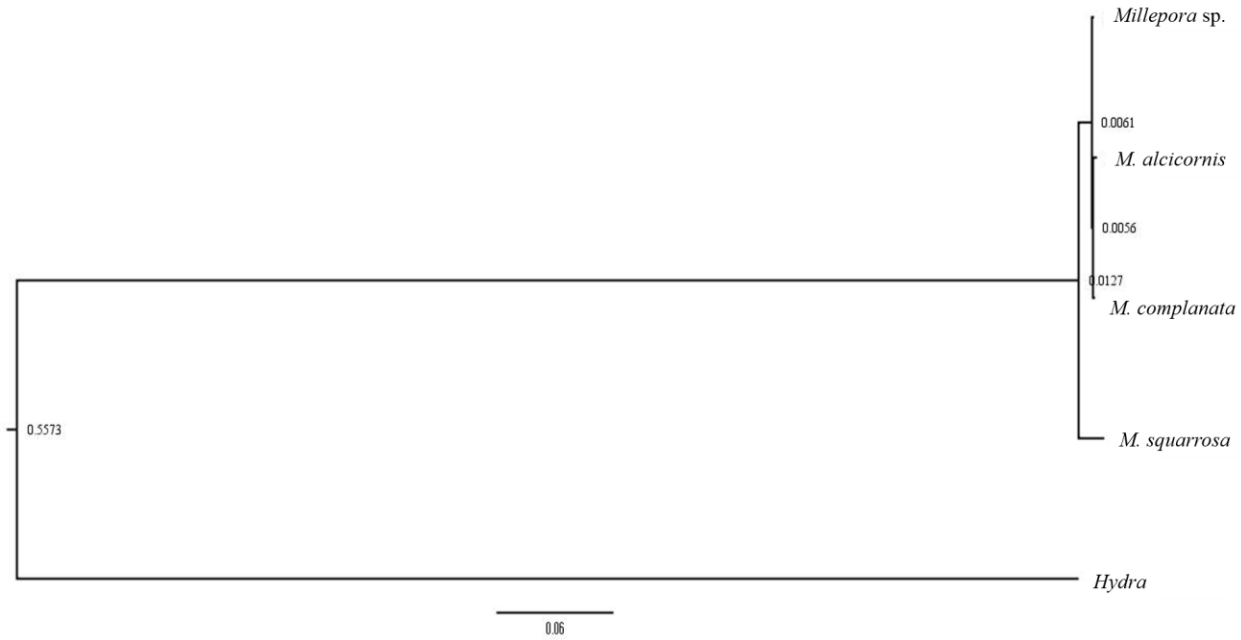


Figure 6: Phylogenetic tree of *Millepora* species based on the Maximum Likelihood method in RAxML. The amino acid substitution model LG+G+F was used as the most appropriate model of evolution. *Hydra* was used as the outgroup. The RAxML program was run with the rapid bootstrapping method search option (-f a) with 1,000 bootstraps.

4.8.1.2 *Millepora squarrosa* as the outgroup

In the concatenation method, Maximum likelihood and Bayesian analyses were performed producing similar topologies to those when *Hydra* was used as the outgroup (**Figures 13 and 14**). Both phylogenetic methods show a close relationship between *M. complanata* and *M. alcicornis*, *Millepora* sp. as different and *M. squarrosa* as the most different species (**Figures 7 and 8**).

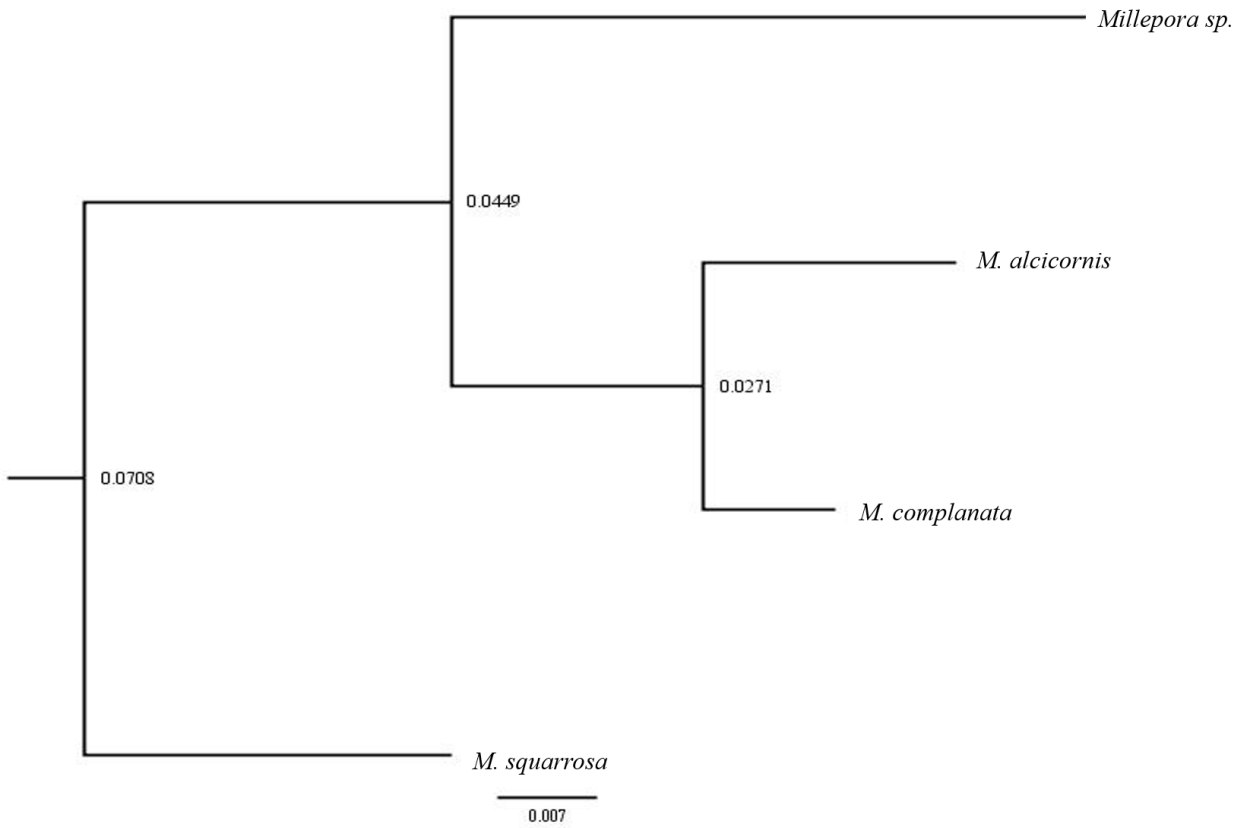


Figure 7: Phylogenetic tree of *Millepora* species based on the Maximum Likelihood method in RAxML. *Millepora squarrosa* was used as the outgroup. The amino acid substitution model JTT+I+F was used as the most appropriate model of evolution. RAxML was run with the rapid bootstrapping method search option (-f a) with 1,000 bootstraps.

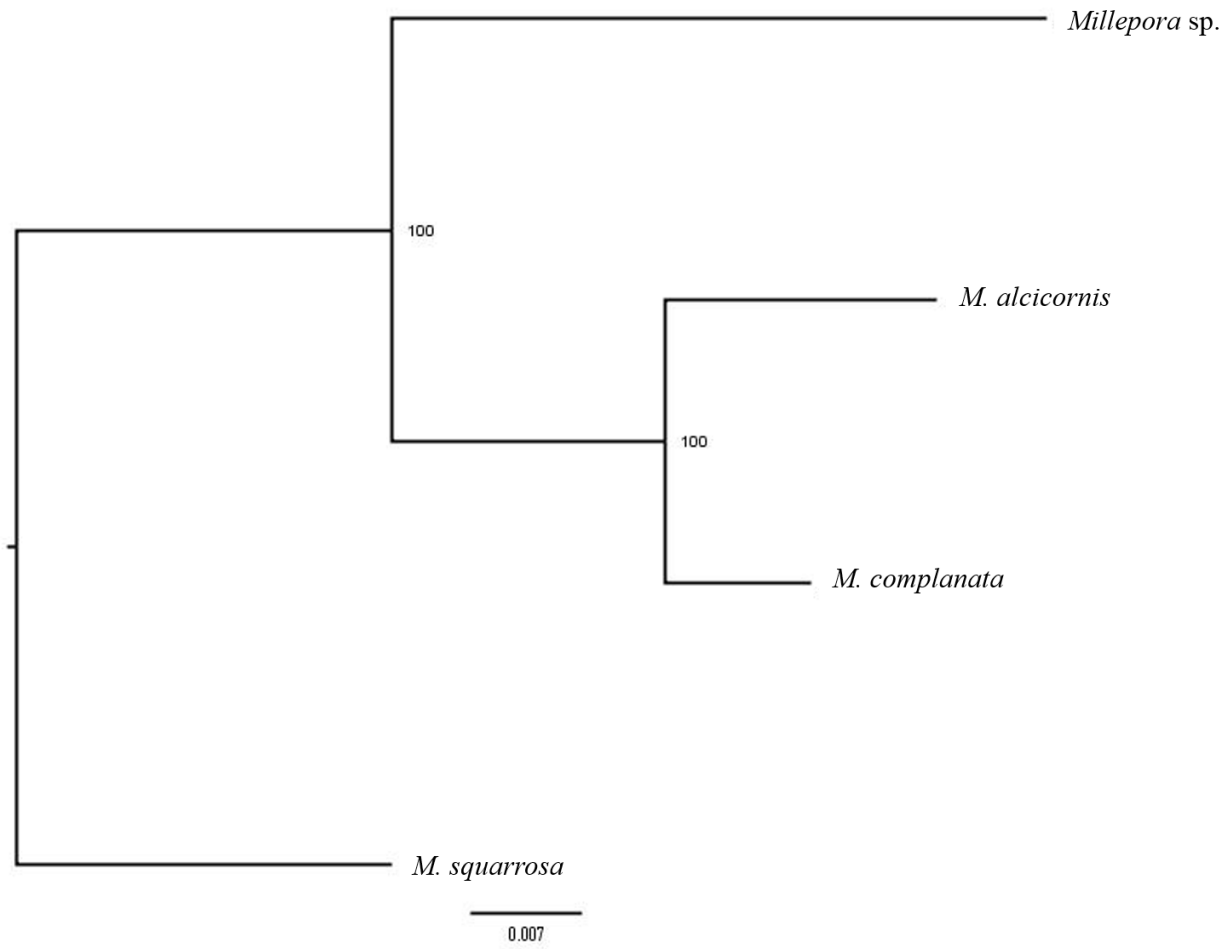


Figure 8: Phylogenetic tree based on Bayesian inference as implemented in MrBayes (Ronquist *et al*., 2012). *Millepora squarrosa* as outgroup. GTR+GAMMA was used as the most appropriate amino acid substitution model for this data set. The analysis was run for 10,000,000 generations with a sample frequency every 10,000; all other parameters were left at default. Nodes are supported with 100% bootstrap support unless otherwise noted.

4.8.2 Coalescence

Because the input for ASTRAL are unrooted trees from the Maximum Likelihood analysis, no Bayesian analysis was performed at this step. The *Hydra* dataset was excluded at this stage because of the high sequence divergence and the high number of gene dropouts. The unrooted trees from the 10,797 proteins data set were used as input for ASTRAL. In this analysis, *Millepora alcicornis* was more closely related to *Millepora* sp. than *M. complanata*, contrary to the topology we recovered with the concatenated approach (**Figure 9**).

The resulting tree from the coalescence analysis has a different topology from the concatenation analysis (**Figure 9 versus Figures 7 and 8**).

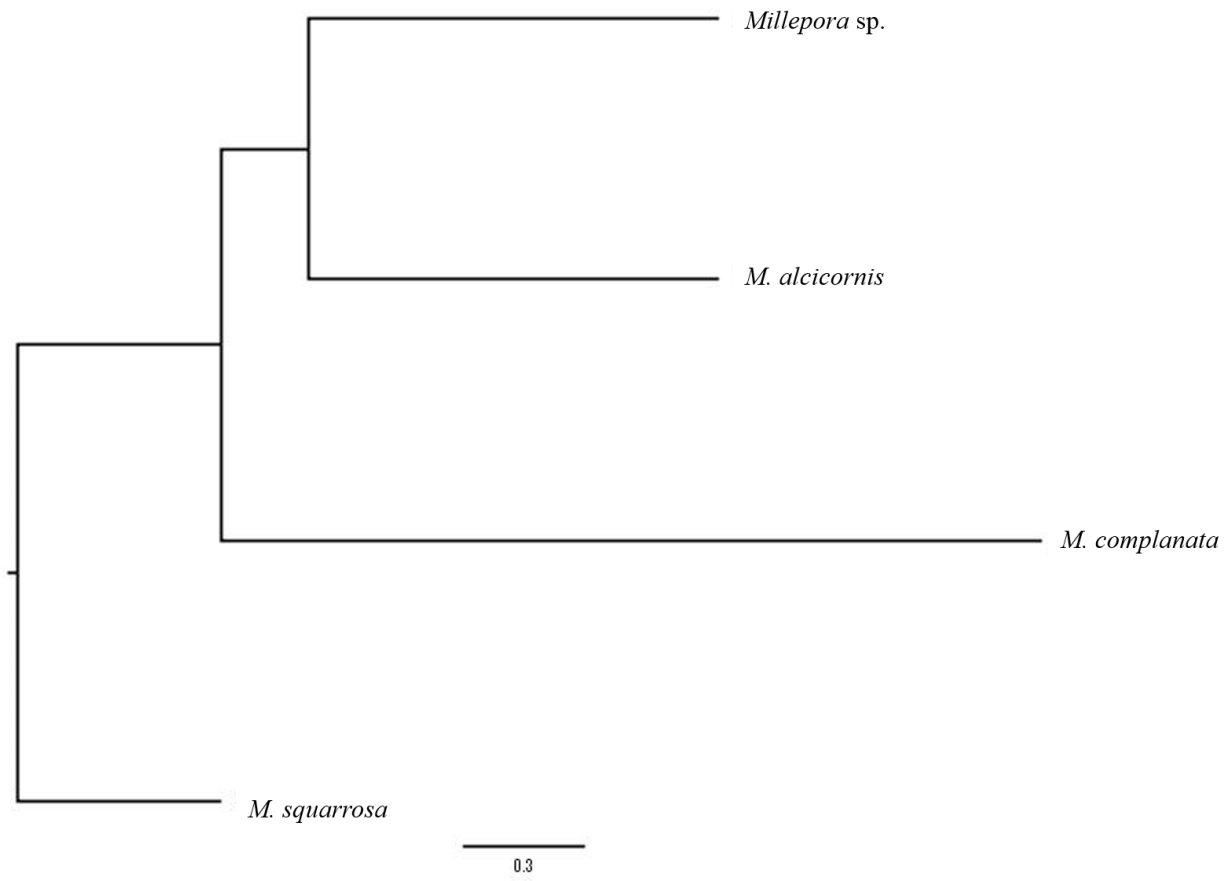


Figure 9: Phylogenetic tree based on the Maximum Likelihood Method as implemented in RAxML. Dataset was constructed with the coalescence method. Ran with for every single gene file with GTR+GAMMA as substitution model and the rapid bootstrapping method (-f a) with 1,000 bootstraps. The trees were used as input for the program Astral-II and was run with the full annotation mode (-t 4) and 100 replicates. Nodes are supported with 100% bootstrap support unless otherwise noted.

5 Discussion

5.1 The taxonomic controversy in *Millepora*

Millepora belongs to the Order Anthoathecata (Cornelius 1992), Suborder Capitata, Family Milleporidae (Fleming 1828). The species problem in *Millepora* was noted as early as 1767 by Linnaeus, who mentioned that the genus contains a heterogeneous group of organisms (Lewis,

2006). After a taxonomic revision was accomplished by Duchassaing and Michelotti (1860) and Edwards (1860), *Millepora* still consisted of an excess of species (Lewis, 2006). It was only after Klunzinger's (1879) taxonomic work, where he established species with their appropriate name and listed all the incidences of synonymy, that the number of *Millepora* species was reduced.

During the 20th century, multiple new species of *Millepora* were described and several intermediate morphotypes recognized (**Figure 1, E and F**). Boschma (1948) recognized seven species from the Indo-Pacific region and three others from the Atlantic Ocean. Vernon (2000) described at least 50 species of milleporids. Tepper *et al*. (2012) recognized the 17-species described by de Weerdt and Glynn (1991) and Cairns *et al*. (1999); 11 species from the Pacific and six from the Atlantic Ocean. More recently, Ruiz-Ramos *et al*. (2014) recognized the Caribbean species described by Boshma (1948), Cairns *et al*. (1999), and Amaral *et al*. (2008). Here, we recognize those species presented in Lewis (2006) and the new *Millepora* species in Amaral *et al*. (2008). It is apparent that there is no consensus on the recognized number of species in the *Millepora* genus, and it is clear that the morphological characters are not sufficient to distinguish the different *Millepora* species, even when combined with the use of few molecular markers (de Weerdt, 1984; Lewis, 1989; Amaral *et al*., 2002; Ruiz-Ramos *et al*., 2014). Furthermore, there are several descriptions of the phenotypic plasticity in *Millepora* (**Figure 1, E and F**), potentially induced by environmental condition or depth (de Weerdt, 1981; Vago *et al*., 1998; Meroz-Fine *et al*., 2003; Lewis, 2006; Tepper *et al*., 2012). Phenotypic plasticity is the most important problem muddling the taxonomic landscape in *Millepora* if morphometrics is the only approach.

Amaral *et al*. (2002) indicated that the skeletal morphological characters are not adequate to distinguish species and Lewis (2006) suggested that studies of genetic population structure might

shed light in the *Millepora* taxonomic problem. Furthermore, Tepper *et al*. (2012) and Ruiz-Ramos *et al*. (2014) did not find sufficient resolution in the genetic markers ITS, 16S and 18S, and COI to differentiate three of the Caribbean species, with the exception *M. squarrosa*, which was easily differentiated with single mitochondrial markers (Ruiz-Ramos *et al*., 2014). Lewis (2006) presented a personal communication with C.S. Tepper (author of Tepper *et al*., 2012) mentioning the possibility of hybridization between *M. complanata* and *M. alcicornis*. Lack of information on the reproductive cycle of milleporids (e.g. spawning time of medusa, larvae) hinders the testing of the hybridization hypothesis.

Next Generation Sequencing has increased the DNA sequencing capabilities exponentially and now transcriptome, proteomes and even full genome sequencing are becoming routine for small labs. Most scientists have adopted the new sequencing technology, leaving behind the single/multiple gene sequence approaches. Transcriptome data is routinely used to resolve the phylogeny of organisms (Kocot *et al*., 2011; Weigert *et al*., 2014; Goodheart *et al*., 2015; Janouškovec *et al*. 2016). Similarly, we are taking advantage of the ubiquitous NGS technologies and we are applying them to resolve the species problem in Caribbean *Millepora*. Here, we present the first transcriptome assembly of four Caribbean species of *Millepora* with thousands of genes for comparison and phylogenetic reconstruction. All species in this study were acclimated for eight weeks to the same environmental conditions in aquaria; therefore, we expect that observed differences between them are not indicative of environmental variability.

5.2 Gene ontology analysis and comparison

Based on the results of gene ontology, most of genes present for the transcriptome of *H. symbiolongicarpus*, the medusa of *A. aurita* and the *Millepora* cytotoxin were found in all *Millepora* transcriptomes, with a few exceptions (**Table 2**). For example, the TGF-β gene was

absent in *Millepora* sp. and present in all other *Millepora*. Sanders *et al*., (2014) and Pang *et al*., (2011), mentioned that Cerberus gene is an antagonist of TGF-β and Wnt. Both genes, Wnt, the Cerberus-like or Cerberus Protein Family, were found in all four transcriptomes. The TGF-β, known as the **T**ransforming **G**rowing **F**actor **B**eta, is a superfamily of genes found only in metazoans and encodes for an intracellular signaling pathway regulating different processes including developmental processes in the cell. In the TGF-β superfamily, we can observe two major classes of genes, the TGF-β-like and BMP-like (Bone Morphogenetic Protein) (Pang *et al*., 2011). The latter was found in all transcriptomes assemblies (**Table 2**). However, when sequences of *Millepora* sp. and *M. complanata* were blasted, the ontology of Bone Morphogenetic Protein (BMP) signaling pathway was found. The BMP pathway interacts with canonical Wnt signaling pathway; both pathways are important in organogenesis.

Another protein that was found in most *Millepora* transcriptomes, except in *M. alcicornis*, was the von Willebrand factor type A domain (vWFA). The vWFA domain has similarities to the hemicentin protein (Brekhman *et al*., 2015). Despite the fact that *M. complanata* and *M. alcicornis* expressed the hemicentin protein, it was not found in the other two transcriptomes (**Table 2**). In the transcriptome of the scyphozoan *Aurelia aurita*, the vWFA domain is expressed in the medusa stage but its function has not been described. The interactions of vWFA and the GpIbα are associated with the movement of the platelets in the vascular system of humans (Huizinga *et al*., 2002).

The protein capicua was not found in all transcriptomes. This protein is expressed during the development of the oocytes in females and in the gastrodermis of the male gonopores in the hydrozoan *Hydractia symbiolongicarpus* (Sanders *et al*. 2014). Capicua (Cic) proteins have structural properties and, in vertebrates, they repress transcription. In *Drosophila*, capicua

recruits Groucho proteins, which are transcriptional co-repressors expressed during the embryo stage but not in other developmental stages (Forés *et al*., 2015). In the hydrozoan *H. symbiolongicarpus* Cic was found in the gonozoid polyps, and it was also found expressed in *Millepora* transcripts, for the exception of *M. squarrosa* and *Millepora* sp. (**Table 2**).

5.3 Phylogenetic reconstruction

The phylogeny reconstruction with *Hydra* as an outgroup is characterized, by the large (**Figure 6**) taxonomic and sequence divergence between *Hydra* and *Millepora* (*Hydra,* Suborder: Aplanulata; *Millepora*, Suborder: Capitata). We identified 3,888 vs. 10,797 orthologous genes when including and excluding *Hydra,* respectively. *Hydra* is a well-studied model freshwater cnidarian with multiple genomes sequenced, whereas no genomic resources existed for *Millepora* before this study. The premise of big data sets stemming from NGS methods was to resolve recalcitrant nodes, such as those of the *Millepora* spp., where traditional approaches have failed to resolve. However, the exponentially larger data sets and their inherent complexity introduce significant phylogenetic noise and conflict in the analyses. Concatenation of sequences is especially useful in cases where noise is presumed to be the dominant source of conflict, and gene-tree methods for reconstruction are useful when conflicted phylogenetic signal is attributed to incomplete lineage sorting (ILS). Even though we have increased the data sets used for phylogenetic analysis in *Millepora* by several orders of magnitude, we still observe conflicted topological signals between the coalescence and the concatenated approaches. The two approaches fail to reach a consensus topology, except for *M. squarrosa*, which is clearly differentiated. The topological differences from the coalescence and concatenation analyses (**Figures 7 and 8 versus Figure 9**), demonstrate that *M. alcicornis*, *M. complanata* and

Millepora sp. are genetically more closely related than expected from their distinct morphologies and support previous results (Tepper *et al*., 2012; Ruiz-Ramos *et al*., 2014). The *Millepora* lineages exhibit ILS as shown by the incongruent results between the concatenation and coalescence analyses (**Figures 7 and 8 versus Figure 9**). But other sources of conflict such as hybridization, which has been hypothesized in *Millepora* (Lewis, 2006), or horizontal gene transfer can cause such topological incongruence. Mirarab and Warnow (2015) indicated that the coalescent method is more accurate when there are large numbers of genes and Incomplete Lineage Sorting (ILS) is not present. They also showed that the concatenation method cannot be discarded, and the tree obtained can be a useful insight in the evolutionary history of the species being studied.

When ILS is observed in a group of species, topological differences can be observed from gene tree to gene tree (Maddison, 1997). ILS is the result of recent divergence between species; in our case, *M. complanata* and *M. alcicornis* and *Millepora* sp. have probably diverged recently, perhaps several hundreds of thousand years ago, and still diverging. The use of fossils may be key in the correct phylogenetic assignment of these species. However, fossils can only be recognized by morphological characters, therefore species assignments of fossil *Millepora* are still hampered by the same limitations as those that exist during species identification of living *Millepora* species. Nonetheless, the existing fossil record may shed some light on the large amount of ILS found in *Millepora* species. The oldest *Millepora* fossil record found was from the Devonian (416 Million Years Ago) (Machel, 1990). Fossils of *M. alcicornis* (Florida - Peteuch, 1986) and *M. complanata* (Costa Rica - Budd *et al*., 1999) were discovered from the Pliocene (5.3 to 2.6 Million Years Ago) and Pliocene to Pleistocene (2.50 Million Years Ago to 11,700 Years Ago), respectively. From the Quaternary Period (from 2,588,000 Years Ago to

present), *M. complanata* and *M. alcicornis* have been reported in various places of the Atlantic: *M complanata* from San Salvador Island in Bermuda (Hatting and Warren, 1989; Greenstein, *et al*., 1998), Panama (Macintyre and Aronson, 2013), and Florida (Lighty, 1977); *M. alcicornis* from Cape Verde Islands (Boekschoten and Best, 1988; Baarli *et al*., 2013), Dominican Republic (Vaughan *et al*., 1921), Haiti (Woodring *et al*., 1924), and Florida (Weinstein, 2007). However, other Quaternary fossil records of *Millepora* spp. were reported from different areas of the Atlantic. The fossil record collectively indicates that *M. alcicornis* and *M. complanata* are relatively recent species with almost two million years of common history.

Today, with the implementation of new techniques and sequencing efforts, we have access to thousands of times more data than previously collected, yet we cannot unequivocally characterize the phylogenetic relationships among *Millepora* species in the Caribbean. We need to focus on identifying genomic areas with signals of strong selection that may be indicative of pre- and postzygotic isolation. We also need to explore the role of hybridization, if it does exist between the *Millepora* species.

Overall, results from this study confirm previous observations by Ruiz-Ramos *et al*. (2014) that *M. squarrosa* is the only easily distinguished species in the Caribbean both morphologically and genetically, but we still cannot differentiate unequivocally between *M. alcicornis* and *M. complanata*. In addition, the *Millepora* sp. represents a new morphotype in the Caribbean. *Millepora* sp. is different from both *M. alcicornis* and *M. complanata* and can be grouped within the "*Millepora* complex", coined by Ruiz-Ramos *et al*. (2014).

6. Conclusions

- *Millepora squarrosa* is the only morphologically and genetically clearly defined species
- Concatenation and coalesence analyses result in different topologies

- The phylogeny of *Millepora* is not resolved most likely because of Incomplete Lineage Sorting
- The *de novo* transcriptomes presented here are the first genomic resources in *Millepora*

7. Recommendation for Future Studies

- Sequencings of the genomes of all recognized *Millepora* species
- Use differential expression analysis to explore species-specific differences in gene expression
- Determination of the reproductive cycle of *Millepora*
- Analyze the microbiome of different *Millepora* spp.
- Conduct experiments to detect hybridization in *Millepora*.

8. References

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21(9), 2104-2105.

Abe, N. 1938. Feeding behavior and the nematocysts of *Fungia* and 15 other species of corals. Palau Tropical Biology Station Studies 1, 469-521.

Amaral, F.D., Broadhurst, M.K., Cairns, S.D., Schlenz, E., 2002. Skeletal morphometry of *Millepora* occurring in Brazil, including a previously undescribed species. Proceedings of the Biological Society of Washington 115, 681-695.

Amaral, F.D., Stainer, A.Q., Broadhurst, M.K., Schlenz, E., 2008. An overview of the shallow-water calcified hydroids from Brazil (Hydrozoa: Cnidaria), including a description of a new species. Zootaxa 68, 56-68.

Baarli, B.G., Santos, A.G., Mayoral, E.J., Ledesma-Vazquez, J., Johnson, M.E., Silva, C.M. Cachao, M., 2013. What Darwin did not see: Pleistocene fossil assemblages on a high energy coast at Ponta das Bicudas, Santiago, Cape Verde Islands. Geological Magazine 150(1), 183-189.

Banaszak, A.T., Iglesias-Prieto, R., Trench, R. K. 1993. *Scrippsiella velellae* sp. nov. (Peridiniales) and *Gloeodinium viscum* sp. nov. (Phytodinales), dinoflagellate symbionts of two hydrozoans (Cnidaria). Journal of Phycology 29, 517-528.

Banaszak, A.T., Ayala-Schiaffino, B.N., Rodríguez-Román, A., Enríquez, S., Iglesias-Prieto, R., 2003. Response of *Millepora alcicornis* (Milleporina: Milleporidae) to two bleaching events at Puerto Morelos reef, Mexican Caribbean. Revista de Biología Tropical 51 (4), 57-66.

Blake, J. A., 1996. Family Spionida Grube, 1850. In "Taxonomic Atlas of the Santa Maria Basin and Western Santa Barbara Channel. Annelida, Part 3. Polychaeta: Orbiniidae to Cossuriadae" (J.A. Blake, B. Hilbig, and P. H. Scott, eds), Vol 6, 81-223. Santa Barbara Museum of Natural History, Santa Barbara, California.

Boekschoten, G.J., Best, M.B. 1988. Fossil and recent shallow water corals from the Atlantic islands off western Africa. Zoologische Mededelingen 62(8):99-112.

Boshma, H., 1948. The species problem in *Millepora*. Zool. Verh 1, 1-115.

Brekhman, V., Malik, A., Haas, B., Sher, N., Lotan, T., 2015. Transcriptome profiling of the dynamic life cycle of the scyphozoan *Aurelia aurita*. BMC Genetics 16 (84).

Brander, K. M., McLeod, A. A.Q. R., Humphreys, W. F. 1971. Comparison of species diversity and ecology of reef-living invertebrates on Aldabra Atoll and at Watamu Kenya. In "Regional Variation in Indian Ocean Coral Reefs" (D. R. Stoddart and Sir Maurice Yonge, eds.) Symposia of the Zoological Society of London 28, 397-431.

Budd, A.F., Johnson, K.G., Stemann, T.A., Tompkins, B.H., 1999. Pliocene to Pleistocene reef coral assemblages in the Limon Group of Costa Rica. Bulletin of American Paleontology 357:119-159.

Cairns, S.D., Hoeksema, B.W., can der Land, J., 1990. Appendix List of the extant stony corals. Atoll Research Bulletin 459: 13-46.

Chiappone, M., Sullivan, K. M., 1994. Ecological structure and dynamics of Nearshore hard-bottom communities in the Florida Keys. Bulletin of Marine Science 54(3): 747-756.

Chomczynski, P., Mackey, K., 1995. Short technical report. Modification of the TRIZOL reagent procedure for isolation of RNA from Polysaccharide-and proteoglycan-rick sources. Biotechniques 19(6), 942-945.

Ciardelli, A., 1967. The anatomy of the feeding mechanism and the food habits of *Microspathodon chrysurus* (Pisces: Pomacentridae). Bulletin of Marine Science 17, 845-883.

Clemente, A., Rodríguez, A., Brito, A., Ramos, A., Monterroso, Ó., Hernández, J.C., 2011. On the occurrence of the hydrocoral *Millepora* (Hydrozoa: Milleporidae) in the subtropical eastern Atlantic (Canary Islands): is the colonization related to climatic events? Coral Reefs 3(1), 237-240.

Conesa, A., Götz, S., García-Gómez, J.M. Terol, J., Talón, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21(18), 3674-3676.

Cornelius, P.F.S., 1992. Medusa loss in leptolid Hydrozoan (Cnidaria) hydroid rafting and abbreviate life-cycles among their remote-island faunae: an interim review. In: J. Bouillon, F. Boero, F. Cicogna, J.M. Gili & R.G. Hughes, eds., Aspects of hydrozoan biology. Scientia Marina 56, 2-3: 245-261.

Croquer A.,Weil E. 2009. Local and geographic variability in distribution and prevalence of coral and octocoral diseases in the Caribbean II: Genera-level analysis. Distribution of Aquatic Organisms 83: 209-223.

Duchassaing, P., Michelotti, J., 1860. Mémorie sur les coralliaires des Antilles. Mémories de l'Académie des Sciences de Turin 2(19), 1-89.

de Kruijf, H. A. M. 1975. General morphology and behavior of gastrozoids and dactylozoids in two *Millepora* (Milleporina, Coelenterata). Marine and Freshwater Behaviour and Physiology 3,181-192.

de Weerdt, W. H., 1981. Transplantation experiments with Caribbean *Millepora* species (Hydrozoa, Coelenterata), including some ecological observations on growth form. Bijdragen tot de Dierjunde 51(1), 1-19.

de Weerdt, W. H., 1984. Taxonomic characters in Caribbean *Millepora* species (Hydrozoa, Coelenterata). Bijdragen tot de Dierkunde 54, 243-262.

de Weerdt, W. H., 1990. Discontinuous Distribution of the Tropical Atlantic Hydrocoral *Millepora squarrosa*. BEAUFORTIA 41, 27, 195-203.

de Weerdt, A., Glynn, P., 1991. A new and presumably now extinct species of *Millepora* (hydrozoa) in the eastern Pacific. Zoological Mededelingen 65(15), 267-276.

Davey, J.W., Hohenlohe, P. A., Etter, P.D., Boone, J.Q., Catchen, J.M., Blaxter, M.L., 2011. Genome-wide genetic maker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics 12, 499-510.

Edmunds, P., 1999. The role of colony morphology and substratum inclination in the success of *Millepora alcicornis* on shallow coral reefs. Coral Reefs 18(2), 133-140.

Edwards, M. H.,1860. Mistoire Naturelle des Coralliaires ou Polypes Proprement Dits. Tome Troisième. Lobrairie Encyclopédique de Rorot, Paris.

Fenner, D., 1999. New observations in the stony coral (Scleractinia, Milleporidae, and Stylasteridae) species of Belize (Central America) and Cozumel (Mexico). Bulletin of Marine Science 64(1),143-154.

Fitt, W. K., 2012. Bleaching of the fire corals *Millepora*. Proceedings of the 12th International Coral Reef Symposium. Cairns, Australia, 9-13 July 2012.

Fleming, J. (1828). A History of British Animals, Exhibiting the Descriptive Characters and Systematical Arrangement of the Genera and Species of Quadrupeds, Birds, Reptiles, Fishes, Mollusca, and Radiata of the United Kingdom; Including the Indigenous, Extirpated, and Extinct Kinds, Together with Periodical and Occasional Visitants. Bell & Bradfute, Edinburgh. Pp. 1–565.

Forés, M., Ajuria, L., Samper, N., Astigarraga, S., Nieva, C., Grossman, R., González-Crespo, S., Paroush, Z., Jiménez, G., 2015. Origins of Context-Dependent Gene Repression by Capicua. PlosOne, 11(1): e1004902. http://doi.org/10.1371/jornal.pgen.1004902

Glynn, P. W., 1973. Aspects of the ecology of coral reefs in the western Atlantic region. In "Biology and Geology of Coral Reefs" (O. A. Jones and R. Endean, eds), Vol. II, Biology 1, 271-324. Academic Press, New York.

Goodheart, J.A., Bazinet, A.L., Collins, A.G., Cimmings, M.P., 2015. Relationships within Cladobranchia (Gastropoda: Nudibranchia) based on RNA-Seq data: an initial investigation. Royal Society Open Science 2, 150196.

Grabherr, M.G., Hass, B. J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nasbaum, C., Tindblad-Toh, K., Friedman, N., Regev, A., 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nature Biotechnology 29(7), 644-652.

Greenstein, B.J., Harris, L.A., Curran, H.A., 1998. Comparison of recent coral life and death assemblages to Pleistocene reef communities: implications for rapid faunal replacement on recent reefs. Carbonates and Evaporites 13(1), 23-31.

Hatting, D.E., Warren, V.L., 1989. Stratigraphic analysis of a fossil Neogoniolithon-capped patch reef and associated facies, San Salvador, Bahamas. Coral Reefs 8, 19-30.

Hickson, S. J., 1891. The medusae of *Millepora murryi* and the gonophores of *Allopora* and *Distichophora*. Proceedings of the Quarterly Journal of Microscopical Science 32, 375-407.

Hoeksema, B.W., Nunes, F.L.D., Linder, A., De Souza, J.N., 2014. *Millepora alcicornis* (Hydrozoa: Capitata) at Ascension Island: confirmed identity based on morphological and molecular analyses. Journal of the Marine Biological Association of the United Kingdom 97(4), 709-712.

Huizinga, E.G., Tsiju, S., Romijn, R.A.P., Schipgorst, M.E., de Groot, P.G., Dixma, J.J., Gros, P., 2002. Structures of lyncoprotein Ibα and its Complex with von Willebrand Factor A1 Domain. Science 297, 1176-1179.

Hunte, W., Conlin, B, E., Marsden, J., R. 1990. Habitat selection in the tropical polychaete *Spirobranchus giganteus*. I. Distribution on corals. Marine Biology 104, 87-92.

Iguchi, A., Iwanaga, S., Nagai, H., 2007. Isolation of a novel protein toxin from fire coral. Biochemical and Biophysical Research Communications 365(1), 107-112.

Jaap, W.C., 1979. Observations on zooxanthellae expulsion at Middle Sambo Reef, Florida Keys. Bulletin of Marine Science 29, 414-422.

Janouškovec, J., Gavelis, G.S., Burki, F., Dinh, D., Bachvaroff, T.R., Gornik, S. G., Bright, K.J., Imanian, B., Storm, S.L., Delwiche, C.F., Waller, R.F., Fensome, R.A., Leander, B.S., Rohwer, F.L., Saldarriaga, J.F., 2016. Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. PNAS, E171-E180.

Jones, D.T., Taylor, W.R. and Thornton, J.M. 1992. The rapid generation of mutation data matrices from protein sequences. Computer Applications in the Biosciences 8, 275-282.

Katoh, S., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 30, 772-780.

Klunzinger, C.B., 1879. Die Jorallthiere des Rothen Meeres. Vol 3. Gutman's chen Buchhandlung, Berlin.

Kocot, K. M., Cannon, J. T., Todt, C., Citarella, M. R., Kohn, A. B., Meyer, A., Santos, S. R., Schander, C., Moroz, L. L., Lieb, B., Halanych, K. M., 2011. Phylotranscriptomics reveals deep molluscan relationships. Nature 477, 252-257.

Le, S.Q., Gascuel. O., 2008. An improved general amino acid replacement matrix. Molecular Biology and Evolution 25, 1307-1320.

Lewis, J. B., 1989. The ecology of *Millepora*. Coral Reefs 8, 99-107.

Lewis, J. B., 1991. Testing the coral fragment size-dependent hypothesis for the calcareous hydrozoan *Millepora complanata*. Marine Ecology Progress Series 70, 101-104.

Lewis, J.B. 1992A. Heterotrophy in corals: Zooplankton predation by the hydrocoral *Millepora complanata*. Marine Ecology Progress Series 90, 251-256.

Lewis, J. B., 1992B. Recruitment, growth and mortality of a coral-inhabiting barnacle *Megabalanus stultus* (Darwin) upon the hydrocoral *Millepora complanata* Lamarck. Journal of Experimental Marine Biology and Ecology 162, 51-64.

Lewis J B., 1996. Spatial distribution of the calcareous hydrozoan *Millepora complanata* and *Millepora squarrosa* on coral reefs. Bulletin of Marine Science 59, 853-856.

Lewis, J.B., 1998. Reproduction, larval development and functional relationships of the burrowing spionid polychaete *Dipolydora armata* with the calcareous hydrozoan *Millepora complanata*. Marine Biology 130, 651-662.

Lewis, J.B., 2006. Biology and ecology of the hydrocoral *Millepora* on coral reefs. Advances in Marine Biology 50(05), 1-55.

Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Genomics 12(323), 1-16.

Li, B., Ruotti, V., Stewart, R.M., Thompson, J.A., Dewey, C.N., 2010. RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics 26(4), 493-500.

Lighty, R.G., 1977. Relict shelf-edge Holocene coral reef: southeast coast of Florida. Proceedings of the 3rd International Coral Reef Symposium 2, 215-221.

Macintyre, I.G., Aronson, R.B., 2013. Diving in the Past: Scuba and the Temporal Dimension of Coral Reefs. Smithsonian Contributions to the Marine Sciences 39.

Machel, H.G., 1990. Faziesinterpretation des Briloner Riffs mit Hilfe eines Faziesmoddells für devonische Riffkarbonate. Geologisches Jahrbuch D95, 43-83.

Madison, W., 1997. Gene trees in species trees. Systematic Biology 46, 523-536.

Marsden, J. R. 1992. Reproductive isolation in two forms of the serpulid polychaete, *Spirobranchus polycerus* (Schamarda) in Barbados. Bulletin of Marine Science 51, 14-18.

Martins, M., 2011. Cutadapt remove adapter sequences from high-throughput sequencing reads. EMBnet.jornals 17(1), 10-12.

Martins Garcia, T., Matthews-Cascon, H., Frankling-Junior, W., 2008. Macrofauna associated with branching fire coral *Millepora alcicornis* (Cnidaria: Hydrozoa). Thalassas 24(1), 11-19.

Martins Garcia, T., Matthews-Cascon, H., Frankling-Junior, W., 2009. *Millepora alcicornis* (Cnidaria: Hydrozoa) as substrate for benthic fauna. Brazilian Journal of Oceanography 57(2), 153-155.

Mattraw, H.C.Jr., 1969. Variations of *Millepora alcicornis* on Bermuda platform. Bermuda Biology Statistics Special Publishing 6, 29-38.

Meroz-Fine, E., Vrickner, I., Loya, Y., Ilan, M., 2003. The hydrozoan coral *Millepora dichotoma*: Speciation or phenotypic plasticity? Marine Biology 143, 175-1183.

Miller, M. A., Pfeiffer, W., Schwartz, T., 2010. "Creating the CIPRES Science Getway for inference of large phylogenetics trees" in Proceedings of Gerway Computing Enviroments Workshop (GCE), 14 Nov. 2010, New Orleans, LA. Pp 1-8.

Mirarab, A., Warnow., 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31, 44-52.

Morri, C., Bianchi, C.N., Cattaneo-Vietti, R., Sartoni, G., 2000. Shallow Epibenthic Communities of Ilha Do Sal (Cape Verde Archipelago, Eastern Atlantic). III Simpósio Fauna e Flora Das Ilhas Atlânticas, Supplement. Pp. 157-165.

Nagalakshmi, U., Waren, K., Snyder, M., 2010. UNIT 4.11 RNA-Seq: A Method for Comprehensive Transcriptome Analysis. Current Protocols in Molecular Biology 89 (11), 1-4.

Oakley, T. H., Wolfe, J.M., Lindgren, A.R., Zaharoff, A.K., 2013. Phylotranscriptomics to Bring the Understudied in to the Fold: Monophyletic Ostracoda, Fossil Placement and Pancrustacean Phylogeny. Molecular Biology Evolution 30(1), 215-233.

Ortiz-González, I.C., Rivera-Vicéns, R.E., Schizas, N.V., 2017. *De novo* transcriptome assembly of the hydrocoral *Millepora alcicornis* (branching fire coral) from the Caribbean. Marine Genomics 32, 27-30.

Pang, K., Ryan, J.F., Baxevanis, A.D., Martindale, M.Q., 2011. Evolution of the TGF-β Signaling Pathway and Its Potential Role in the Ctenophore, *Mnemiopsis leidyi*. PlosOne 6(9): e24152. https://doi.org/10.1371/journal.pone.0024152

Pasternak, Z., Rix, A., Abelson, A. 2001. Episymbionts as possible anti-fouling agents on reef-building hydrozoans. Coral Reefs 20, 318-319.

Perez, C. D., Gomez, P. B., 2012. First records of the fireworm *Hermodice carunculata* (Annelida, Polychaeta) preying on colonies of the fire coral *Millepora alcicornis* (Cnidaria, Hydrozoa). Biota Neotropical 12(2), 217-219.

Petuch, E.J., 1986. The Pliocene Reefs of Miami: Their Geomorphological Significance in the Evolution of the Atlantic Coastal Ridge, Southeastern Florida, U.S.A. Journal of Coastal Research 2(4), 391-408.

Randall, J. E.,1967. Food habits of reef fishes of the West Indies. Studies in Tropical Oceanography 5, 665-847.

Razak, T.B., Hoeksema, B.W., 2003. The hydrocoral genus *Millepora* (Hydrozoa: Capitata: Milleporidae) in Indonesia. Zool. Berh. Leiden 345, 313-346.

Ronquis, F., Teslenko, M., van der Mark, P., Ayres, D, L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61(3), 539-542.

Ross, A. 1999. On the occurrence of *Megabalanus stultus* (Darwing), 1854 (Cirripedia: Balanomorpha) in Taiwan: A reappraisal. Zoological Studies 38, 275-278.

Ross, A., Newman, W. A., 1973. Revision of the coral-inhabiting barnacles (Cirripedia: Balanidae). Transactions of the San Diego Society of Natural History 17, 137-174.

Ruiz-Ramos, D.V., Weil, E., Schizas, N.V., 2014. Morphological and genetic evaluation of the hydrocoral *Millepora* species complex in the Caribbean. Zoological Studies 53, 4; https://doi.org/10.1186/1810-522X-53-4

Sanders, S.M., Shcheglovitova, M., Carwright, P., 2014. Differential gene expression between functionally specialized polyps of the colonial hydrozoan *Hydractia symbiolongicarpus* (Phylum Cnidaria). BMC Genomics 15(1), 1.

Schonwald, H., Dubisnky, Z., Achituv, Y., 1997. Diel carbon budget of zooxanthellate hydrocoral *Millepora dichotoma*. In "Proceedings of the Eighth International Coral Reef Symposium, Panamá" (H. A. Lessios and I. G. Macintyre, eds) Vol. 1., 939-946. Smithsonian Tropical Research Institute, Panamá.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19), 3010-3012.

Soong, K. Chaglai, M. 1992. Rediscovery of *Megabalanus stultus* (Darwin, 1854) (Cirripedia, Rhoracica, Balanidae), a fire coral symbiont in the Pacific. Crustaceana 63, 23-28.

Soong, K., Cho, L.C., 1998. Synchronized release of medusa from three species of hydrozoan fire corals. Coral Reefs 17, 145-154.

Stamatakis, A., 2014. RAxML Version 8: A tool for Phylogeny Analysis and Post-Analysis of Large Phylogenies. Bioinformatics, 30(9), 1312-1313.

Strömgern, T., 1976. Skeleton growth of the hydrocoral *Millepora complanata* Lamarck in relation to light. Limnology and Oceanography 21, 100-104.

Tepper, C., Squiers, L., Hay, C., Gorbach, D., Friend, D., Black, B., Greenstein, B., Strychar, K., 2012. Cryptic species: a mismatch between genetics and morphology in *Millepora*. Marine Sciences 2(5), 57-65.

Vago, R., Achituv, Y., Vaky, L., Dubinsky, Z. Kizner, Z., 1998. Colony architecture of *Millepora dichotoma* Foskal. Journal of Experimental Marine Biology and Ecology 224, 225-235.

Vaughan, T.W., Cooke, W., Condit, D.D., Ross, C.P., Woodring, W.P., Calkins, F.C., 1921. A geological reconnaissance of the Dominican Republic. Geological Survey of the Dominican Republic Memoir 1, 1-268.

Veron, J.E.N., 2000. In: Staffort-Smith, M. eds., Corals of the World. Vols 1-3. Australian Institute of Marine Science, Townsville, Australia.

Weigert, A., Helm, C., Meyer, M., Niclel, B., Arendt, D., Hausdorf, B., Santos S. R., Halanuch, K. M., Purschke, G., Bleidorn, C., Struck, T. H., 2014. Illuminating the Base of the Annelid Tree Using Transcriptomics. Molecular Biology Evolution 36(6), 1391-1401.

Weil, E., Croquer, A., Urreiztieta, I. 2009. Temporal variability and consequences of coral diseases and bleaching in La Parguera, Puerto Rico from 2003-2007. Caribbean Journal of Science 45 (2-3): 221-246

Weil, E., Rogers, C.C., 2011. Coral Reef disease in the Atlantic-Caribbean. I Z. Dubinski and N. Stambler Eds. Coral Reefs: An Ecosystem in Transition Chapter 27. Pp 465-492. Springer-Verlag.

Weinstein, D., 2007. Taphonomy of the Late Pleistocene Key Largo Limestone: A Comparison of Modern and Ancient Coral Reef Ecosystems. PhD Dissertation, University of Miami. Pp. 56

Whale, C.M. 1980. Detection, pursuit, and overgrowth of tropical gorgonians by milleporids hydrocorals: Perseus and Medusa revisited. Science 209, 689-691.

Williams E. H., Buncley-Williams, L. 1990. The world-wide coral reef bleaching cycle and related sources of coral mortality. Atoll Research Bulleting 335.

Woodring, W.P., Brown, J.S., Burbank, W.S., 1924. Geology of the Republic of Haiti. Pp. 1-163.

Zapata, F., Goetz, F.E., Smith, S.A., Howison, M., Siebert, S., Church, S.H., Sanders, S.M., Lewis Ames, C., McFadden, C.S., France, S.C., Daly, M., Collins, A. G., Haddock, S.H.D., Dunn, C.W., Cartwright, P., 2015. Phylogenomic Analyses Support Traditional Relationships within Cnidaria. PlosOne 10(10): e0139068. https://doi.org/10.1371/journal.pone.0139068

11. Appendix

9.1 MixS

Appendix, Table 3: MixS information of *Millepora alcicornis*

Item	Definition
Classification	Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Cnidaria; Hydrozoa; Hydroidolina; Anthoathecata; Capitata; Milleporidae; *Millepora*; *alcicornis*
Investigation type	Eukaryote transcriptome
Project name	*Millepora alcicornis* transcriptome
Environment	
Geographic location	Caribbean Sea:

	Puerto Rico, Guánica: Coral Key
	Puerto Rico, Guánica: La Jungla Reef
	Puerto Rico, Lajas: Enrique Key
	Puerto Rico, Cabo Rojo: Playita Azul Reef
Latitude Longitude	Coral Key: 17° 56.280'N, 66° 53.406'W
	La Jungla Reef: 17° 56.650'N, 66° 58.033'W
	Enrique Key: 17° 57.288'N, 67° 3.155'W
	Playita Azul Reef: 18° 8.072'N, 67° 11.245'W
Collection date	2015-10
Environment properties	Shore reef
Depth	<3 meters
Collector	Ingrid C. Ortiz González and Orlando Espinosa Ortiz
Sequencing method	Illumina 4000; Paired-end (2x150): Poly(A) caption
Estimated size Assembly	500Mbp
Method	De novo assembly
Program	Trinity 2.1.1
Finishing Strategy	High quality transcriptome assembly

Appendix, Table 4: MixS information of *Millepora complanata*

Item	
Classification	Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Cnidaria; Hydrozoa; Hydroidolina; Anthoathecata; Capitata; Milleporidae; *Millepora complanata*
Investigation type	Eukaryote transcriptome
Project name	*Millepora complanata* transcriptome
Environment	
Geographic location	Caribbean Sea
	Puerto Rico, Guánica: Coral Key
	Puerto Rico, Guánica: La Jungla Reef
	Puerto Rico, Lajas: Enrique Key
	Puerto Rico, Cabo Rojo: Playita Azul Reef
Latitude Longitude	Coral Key: 17° 56.280'N, 66° 53.406'W
	La Jungla Reef: 17° 56.650'N, 66° 58.033'W
	Enrique Key: 17° 57.288'N, 67° 3.155'W
	Playita Azul Reef: 18° 8.072'N, 67° 11.245'W
Collection date	2015-10

Environment properties	Shore reef
Depth	<3 meters
Collector	Ingrid C. Ortiz González and Orlando Espinosa Ortiz
Sequencing method	Illumina 4000; Paired-end (2x150): Poly(A) caption
Estimated size Assembly	500Mbp
Method	De novo assembly
Program	Trinity 2.1.1
Finishing Strategy	High quality transcriptome assembly

Appendix, Table 5: MixS information of *Millepora squarrosa*

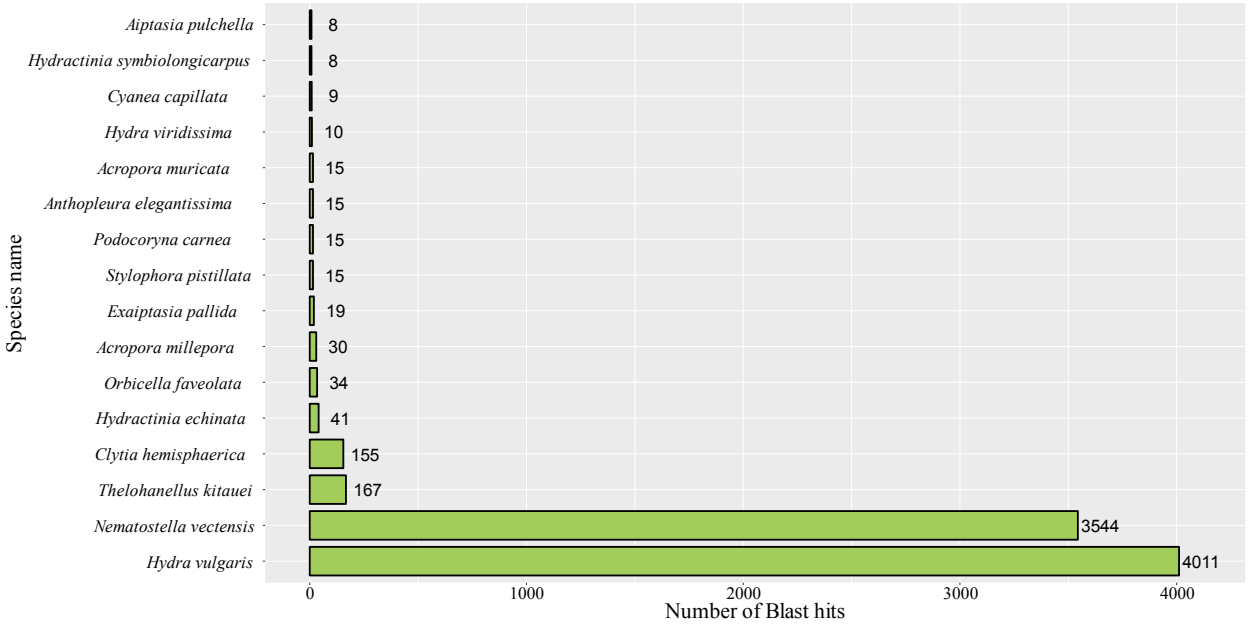
Item	
Classification	Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Cnidaria; Hydrozoa; Hydroidolina; Anthoathecata; Capitata; Milleporidae; *Millepora squarrosa*
Investigation type	Eukaryote transcriptome
Project name	*Millepora squarrosa* transcriptome
Environment	
Geographic location	Caribbean Sea
	Puerto Rico, Guánica: Turrumote II Key
Latitude Longitude	Turrumote II Key: 17° 55.787'N, 66° 58.424'W
Collection date	Nov-4-2015
Environment properties	Barrier reef
Depth	4-5 metes
Collector	Orlando Espinosa and Jaaziel García
Sequencing method	Illumina 4000; Paired-end (2x150): Poly(A) caption
Estimated size Assembly	500Mbp
Method	De novo assembly
Program	Trinity 2.1.1
Finishing Strategy	High quality transcriptome assembly

Appendix, Table 6: MixS information of *Millepora* sp.

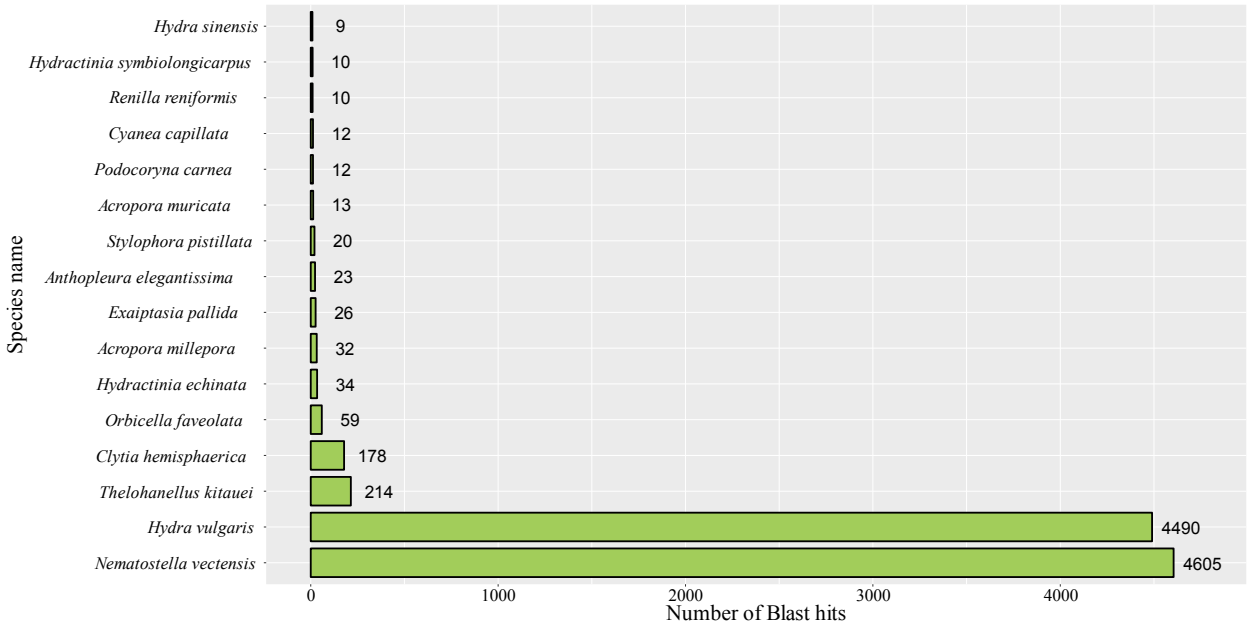
Item	

Classification	Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Cnidaria; Hydrozoa; Hydroidolina; Anthoathecata; Capitata; Milleporidae; *Millepora* sp.
Investigation type	Eukaryote transcriptome
Project name	*Millepora* sp. Transcriptome
Environment	
Geographic location	Caribbean Sea
	Puerto Rico, Guánica: Mata Seca Key
Latitude Longitude	Mata Seca Key: 17° 57.636'N, 67° 0.534'W
Collection date	Oct-22-2015
Environment properties	Shore reef
Depth	Less than a meter
Collector	Ingrid C. Ortiz González
Sequencing method	Illumina 4000; Paired-end (2x150): Poly(A) caption
Estimated size	500Mbp
Assembly Method	De novo assembly
Program	Trinity 2.1.1
Finishing Strategy	High quality transcriptome assembly

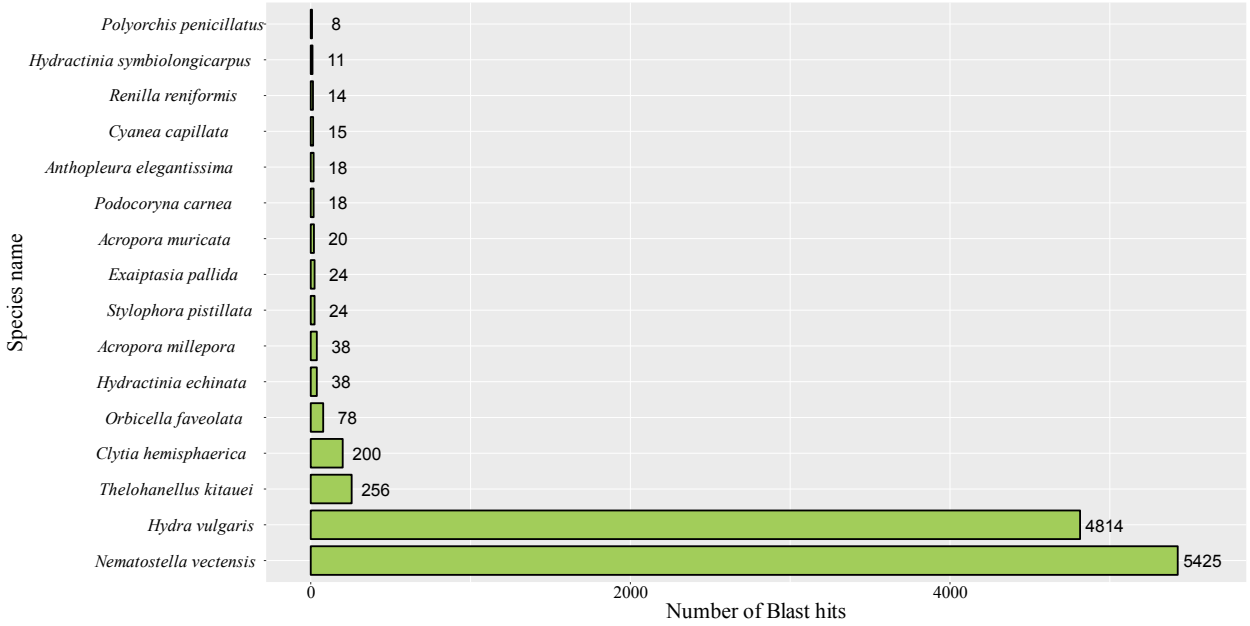
9.2 Blast Hits in Blas2Go



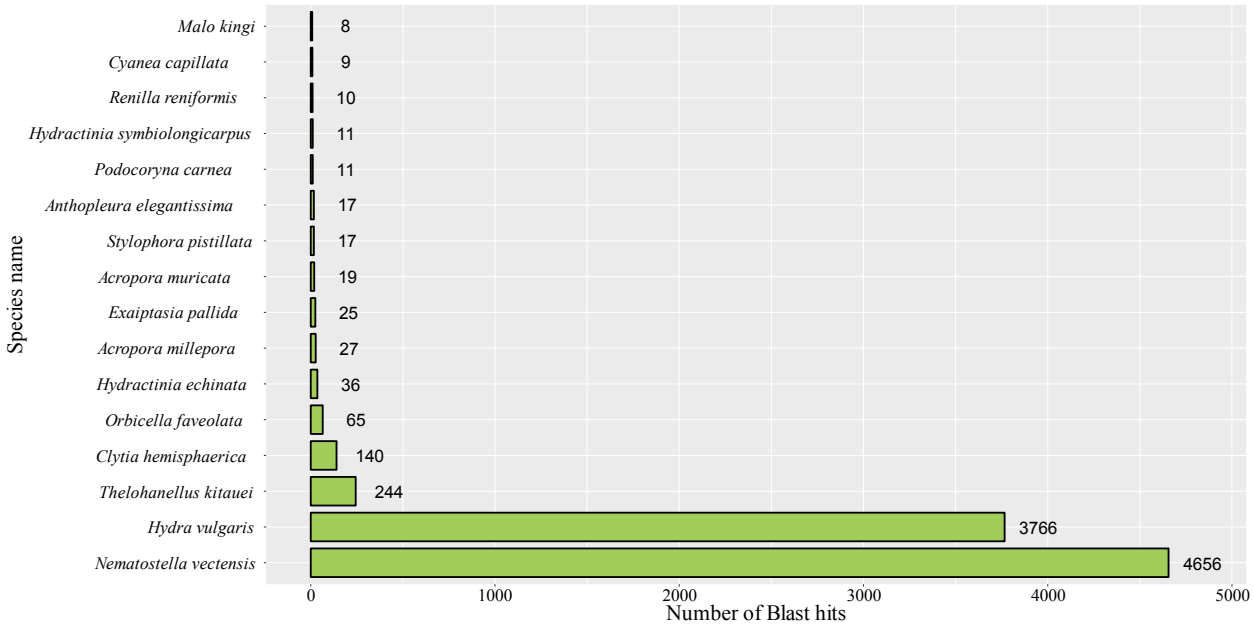
Appendix, Figure 10: Blast hits of *Millepora alcicornis* transcriptome in Blast2Go.



Appendix, Figure 11: Blast hits of *Millepora complanata* transcriptome in Blast2Go.



Appendix, Figure 12: Blast hits of *Millepora squarrosa* transcriptome in Blast2Go.



Appendix, Figure 13: Blast hits of *Millepora* sp. transcriptome in Blast2Go.

9.3 Script

```
#!/usr/bin/bash

###Script to get the reciprocal blast match of millepora sequences

#Loop to rename sequences
#Programs needed in folder:
#       rename_sequences.py

for x in `ls *.original`;do

        a=$( cat $x | grep -c "^>" )

        b=$( echo $x | cut -f 1 -d "." )

        for y in `seq 1 $a`;do

                echo ">"$b"_orf_"$y >>$b.rename

        done

        ./rename_sequences.py $x $b.rename $b.rename.sequences

done

for x in `ls *.rename.sequences`;do
```

```
        cat $x >>millepora_proteins_filter.fasta

done

#Blast database creation and run

makeblastdb -in millepora_proteins_filter.fasta       -input_type fasta -dbtype prot

mv millepora_proteins_filter.fasta millepora_proteins_filter_input.fasta

blastall -p blastp -i millepora_proteins_filter_input.fasta -d
millepora_proteins_filter.fasta -e .00001 -o blast_v_blast_millepora_proteins_filter.txt
-a 15 -m 8


###LOOP TO GET RECIPROCAL BLAST USING MULTIPLE SORTING K3 K11
#Programs needed in the folder:
#       extract_final_sequences.py

#Initial loop using malc as reference.
#Possible isoforms will not pass this step

for x in `cat malc.rename | cut -c 2-`;do

        #Multiple sorting of blast result. Using all matches with malc in 1st column and
not present in 2nd column.
        cat blast_v_blast_millepora_proteins_filter.txt | awk '{if ($1~/^malc/) print $0}' |
awk '{if ($2!~/^malc/) print $0}' | \
        grep -w "$x" | sort -k 3 -g | sort -k 11 -g >sort_k3k11_malc

        #Variables to know when a sequence is already taken
        #a for mcom
        a=$( echo "0")
        #b for msp
        b=$( echo "0")
        #c for msqu
        c=$( echo "0")

        #Checking if they are reciprocal matches and writing name to a file

        cat sort_k3k11_malc | while read line;do

                echo $x >malc.name.match

                #Reciprocal check with mcom
```

```
		if [ `echo $line | awk '{print $2}' | cut -f 1 -d "_"` == "mcom" ] && [ $a -eq 0 ];then

			echo $line | awk '{print $2}' >mcom.name.match

			a=$( echo "1")

			name_malc=$( cat malc.name.match )
			name_mcom=$( cat mcom.name.match )

			cat blast_v_blast_millepora_proteins_filter.txt | awk '{if ($1~/^mcom/) print $0}' | awk '{if ($2!~/^mcom/) print $0}' | \
			grep -w "$name_mcom" | sort -k 3 -g | sort -k 11 -g >sort_k3k11_mcom

			if [[ `cat sort_k3k11_mcom | grep -w "$name_malc" | awk '{print $1}' | head -n 1` == $name_mcom ]] && [[ `cat sort_k3k11_mcom | grep -w "$name_malc" | awk '{print $2}' | head -n 1` == $name_malc ]];then

				echo $name_malc >>$x.name
				echo $name_mcom >>$x.name

			fi

		#Reciprocal check with msp

		elif [ `echo $line | awk '{print $2}' | cut -f 1 -d "_"` == "msp" ] && [ $b -eq 0 ];then

			echo $line | awk '{print $2}' >msp.name.match

			b=$( echo "1")

			name_msp=$( cat msp.name.match )

			cat blast_v_blast_millepora_proteins_filter.txt | awk '{if ($1~/^msp/) print $0}' | awk '{if ($2!~/^msp/) print $0}' | \
			grep -w "$name_msp" | sort -k 3 -g | sort -k 11 -g >sort_k3k11_msp

			if [[ `cat sort_k3k11_msp | grep -w "$name_malc" | awk '{print $1}' | head -n 1` == $name_msp ]] && [[ `cat sort_k3k11_msp | grep -w "$name_malc" | awk '{print $2}' | head -n 1` == $name_malc ]];then

				echo $name_msp >>$x.name
```

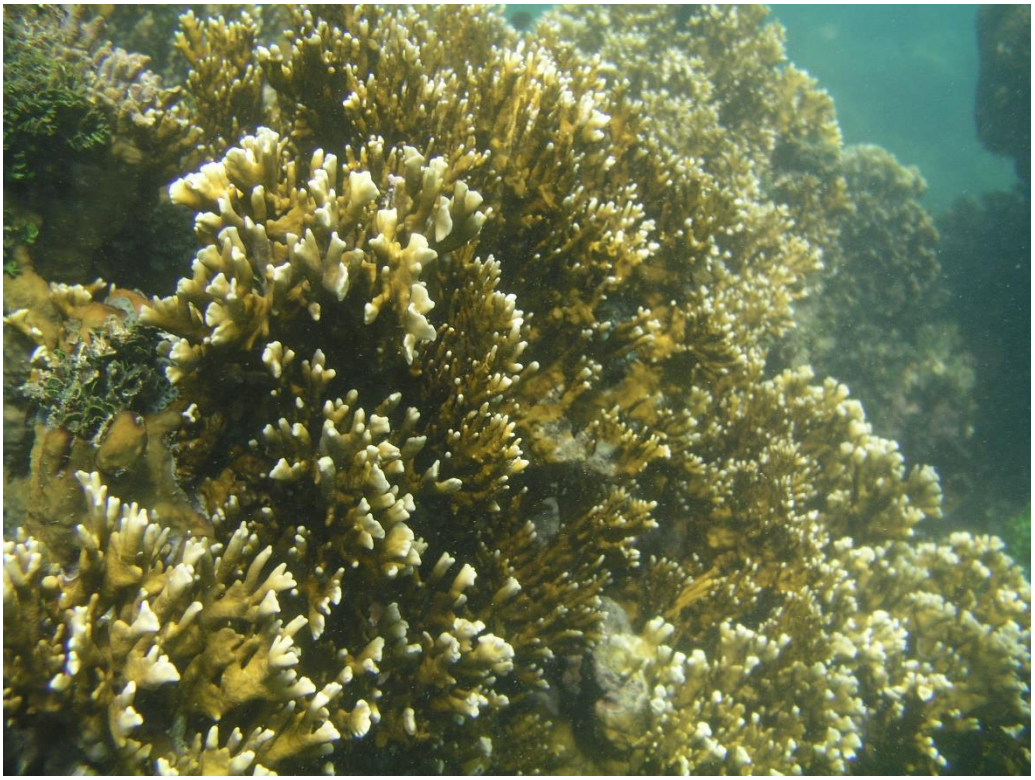
```
                fi

        #Reciprocal check with msqu

        elif [ `echo $line | awk '{print $2}' | cut -f 1 -d "_"` == "msqu" ] && [ $c -
eq 0 ];then

                echo $line | awk '{print $2}' >msqu.name.match

                c=$( echo "1")

                name_msqu=$( cat msqu.name.match )

                cat blast_v_blast_millepora_proteins_filter.txt | awk '{if
($1~/^msqu/) print $0}' | awk '{if ($2!~/^msqu/) print $0}' | \
                grep -w "$name_msqu" | sort -k 3 -g | sort -k 11 -g
>sort_k3k11_msqu

                if [[ `cat sort_k3k11_msqu | grep -w "$name_malc" | awk '{print
$1}' | head -n 1` == $name_msqu ]] && [[ `cat sort_k3k11_msqu | grep -w
"$name_malc" | awk '{print $2}' | head -n 1` == $name_malc ]];then

                        echo $name_msqu >>$x.name

                fi

        fi

    done

done


#Loop for extracting sequences that are reciprocal to each other

for x in `ls -1 malc_orf_*.name`;do

    if [ `cat $x | cut -f 1 -d "_" | sort -u | wc -l` -eq 4 ];then

        cat $x | sed 's/^/>/g' >$x.edit

        a=$( echo $x | cut -f 1 -d "." )

        ./extract_final_sequences.py millepora_proteins_filter_input.fasta $x.edit
$a.seqs
```

fi

done

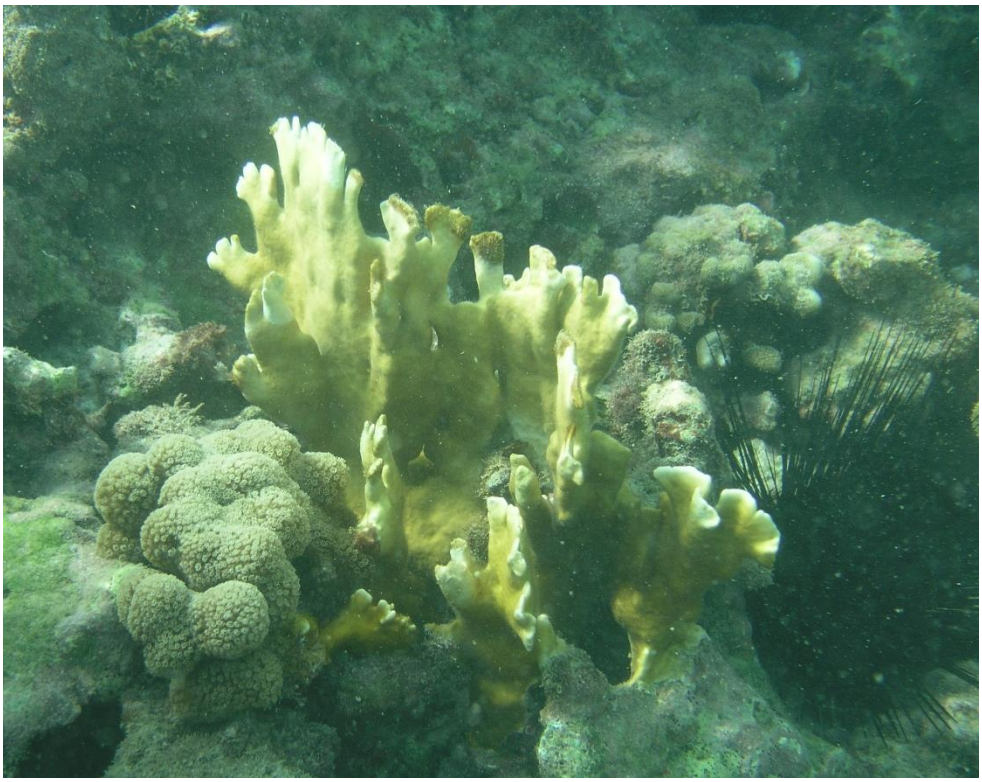
10 Pictures

10.1*Millepora alcicornis*



Appendix, Figure 14: *Millepora alcicornis* from Playita Azul Reef, southwest Puerto Rico.

10.2 Millepora complanata



Appendix, Figure 15: *Millepora complanata* from La Jungla Reef, southwest Puerto Rico.

10.3 Millepora squarrosa

Appendix, Figure 16: *Millepora squarrosa* from Turrumote II reef, southwest Puerto Rico.

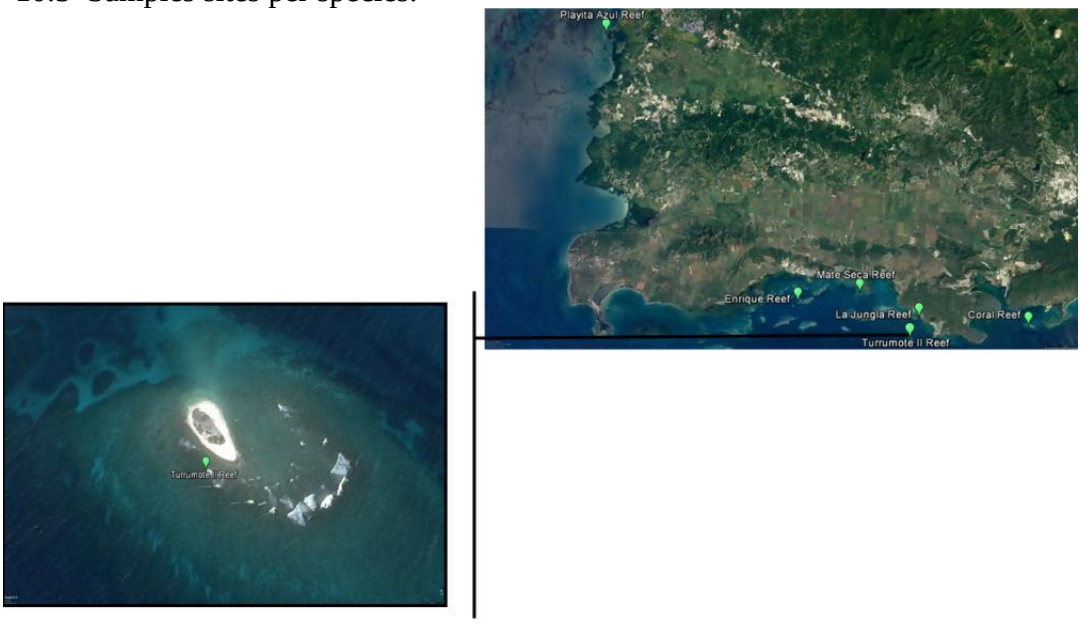
Appendix, Figure 17: *Millepora squarrosa* from Turrumote II reef, southwest Puerto Rico.

10.4 *Millepora* sp.

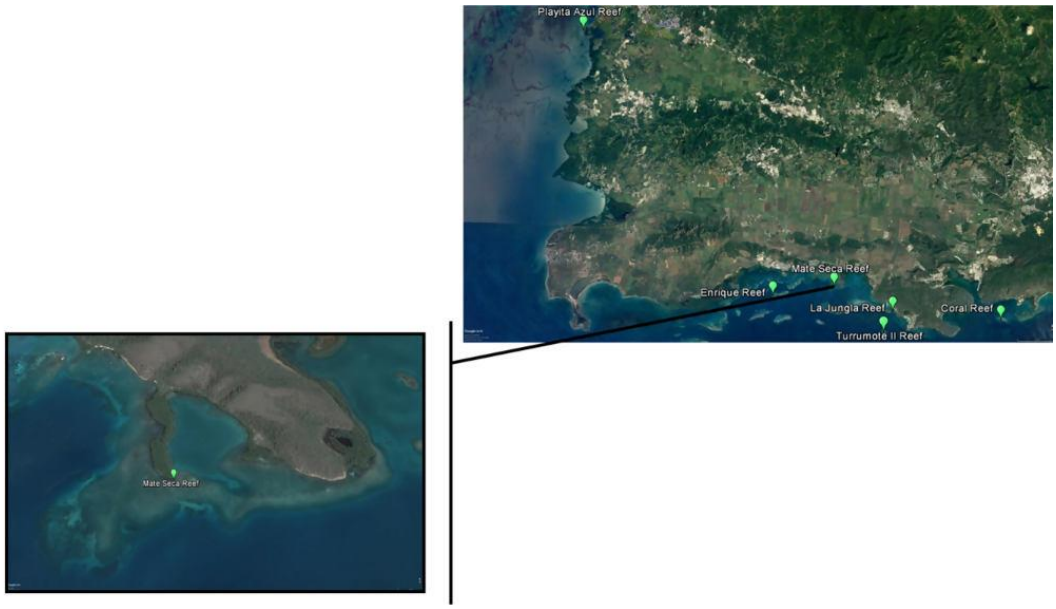
Appendix, Figure 18: *Millepora* sp. from Mata Seca Key.

Appendix, Figure 19: *Millepora* sp. from Mata Seca Key, southwest Puerto Rico.

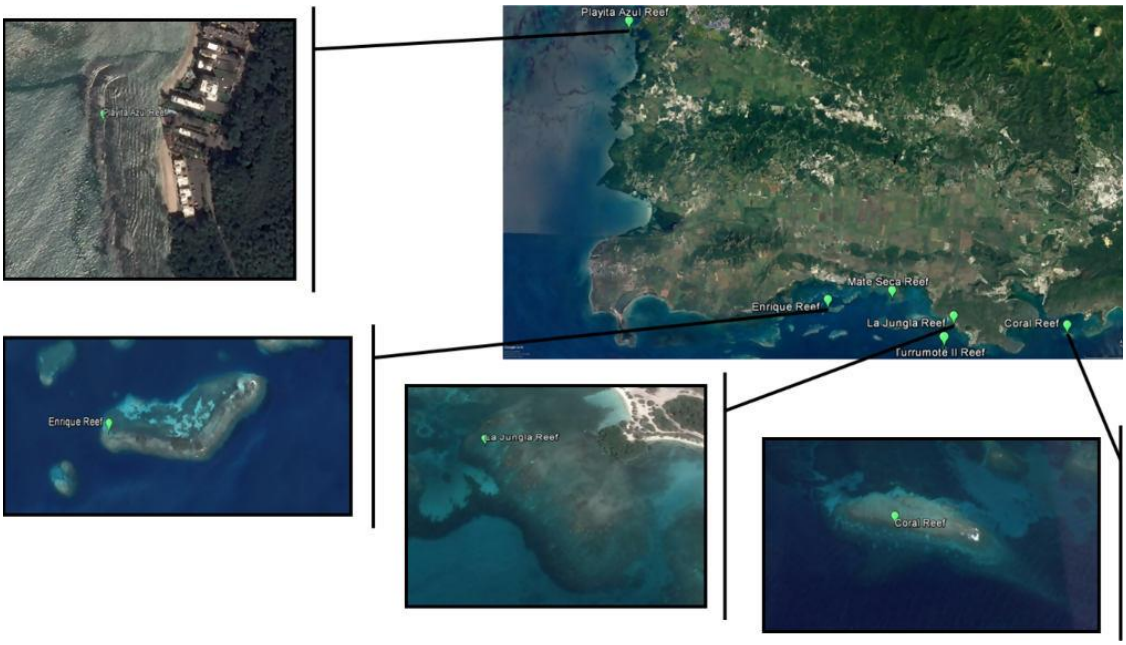
10.5 Samples sites per species.



Appendix, Figure 20: Sampling site of *Millepora squarrosa*, Turrumote II.



Appendix, Figure 21: Sampling site of *Millepora* sp., Mata Seca Key, southwest Puerto Rico.



Appendix, Figure 22: Sampling site of *Millepora alcicornis* and *Millepora complanata*, Coral Key, La Juagla Reef, Enrique Key, Playita Azul Reef.