OLLSCOIL NA GAILLIMHE
UNIVERSITY OF GALWAY

Discovery and Characterization of Novel Poriferan Biosynthetic Pathways *via* Next-Generation Sequencing

PhD candidate: Kenneth Sandoval (BSc, MSc)

Supervisor: Prof. Grace P. McCormack

Co-Supervisor: Prof. Olivier P. Thomas

A thesis submitted to the Zoology Department, Faculty of Science, University of Galway in fulfilment of the requirements of the Degree of Doctor of Philosophy

December 2022

Table of Contents

Declaration

This thesis has not been submitted, in whole or in part, to this, or any other university for degree. This work, except where otherwise stated, is the original work of the author.

Signed Kenneth W Sandoval

Date 20-12-2022

Acknowledgements

Although this thesis bears my name, I could not have done it without the support of many I have known before or met after I began this endeavour.

I would first like to thank my supervisor Prof. Grace McCormack for choosing me to carry out this project. Your continued support, even during my low points of this journey, have been fundamental for me to persevere. I believe that your curiosity, enthusiasm, and optimism were essential complements to the methodological way I approach science. I am fortunate that during my PhD I not only had a supervisor, but also a friend.

I would also like to thank my co-supervisor Prof. Olivier Thomas for your helpful knowledge on marine natural products chemistry and for your assistance with the collection of samples. This project was a multidisciplinary one and I am happy that such a collaboration existed during my time here.

My graduate research council, comprised of Prof. Anne Marie Power, Prof. Louise Allcock and Dr Aoife Boyd, also has my thanks. Each year, the three of you offered insightful support and constructive criticism on the progress of my PhD. I consider myself lucky to have had three of the best women in science watching out for me over the years.

My second family, the University of Galway Zoology Department, also has my sincerest gratitude and respect. I have never worked in a more fun, social, and quirky environment than this one. I especially owe thanks to the technicians Eoin, Maeve, Darek and Aedín. Whether it be sample collection, operating equipment, or finding the right material for an experiment you four had my back. I would further like to thank all the rest of the staff in our department for your friendly nature, helpful assistance and for often indulging my scientific curiosities.

I will always believe that one of the most amazing aspects of the PhD is the opportunity to befriend your fellow postgraduates. Yet, while you want them to succeed, one of the hardest things is seeing them finish before you and move on to new ventures. Some of you have done so, while others remain here at the end of my own journey. From talking science to enjoying pints on Friday I'm grateful to have met every one of you. I'd especially like to thank Maria Vittoria, Keith, Jose Maria, Roberto, Cesar, Michael, Edel, Kenan, Ryan, Declan, and Jamie for all the good times both inside and outside of academia.

I also owe thanks to several other individuals at the University of Galway who aided me during my project. First, Dr Laurence Jennings and Daniel Rodrigues of the Marine Biodiscovery Laboratory who respectively helped with the analysis of chemical extracts and the collection of samples. Second, Dr Antoine Fort of the Plant Systems Biology Laboratory who aided me with the planning and methodology of sequencing a genome. Third, Prof. John Pius Dalton, Dr Krystyna Cwiklinski and Jesús López of the Molecular Parasitology Laboratory who taught me techniques in synthetic biology.

This project was funded by several institutions which enabled me to carry out my work. Most significantly is the Marie Skłodowska-Curie Innovative Training Network IGNITE – Comparative Genomics of Non-Model Invertebrates. Furthermore, I was fortunate to receive funding from the Thomas Crawford Hayes 2021 Research Fund and the Higher Education Authority 2021 Covid-19 related Research Cost Extension.

Expanding on IGNITE, I would like to thank Prof. Gert Wörheide for conceptualizing and leading this incredible program. I would also like to thank all the other professors as well as my fellow early-stage researchers who were involved in IGNITE. The lectures and advice on topics such as bioinformatics and evolution were valuable to someone like me who came from a differing background. I especially want to extend gratitude towards Prof. Jean-François Flot and the Evolutionary Biology & Ecology laboratory of the Université Libre de Bruxelles for the amazing secondment. Without IGNITE and the support of its members I would not have developed into the scientist I am today.

I would like to extend my gratitude to Prof. Eric Schmidt and the Marine Natural Products Laboratory of the University of Utah for hosting me for three months. It was a dream come true to learn techniques in synthetic biology for connecting biosynthetic genes to natural products. The experience helped me realize what kind of scientist I will strive to become in the near future. It is my hope that we can continue to collaborate on the fascinating gene I discovered during my PhD.

While this process occurred in Ireland, I must also acknowledge those who have supported me before my time here. I would like to thank my childhood friends David and Jacob for their continued friendship even as I live in another part of the world. And of course, I profoundly thank my family, especially my parents Rudy and Belinda, my brother Rudy Jr, and my sister-in-law Quyen. From letting me play with bugs as a child to encouraging me to pursue my scientific career in Europe you have always supported my passion for nature and biology.

Finally, I feel I should pay homage to the one and only *Haliclona indistincta*—you beautiful, difficult, fascinating, pain in the arse. I couldn't have picked a more challenging organism to kickstart my career in genomics and biosynthesis. But the challenge is what keeps me curious about you and your fellow haplosclerids.

Abstract

Sponges of the phylum Porifera are benthic, filter feeding animals which can be found in waters throughout the world. Because of these qualities, consistent challenges they face include exposure to pathogenic microorganisms, spatial competition with other benthics, and predation from more mobile animals. To protect themselves, many sponges have developed complex chemical defences which display antimicrobial, antifouling, and antifeeding purposes. In turn, isolated chemical compounds, or natural products, from these organisms have been shown to exhibit activity towards clinically relevant targets such as pathogenic microorganisms, parasites, and tumoral cells. While derived from an animal, it has been shown that many sponge natural products are actually produced by associated microorganisms living on or within the host. Traditionally, such natural products would be isolated and characterized *via* chemical extraction, purification, structure elucidation, and bioassays. However, advances in next-generation sequencing and heterologous expression have produced an alternative process to drug discovery. First, the genome of an organism is sequenced and genes responsible for known or unknown natural products are identified *in silico via* a process known as genome mining. Second, these genes are cloned and expressed in a heterologous expression system *in vivo* to determine their connection to a natural product. This second approach to drug discovery has been used to identify the biosynthetic origin of natural products from many organisms including sponges.

The overall aim of this project was to employ this genomics-driven approach to drug discovery to identify genes responsible for new and known natural products with possible therapeutic and biotechnological applications from Irish sponges. A specific focus was to identify the biosynthetic origin of a family of compounds known as 3-alkylpyridine alkaloids (3-APs) which are highly limited to sponges of the Order Haplosclerida. These compounds are hypothesized to be produced *via* a polyketide synthase (PKS) which accepts nicotinic acid (NTA) as a starter unit. Two species found in Ireland, *Haliclona indistincta* and *Haliclona viscosa*, are known sources of 3-APs and thus were chosen to test this hypothesis. Furthermore, chemical extracts from these species have exhibited selective cytotoxicity towards tumoral cells which may be attributed to their 3-APs. However, it was a secondary goal of this project to also identify alternative genes responsible for new natural products such as sterols and larger, bioactive proteins.

Based on these goals, I first sequenced the transcriptomes of *H. indistincta* and *H. viscosa*, mapped protein-coding genes to metabolic pathways, and identified a pathway to produce

NTA from tryptophan in both species. However, a source of the alkyl chain was not identified as no complete pathway for polyketide or fatty caid biosynthesis could be identified. To overcome the limitations of transcriptomics, I then sequenced and mined the metagenome of *H. indistincta*. This allowed for the identification of a single megasynthase gene, a nonribosomal peptide synthetase-polyketide synthase (NRPS-PKS), which fit the PKS hypothesis on 3-AP biosynthesis. This NRPS-PKS gene appears to originate from a chromosome of *H. indistincta* rather than its associated microbiota which is unheard of in the field of marine natural products. By analysing publicly available sponge genomes and transcriptomes, I was able to identify similar enzymes in other sponges of the classes Demospongiae and Homoscleromorpha. This indicates that sponges themselves should not be discounted in comparison to their associated microbiota as sources of nitrogenous natural products. Because *in silico* predictive analysis could not determine whether the *H. indistincta* NRPS-PKS could create 3-APs, I attempted heterologous expression of the entire gene in *Saccharomyces cerevisiae* with the intention of *in vitro* functional characterization. No heterologous expression was observed which may indicate the presence of undetected introns within the synthesized gene. Finally, an alternative source of bioactive molecules from *H. indistincta* and *H. viscosa* was identified: an array of genes encoding for actinoporin-like proteins (ALPs). By characterizing the ALPs with *in silico* methods, I predicted that one likely has membrane binding and cytolytic capabilities similar to actinoporins from cnidarians. These ALP genes are widespread in the phylum Porifera and thus represent an untapped source of bioactive proteins with potential applications.

Chapter 1

General Introduction

Part of this chapter contributed to the as of yet unpublished manuscript:
Sandoval, K., Thomas, O.P., McCormack, G.P., 2022. From Monomers to Manzamines: Observed Patterns, Biosynthetic Considerations and Chemical Ecology of 3-alkylpyridine Alkaloids from the Order Haplosclerida.

Natural Products Research in the Genomics Era

Natural products (NPs) are chemical compounds or substances that are produced by an organism in nature. NPs encompass both central (or primary) and specialized (or secondary) metabolites. The former are directly involved in the growth, development and reproduction of an organism. The latter are not involved in these processes, but instead provide said organism with an advantage for its survival and include pigments, hormones, pheromones and toxins. Many specialized metabolites are often involved in the chemical offense and defence of the producing organism with those it interacts with. As a result, these compounds often display activities towards therapeutic targets of interest such as tumoral cells, pathogenic microorganisms, and parasites; this translates to natural products being of great interest for the development of new medicines. In support of this, as of the year 2014 it is estimated that over half of clinically approved drugs are either unaltered natural products, natural product derivatives, or synthetic mimics of natural products [1]. For the remainder of this introduction, the term "natural product" will specifically be in reference to "specialized metabolite".

Traditionally, NPs were isolated *via* extraction, fractionation, purification, and structure elucidation from target organisms in the laboratory. While this approach is essential for the pipeline of NP discovery and remains crucial today, it is not without drawbacks. For one, the re-isolation of previously discovered compounds is a frequent issue which results in wasted time and necessitates dereplication strategies [2]. Furthermore, often the genes responsible for a bioactive compound are either not expressed or lowly expressed at the time of sampling, resulting in potential NPs being overlooked [3]. Considering this, an alternative approach to NP discovery has developed which relies on the connection they have with the central dogma of molecular biology (Figure 1.1). This dogma illustrates that transcription of DNA produces RNA which is in turn translated to produce protein. Excluding cases of non-enzymatic reactions, the biosynthesis of NPs and other metabolites is overwhelmingly performed by enzymatic proteins thus linking them to this dogma. Based on this connection, the revolution in next-generation sequencing has resulted in a new, complimentary approach to drug discovery in which new or known NPs are linked to the genes responsible for their biosynthesis.

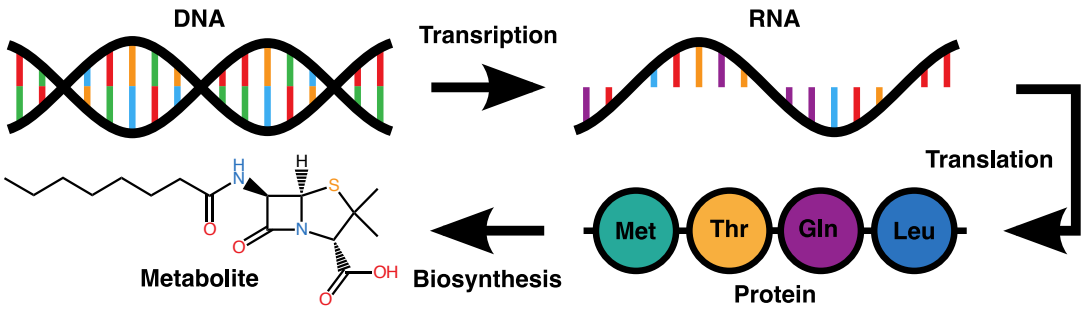


Figure 1.1 The central dogma of molecular biology with the addition of enzymatic metabolite biosynthesis.

Microorganisms have especially been the target of choice for this new approach due to several features. For one, certain taxa such as the actinobacteria are prolific sources of NPs with many having significant portions of their genome devoted to specialized metabolism [4]. Furthermore, these biosynthetic genes are often clustered together in regions of microbial genomes known as biosynthetic gene clusters (BGCs) [5]. This particular feature has allowed the development of genome mining software which screen sequenced genomes for BGCs responsible for the production of several well-studied classes of NPs such as polyketides, nonribosomal peptides, ribosomally synthesized and post-translationally modified peptides, and terpenes [6-8]. Finally, microorganisms, particularly bacteria and fungi, have significantly smaller genomes than other prolific sources of NPs such as plants and animals [9]. Due to this, high-coverage genomes are easily sequenced and assembled from microorganisms. Furthermore, axenic DNA is readily acquirable from those which can be cultured on synthetic media. This alternative approach to drug discovery has been particularly powerful in understanding the biosynthetic potential of microbial "dark matter"; those microorganisms which cannot yet be cultured [10]. By taking environmental samples, extracting and sequencing DNA, and assembling metagenomes, the biosynthetic genes of these microbial communities become available for further analyses [11].

The above approach of sequencing, assembling and mining a genome is a powerful one to estimate the biosynthetic potential of an organism or identify possible genes responsible for a known NP. However, to truly connect a gene to a natural product further *in vivo* or *in vitro* studies are necessary. Bacteria, fungi, plants and animals can be manipulated *via* means such as gene deletion studies to estimate whether a gene plays a role in the biosynthesis of a NP [12-14]. However, as mentioned previously many NPs and their corresponding BGCs are derived from uncultivable microorganisms and thus are unsuitable for such methods. Furthermore, gene editing in certain organisms which are rich sources of NPs, such as sponges, is in its infancy [15]. In cases such as these, cloning of the responsible gene or BGCs into an expression vector, transforming a heterologous host, and expressing the genes of interest represents the most thorough ways to truly link a natural product to its corresponding gene(s). In this way, the NP may be directly produced from the heterologous host which is commonly done *in vivo* with numerous bacterial species that have been modified to function as chassis [16]. Alternatively, the expression and isolation of the responsible biosynthetic enzyme(s) followed by *in vitro* characterization with suitable substrates is another method which has been applied to biosynthetic genes not suitable for microbial fermentation [17].

The Oceans as a Source of Marine Natural Products

Originally terrestrial organisms such as plants, fungi and bacteria represented the major source of NPs, but the 20th century saw new inspiration for drug discovery in the oceans of our planet. In particular, benthic organisms such as macroalgae, sponges, corals, bryozoans,

molluscs and tunicates have been found to be rich sources of marine natural products (MNPs). This is likely derived from the need of these organisms to compete with one another for space to grow, to ward off more mobile predators, and to prevent infection by pathogens. Despite many MNPs having been isolated from the aforementioned benthic macroorganisms, the true origin of these compounds has almost always been traced to microorganisms associated with said macroorganisms or in a symbiotic relationship with them [18]. However, it should be noted that some exceptions have been identified in which marine algae and invertebrates have been identified as the true source of MNPs [17,19-22]. The unique biodiversity of the oceans in comparison to terrestrial environments has also translated into a more unique chemistry with features such as nitrogen and bromine atoms being far more common in MNPs than their terrestrial counterparts [23]. To date, there are 17 clinically approved drugs which trace their origins to MNPs and an additional 29 in clinical trials [24]. For these reasons, the oceans and its denizens represent a proliferative source of new compounds with applicable bioactivities. Due to their high degree of biodiversity, tropical marine environments such as coral reefs were often the focus of choice for the discovery of new MNPs. However, more recently other underexplored regions such as the more frigid waters of Ireland have been recognized as promising sources of new MNPs [25].

To determine whether a MNP is suitable for some form of commercial exploitation, large quantities are necessary for the myriad of bioassays which can be tested. In some cases, the organism of origin has been harvested from the wild. A prime example is the story of halichondrin B from which 1 ton of harvested *Lissodendoryx* sp. yielded an unsustainable 310 mg of compound [26]. To avoid this issue as well as those associated with aquaculture, chemical synthesis is the preferred method to produce pharmaceuticals derived from NPs due to its controlled nature. Still, synthetic chemistry can often not mimic the efficiency of biosynthetic pathways which have resulted from years of evolution as is evident by the semisynthetic process for the production of trabectedin and the completely fermentative process for the production of the pre-clinical salinosporamide [27,28]. Thus, knowledge on the biosynthetic mechanisms by which MNPs are produced plays a crucial role in not just the sustainability but also the possibility of advancing a potential MNP to becoming a commercial product. This is especially so with benthic marine invertebrates which are often slow-growing and thus susceptible to overharvesting.

Sponges are aquatic animals of the phylum Porifera. They can be found in tropical, temperate, and polar waters throughout the world and at depths ranging from the shallow coasts to the deep-sea floor [29]. As adults they are members of the benthic fauna in a variety of biodiverse ecosystems such as coral reefs and the more recently recognized sponge grounds [30]. The majority of these animals derive their nutrition by filtering vast quantities of dissolved organic matter and microorganisms, although several exceptions have been reported including acquisition of nutrition from microbial symbionts or direct, carnivorous predation on smaller animals [31-33]. Due to these qualities, sponges face the

aforementioned challenges of benthic marine organisms which drive them to be prominent sources of MNPs. Indeed, sponges have been the origin of several of the previously mentioned clinically approved MNPs. The first example is the Caribbean sponge *Tectitethya crypta* from which spongothymidine and spongouridine were isolated and served as inspiration for the design of cytarabine and vidarabine [34]. Furthermore, halichondrin B from the Japanese sponge *Halichondria okadai* is the progenitor of the cytotoxic eribulin [35]. Overall sponges and their associated microbiota have been a particularly consistent and lucrative source of MNPs [36]. As such, it is likely that this phylum will continue to serve as the origin of new clinically approved MNPs or MNP-derivatives in the future. In lieu with the biodiscovery efforts being pursued in Ireland, several sponges collected in the waters of the country have been identified as promising sources of bioactive MNPs [37-40].

3-Alkylpyridine Alkaloids from Sponges of the Order Haplosclerida

Sponges of the order Haplosclerida are a particularly well-known source of many different types of MNPs such as polyacetylenes and the renieramycin alkaloids [41]. However, perhaps one of the most well-known families of MNPs from this order are the 3-alkylpyridine alkaloids (3-APs). These alkaloids were first discovered in 1978 upon the identification of the high-molecular weight polymer halitoxin from several Caribbean haplosclerid sponges (Figure 1.2) [42]. Since this initial finding, over three-hundred unique 3-AP structures have been isolated primarily from this order [43-184]. They appear to be a global phenomenon as they have been isolated from specimens throughout the waters of the world (Figure 1.3). Often these compounds, in particular the high-molecular weight polymers, are considered discardable, nuisance compounds in that they are consistently isolated from haplosclerid sponges and exhibit non-selective inhibition of pharmacological targets [185,186]. However, several smaller isolated 3-APs have considerable promise for pharmaceutical or biotechnological applications. One of the first to be discovered, manzamine A from an unidentified *Haliclona* sp., exhibited a potent IC50 value of 0.07 µg/mL towards P388 mouse leukaemia cells; since then it has also been shown to exhibit notable activity towards targets such as other tumoral cell lines, viruses and parasites [47,187]. In addition xestospongin C, originally isolated from *Neopetrosia chaliniformis* (then *Xestospongia exigua*), was found to exhibit an IC_{50} of 55 µM on the inositol 1,4,5-trisphosphate-dependent release of Ca^{2+} ions and is commercially available for assays related to this mechanism [45,188].

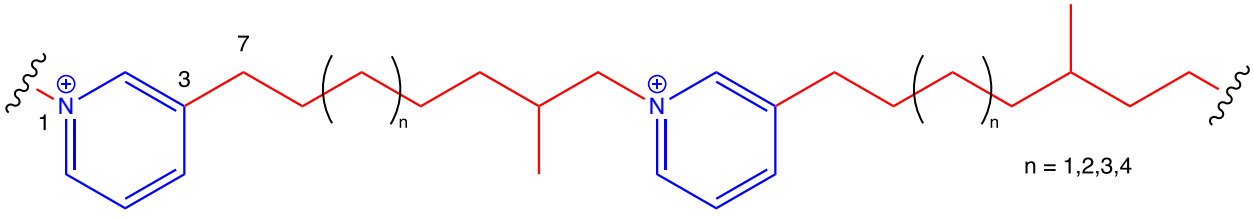


Figure 1.2 Halitoxin, the first structurally characterized 3-AP [42]. The pyridinium ring and the alkyl chain positioned on the 3rd atom of said ring are respectively highlighted in blue and red.

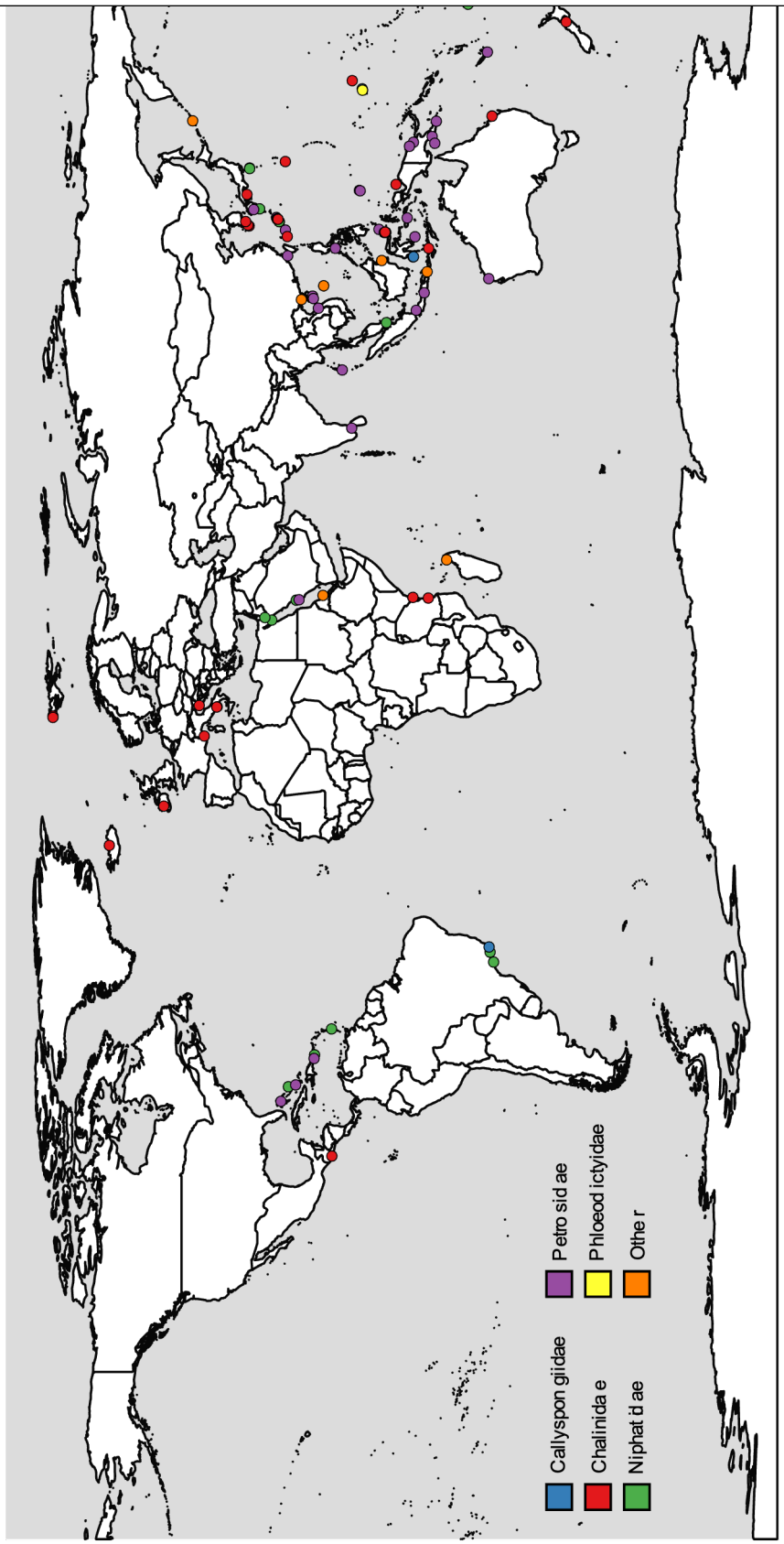


Figure 1.3 Global distribution of geographically recorded sponges from which 3-APs have been isolated. The map was created using SimpleMappr [189].

Some of the haplosclerid sponges from Ireland, namely the sister species *Haliclona indistincta* and *Haliclona viscosa*, are also known sources of 3-APs which exist in a polymerized form of undetermined length (Figure 1.4) (Olivier Thomas, personal communication) [173]. These species are of particular interest due to the selective antitumoral activity their chemical extracts display towards DU145 and MDA-MB-468 cell lines while not causing damage to normal fibroblast cells; it is believed that such activity is due to the 3-APs which are highly concentrated in their tissue (Grace P. McCormack, personal communication). Chemical analyses have indicated that the detergent-like substance exuded by these species upon physical damage comprise the polymeric 3-APs (Olivier Thomas, personal communication).

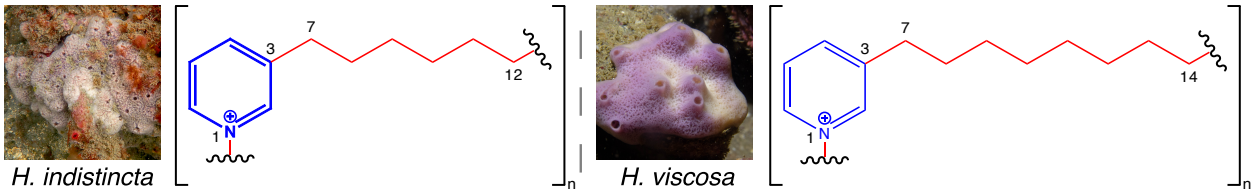


Figure 1.4 *Haliclona indistincta*, *H. viscosa*, and the polymeric 3-APs which have been respectively isolated from these species collected in Irish waters (Photograph credit: Daniel Rodrigues). The pyridinium ring and the alkyl chain positioned on the 3rd atom of said ring are respectively highlighted in blue and red.

As suggested by their name, these alkaloids are united in their common possession of one or more pyridine, pyridinium or piperidine rings as well as the position of their alkyl chains on said rings (Figure 1.2; Figure 1.4; Figure 1.5). This has resulted in a common biosynthetic origin being postulated for the 3-APs. Such a notion was first realized by Cimino and co-authors in that the saraines, xestospongins and petrosins could originate from a hypothetical, cyclic, dimeric 3-AP precursor [190]. Such a precursor was later identified upon the isolation of the cyclostellettamines and haliclamines [54,71]. Furthermore, the link between these smaller dimers with the polymeric halitoxins was also proposed by Cimino and co-authors in that coupling of alkylated pyridine or piperidine units could be a common biosynthetic process [190]. This notion was strengthened upon the identification of the monomeric niphatynes [49]. Later, Baldwin and Whitehead significantly expanded upon the perceived diversity of 3-APs by proposing a link between cyclic dimers and the more structurally complex manzamine alkaloids [191]. This hypothesis was later supported by the isolation of the ingenamine and ircinal skeletons which corresponded to proposed intermediates in the Baldwin and Whitehead hypothesis [64,77]. Based on these hypotheses Anderson and co-workers proposed an organized hierarchy of 3-APs from simple monomers to more complex derivatives which has largely remained stable since its initial publication [192].

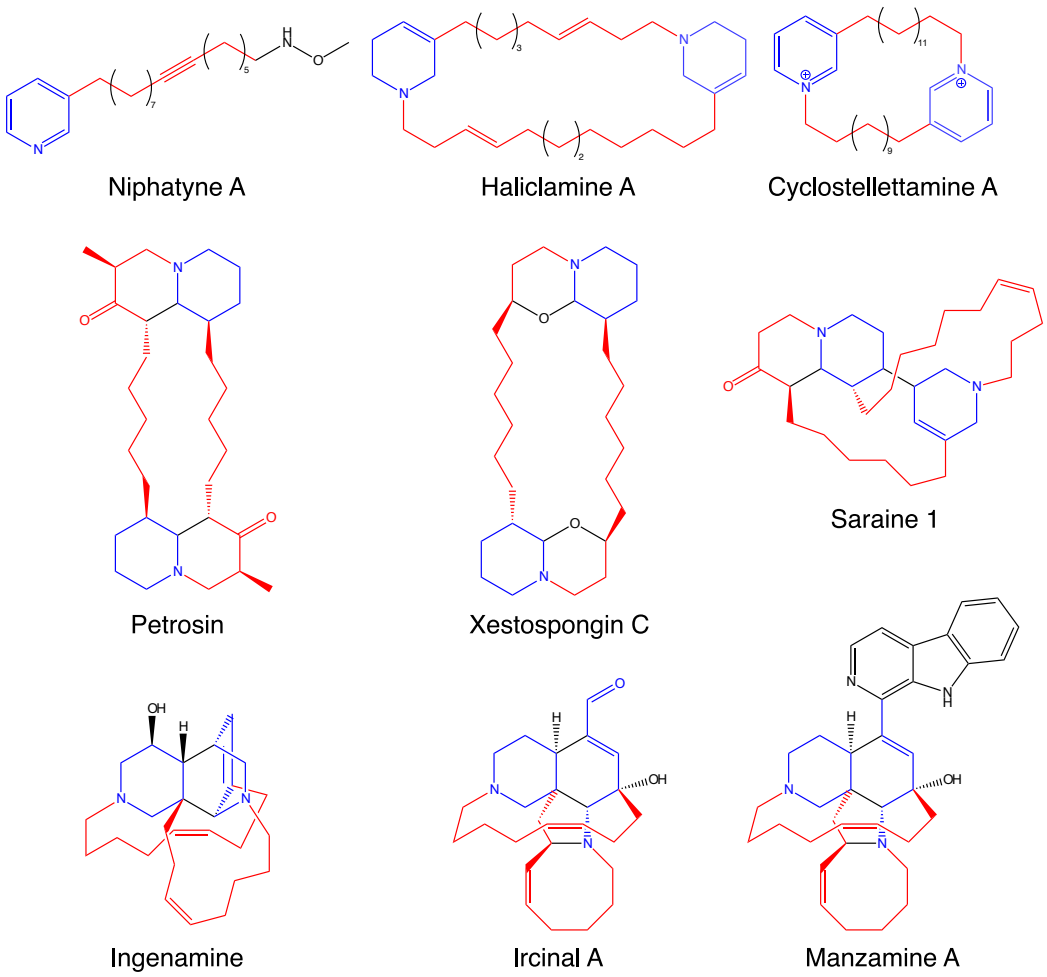


Figure 1.5 Examples of some of the first 3-APs to be discovered [43,45,46,47,49,54,64,71,77]. The common structural motifs of 3-APs, specifically the pyridine, pyridinium or piperidine rings, are highlighted in blue while the alkyl chain positioned on the 3rd atom of said ring is highlighted in red.

Broadly, there appear to be several core biosynthetic steps which differentiate the numerous types of 3-AP skeletons derived from the common building blocks. The simplest are linear monomers which appear to be the result of modification of a precursor monomeric unit without resulting in any oligomerization or cyclization. Second is the dimerization of two monomers in which the terminal ends of both alkyl chains are conjugated in a tail-to-tail fashion. Third is the bonding of the alkyl chain of one monomer with the pyridine nitrogen atom of another (or itself) in a head-to-tail as well as cyclic or linear fashion. Fourth is the formation of a bis-quinolizidine or similar skeleton from a cyclic dimer. Fifth is the intramolecular bonding between two pyridine rings of a cyclic dimer. Sixth is the opening of one pyridine ring of the fifth category resulting in the ircinol, and consequently manzamine, skeletons. Based on these observations, six general, but not exhaustive, biosynthetic categories of 3-APs can be envisioned: linear monomers (LM), linear tail-to-tail dimers (LTTD), linear and cyclic head-to-tail conjugates (LCHTC), cyclic bis-quinolizidines (CBQ), cyclic dimeric pyridine conjugates (CDPC) and ircinols and manzamines (IM) (Figure 1.6).

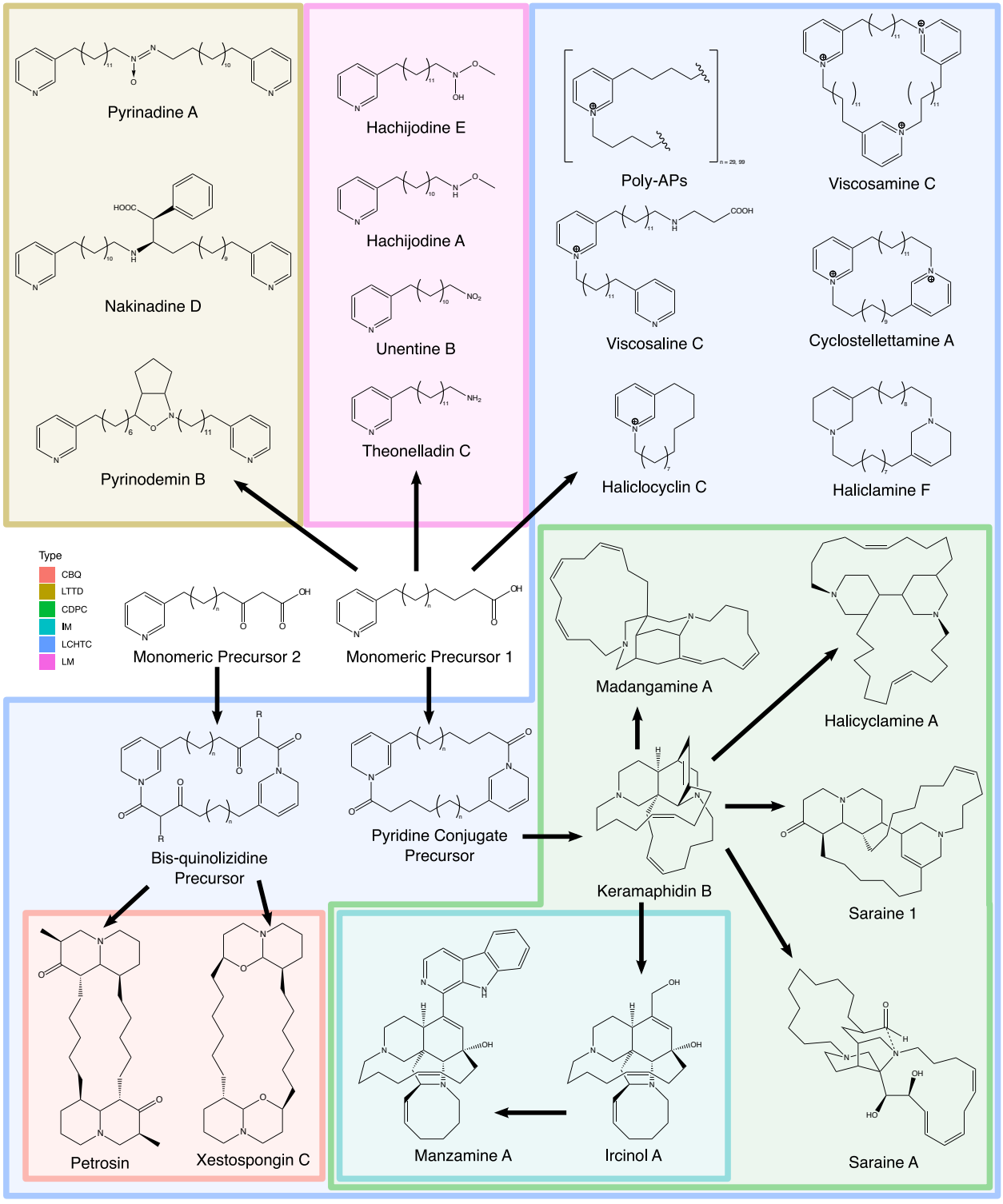


Figure 1.6 Adaptation of the biosynthetic connection of several types of 3-APs as proposed by Fontana (2006) [193]. All 3-APs are grouped into six different categories which respectively require a hypothetical unique biosynthetic step. The categories are as follows: CBQ, cyclic bis-quinolizidines; LTTD, linear tail-to-tail dimers; CDPC, cyclic dimeric pyridine conjugates; IM, ircinols and manzamines; LCHTC, linear and cyclic head-to-tail conjugates; LM, linear monomers.

By considering both instances of initial discovery and re-discovery of 3-APs it is possible to visualize a quantified distribution of the aforementioned six biosynthetic mechanisms across the different families of the Order Haplosclerida as well as identify which occur in multiple families (Figure 1.7) [42-184]. In this way, an idea on which biosynthetic steps are common across the Order or specific to certain families can be determined. As seen in the stacked UpSet plot, a majority of the 3-APs have been isolated from the Petrosiidae, Niphatidae and Chalinidae whereas the Phloedictyidae and Calyspongiidae have been less explored [194]. The LMs are distributed across all five families, although only a single example from a *Haliclona* sp. supports their presence in the Chalinidae. Furthermore, the LMs comprise most 3-APs which have been isolated from non-haplosclerid sponges. The LTTDs represent the most taxonomically specific category as they have strictly been derived from the order Niphatidae. The LCHTCs are also spread across all five families although the majority have been isolated from the Chalinidae. The CBQs, which are some of the simplest derivatives of cyclic dimers, are also taxonomically restricted to the Petrosiidae, Phloeodictyidae, and a single Niphatidae. Both the CDPCs and the IMs are widely distributed across all families except the Phloedictyidae.

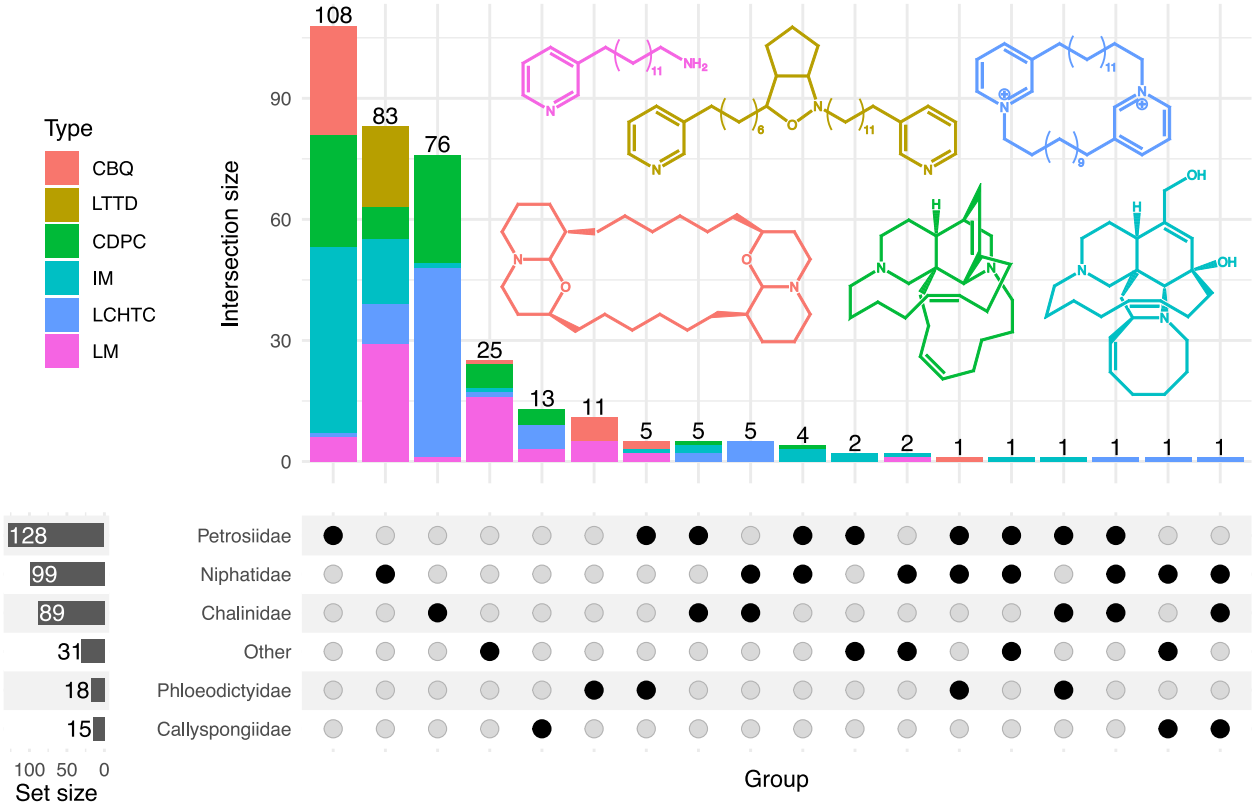


Figure 1.7 Distribution of different types of 3-APs across poriferan taxonomic groups [42-184]. Abbreviations are as follows: CBQ, cyclic bis-quinolizidines; LTTD, linear tail-to-tail dimers; CDPC, cyclic dimeric pyridine conjugates; IM, ircinols and manzamines; LCHTC, linear and cyclic head-to-tail conjugates; LM, linear monomers. The stacked UpSet plot was created using ComplexUpset version 1.3.3 [194].

Biosynthetic Considerations of 3-Alkylpyridine Monomeric Units

Based on their shared structural motifs, it appears likely that the 3-APs also share a common biosynthetic origin. However, what is enigmatic is the biosynthetic formation of the common, monomeric units they share. One of the first proposals for the latter was also by Baldwin and Whitehead in that a dihydropyridine ring could be produced by the condensation of ammonia, an acyclic dialdehyde, and a C3 subunit such as acrolein (Figure 1.8) [191]. Such a proposal also connected the 3-APs with other non-3-AP alkaloids associated with the Haplosclerida such as manzamine C, keramaphidin C, keramamine C, papuamine, haliclorensins and motuporamines A-C [70,195-198]. However, there appears to be no genetic or enzymatic evidence of this process occurring either in sponges, or other organisms.

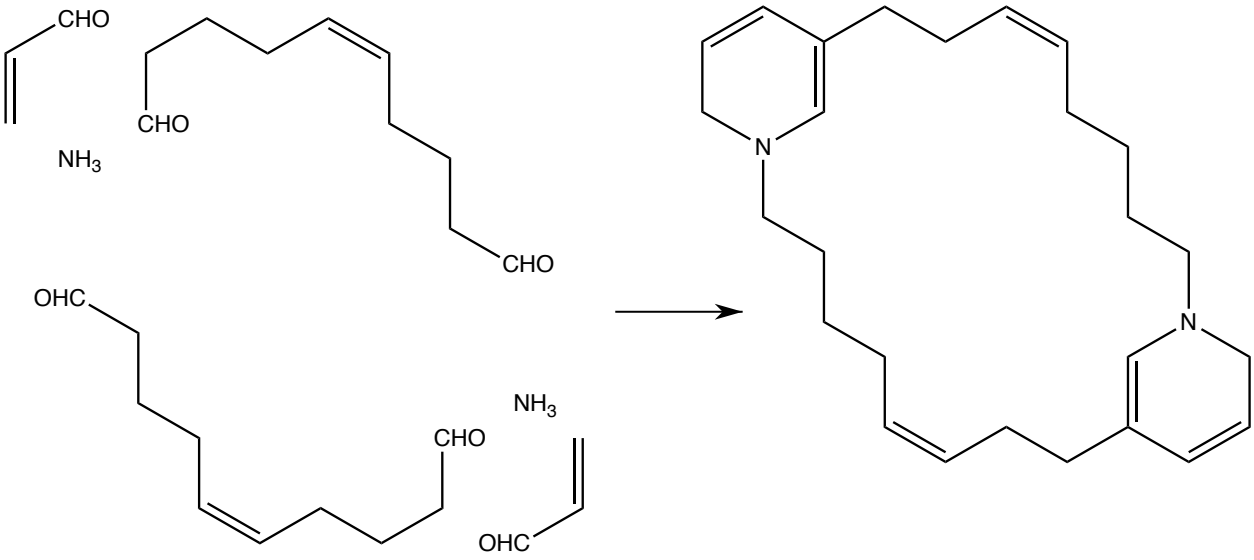


Figure 1.8 The Baldwin and Whitehead hypothesis on the biosynthesis of a 3-AP cyclic dimer [191].

A second proposal was conceived upon the isolation of the pachychaline alkaloids in that the norspermidine moiety could be a precursor to a pyridinium ring (Figure 1.9) [136,142]. In this proposal the norspermidine moiety would first be used to generate a tertiary amine. Following this, pyridinium formation would be achieved by steps of oxidation, cyclization and elimination. Similar to the Baldwin and Whitehead hypothesis, the norspermidine moiety would also provide the equivalent of a C3 subunit. In this manner, both the C-N and C-C connections between pyridinium rings observed in the pachychalines could be accounted for. In addition, the proposed processes of forming a tertiary amine were also thought to be the method in which alkaloids such as the aforementioned motuporamines are created. 3-APs possessing similar norspermidine moieties have yet to be isolated from other haplosclerid sponges which questions their role as a fundamental biosynthetic moiety.

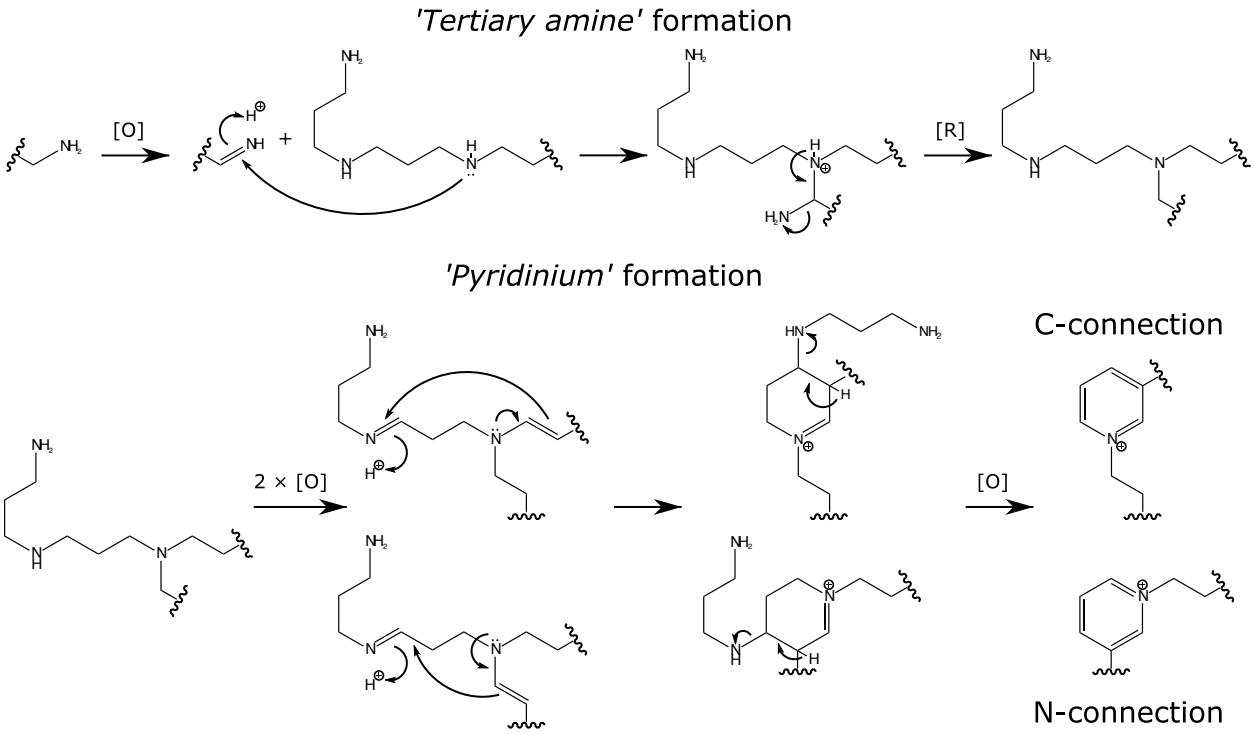


Figure 1.9 The Laville hypothesis on the biosynthesis of pyridine rings accounting for both C-N and C-C oligomerization [136,142].

An alternative theory to the origin of the pyridine ring was briefly proposed by Andersen and co-authors in that nicotinic acid (NTA) may function as a starter unit for chain elongation *via* sequential Claisen condensation reactions in a similar manner to fatty acid or polyketide biosynthesis (Figure 1.10) [192]. NTA plays an important role in the biosynthesis and recycling of the ubiquitous cofactor nicotinate mononucleotide, and it is often acquired *via* diet [199]. Alternatively, it can be recycled from nicotinamide *via* nicotinamidase in certain organisms such as bacteria, fungi and plants; genome sequencing of the haplosclerid *Amphimedon queenslandica* indicates that this sponge, and thus possibly others, are also capable of this feat [200]. This hypothesis was later used to illustrate how linear monomers derived from a hypothetical, highly reducing polyketide synthase (PKS) could function as building blocks for more complex 3-AP skeletons [193]. Of the proposed hypotheses, this is the most well-studied in biological systems in that there is a precedence in nature for NTA to be incorporated into polyketides. However, many of these observations have been primarily derived from metabolites of organisms other than sponges of the Order Haplosclerida. In this way, their biosynthesis mainly serves as inspiration for how the 3-APs may be formed.

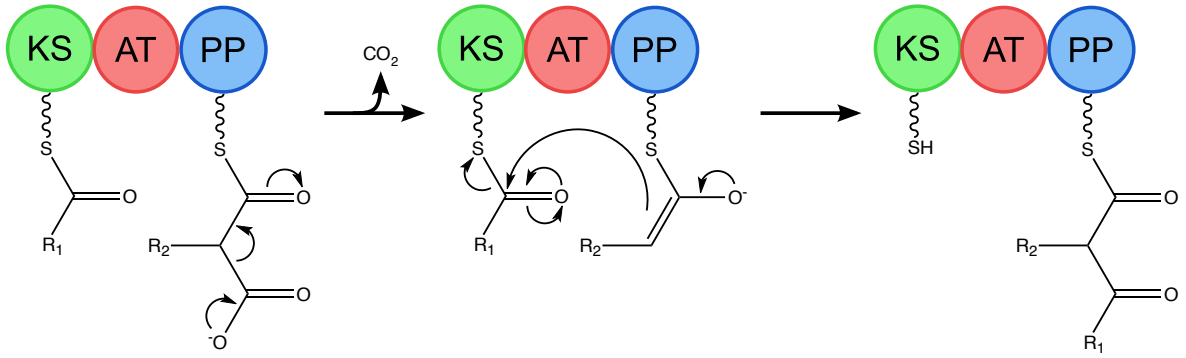


Figure 1.10 Simplified example of a Claisen condensation reaction which could be catalysed by enzymatic domains associated with polyketide or fatty acid biosynthesis. In the case of 3-AP biosynthesis, unit R_1 would be a pyridine ring. Abbreviations are as follows: KS, ketosynthase; AT, acyltransferase; PP, phosphopantetheine attachment site.

Prior to the availability of genomics and a greater understanding of biosynthetic gene clusters, isotope-labelling studies provided the best insight into how the 3-APs may be created. Initially, this was shown when the biosynthesis of haminols, alkaloids isolated from the mollusc *Haminoea orbignyana* which are structurally similar to a hypothetical, monomeric 3-AP precursor, was determined to be *via* a PKS [201]. In this biosynthetic pathway the pyridine ring and the first carbon of the alkyl-chain at C-12 were determined to originate from NTA as a deuterated form of this molecule was incorporated into the haminols without loss of deuterium. The pyridine ring and first carbon of the alkyl-chain at C-12 were found to originate from NTA. Retaining the deuterium atoms implied that the oxidation state of the pyridine ring was not changed during biosynthesis. In turn the C-2, C-4, C-6, C-8 and C-10 positions of alkyl chain were found to be constructed with sequential additions of [1-^{13}C]-acetate in a manner similar to polyketide biosynthesis. A follow-up study utilizing [1,2-^{13}C]-acetate revealed that, again like polyketide biosynthesis, the entire unit of acetate was utilized as an extender unit for the alkyl chain [202]. This study represented the first instance of a polyketide synthase (PKS) utilizing NTA as a starter unit and is the likely mechanism by which additional monomeric 3-APs from other cephalaspid molluscs are generated [203,204]. Similarly, isotope-labelling studies have shown that NTA and acetate are respectively responsible for the piperidine rings and alkyl chains of the alkaloid saraine A from *H. sarai* (Olivier Thomas, personal communication) [170]. Furthermore, tryptophan and aspartate, which are animal and bacterial precursors to NTA respectively, were both observed to be incorporated, although the former at a greater amount. While incredibly informative, these observations leave some remaining uncertainties. For one, it is unclear if the alkyl chains are derived from a PKS or a fatty acid synthase (FAS) as both would utilize acetate. Second, the incorporation of both tryptophan and aspartate raises the question as to whether the host organism, a microbial symbiont or both were creating the saraines.

Overall, despite these relevant observations on what substrates are utilized there remains a lack of evidence on the genetic level as to what enzymes are responsible for the 3-APs.

Several BGCs have been identified in bacteria and fungi which involve the incorporation of NTA into polyketide biosynthesis. While none of these microbial isolates were derived from a haplosclerid sponge, their BGCs may provide clues on possible similar genes that could be responsible for the 3-APs. The first, which is most reminiscent of a hypothetical 3-AP precursor, are the pyridinopyrones which were originally extracted from a *Streptomyces* sp. isolated from marine sediment [205]. These compounds were shown to be produced by a similar mechanism to the haminols *via* feeding experiments with nicotinic acid-d_4, ^{15}N-aspartic acid (a bacterial building block for NTA), and [1-^{13}C]-acetate. This compound was later isolated from *Streptomyces albus* J1074 and linked to a BGC encoding for an iterative type 1 PKS (Figure 1.11A) [206]. The authors reported that upstream of this PKS is a gene with the proposed function 2,3-dihydroxybenzoyl adenylate synthase which is most likely responsible for activating the NTA as a starter unit. The biosynthesis of a linear, monomeric haplosclerid 3-AP precursor may be accomplished *via* a similar system (Figure 1.11B).

There are also several instances of NTA being utilized by a PKS for compounds which are not structurally similar to 3-APs (Figure 1.12). The biosynthesis of the meroterpenoid pyripyropene A from *Aspergillus fumigatus* strain Af293 was found to involve a CoA-ligase which prepares NTA as a starting unit for PKS chain-extension through cloning and expression of the *pyr1* and *pyr2* genes [207]. In turn a set of homologous genes, *olcI* and *olcA*, were shown to be involved in the biosynthesis of the second fungal meroterpenoid 15-deoxyoxalicine B from *Penicillium canescens* strain ATCC 10419 *via* gene deletion studies [208]. Finally, there exists an instance of a standalone NRPS adenylation domain, which is involved in the biosynthesis of kosinostant by *Micromonospora* sp. TP-A0468, utilizing nicotinic acid [209]. However, it should be noted that this is promiscuous activity of the adenylation domain as its primary function is to activate aminopyrrolinic acid for the biosynthesis of the aminopyrrole moiety [210].

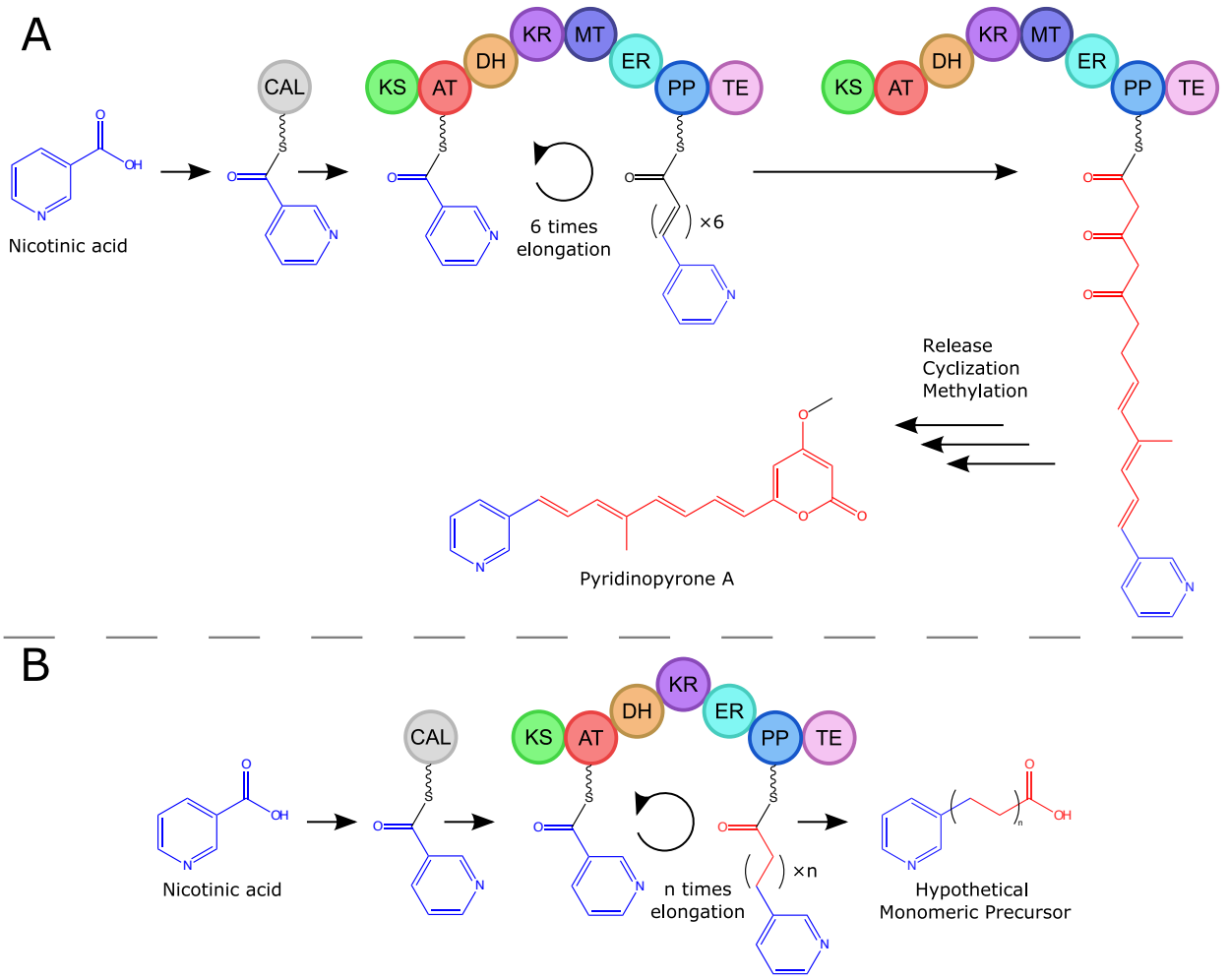


Figure 1.11 (**A**) Proposed biosynthesis of pyridinopyrone A [206]. (**B**) Hypothetical biosynthesis of a 3-AP monomer *via* a CoA ligase and a fully reducing polyketide synthase as inspired by the biosynthesis of pyridinopyrone A. Abbreviations are as follows: CAL, CoA ligase; KS, ketosynthase; AT, acyltransferase; DH, dehydratase; KR, ketoreductase; MT, methyltransferase; ER, enoylreductase; PP, phosphopantetheine attachment site; TE, thioesterase. Parts of the compound derived from NTA and a PKS are coloured blue and red respectively.

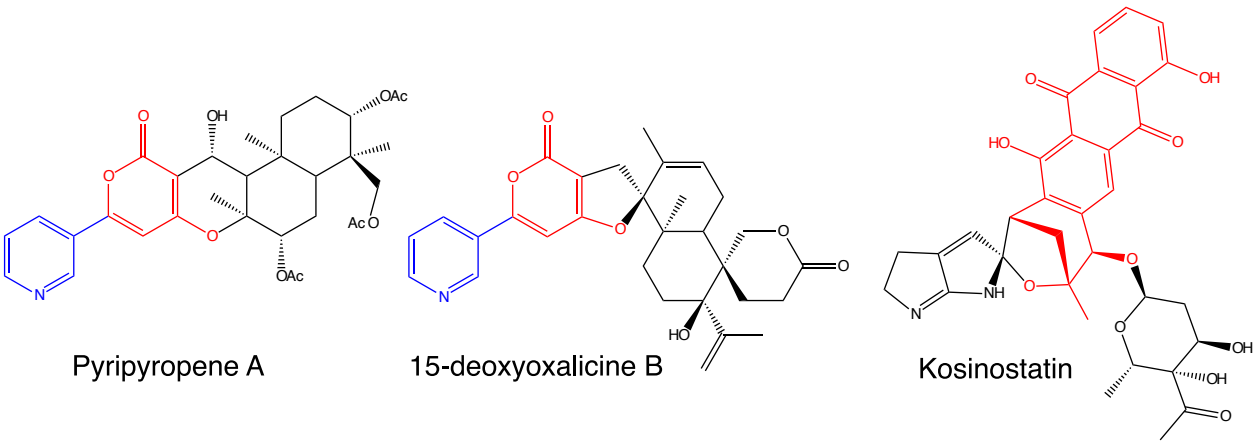


Figure 1.12 Additional polyketide natural products whose biosynthetic pathways involve the incorporation NTA or possess the capability to solely activate it. Parts of the compound derived from NTA and a PKS are coloured blue and red respectively.

An alternative to the PKS hypothesis is that 3-AP monomeric units could be the result of a condensation between NTA and an already-formed carbon chain such as a fatty acid. There are a few pieces of evidence to back up such a concept. For one, the alkyl-chains of the 3-APs are often fully or highly reduced with very few instances of ketones or alcohol groups. In addition, rings that are a result of cyclization after the release of the polyketide, such as is seen with pyridinopyrone A, are completely absent in all reported 3-APs [205]. Furthermore, skipped polyenes, a common feature of polyunsaturated fatty acids, can be observed on the alkyl-chains of several 3-APs such as madangamine A [211,212]. Typically, the ketosynthase domain of a PKS or FAS is involved in catalysing a Claisen condensation of two acyl-CoA derivatives which results in the elongation of a polyketide or fatty acid respectively (Figure 1.10). The most common units for this process are small building blocks such as malonyl-CoA. However, several examples of a ketosynthase domain utilizing a fully formed fatty acid as a substrate have been reported such as that for the biosynthesis of norsolorinic acid anthrone [213]. More recently, a ketosynthase, NocG, which catalyses the condensation of two fatty acids activated by a fatty acyl-AMP ligase at their respective carboxylic acids was reported to be involved in the biosynthesis of the cyanobacterial metabolite nocuolin A (Figure 1.13A) [214]. A similar mechanism for the condensation of NTA and a fatty acid could be envisioned followed by decarboxylation and reduction of the remaining carboxylic acid and ketone moieties respectively (Figure 1.13B). While a fatty acid origin of the alkyl chain of 3-APs has not been determined, there is a precedence for fatty acids to be building blocks for sponge-derived alkaloids such as crambescin C1-480 [215].

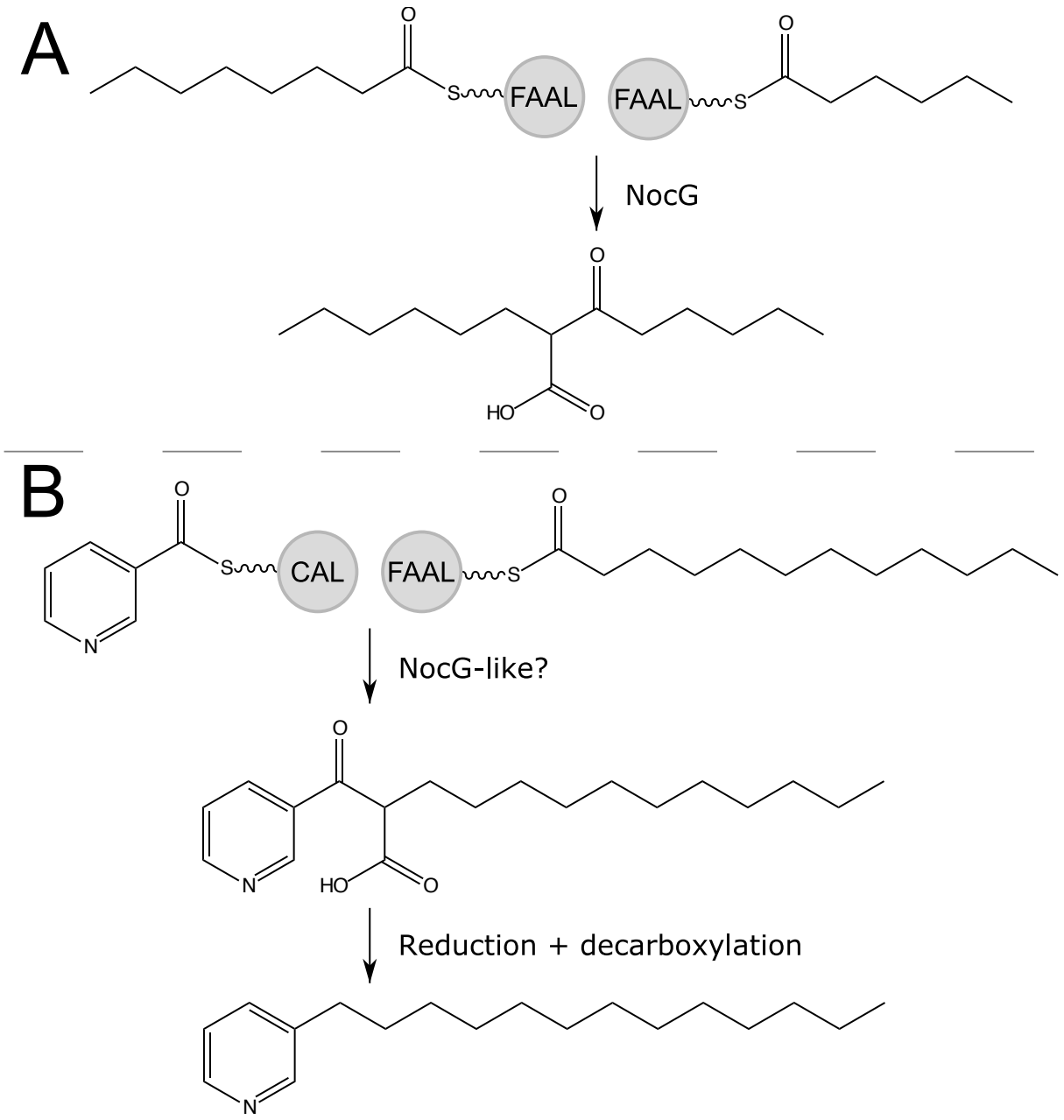


Figure 1.13 (**A**) Biosynthesis of the nocuolin A precursor [214]. (**B**) Hypothetical biosynthesis of a 3-AP monomer *via* a NocG-like pathway which involves condensation of NA with a fully formed fatty acid. Abbreviations are as follows: FAAL, Fatty acyl-AMP ligase; CAL, CoA-ligase.

Inspiration can also be drawn from plant alkaloids. Specifically, the well-known pyridine alkaloid nicotine represents a rare instance in which a condensation occurs between the 3rd atom of a pyridine ring relative to the nitrogen atom with an additional substrate [216]. While the biosynthetic pathway is not completely understood, two enzymes appear to be involved: the NADPH-dependent reductase A622 and a series of berberine bridge enzyme-like (BBE-like) protein paralogs (Figure 1.14A). A622 is thought to be involved in the activation of NTA and is hypothesized to reduce the pyridine ring [217,218]. In turn, this may

result in a decarboxylation and sequential nucleophilic attack upon an electron-deficient *N*-methylpyrollinium. This sequential condensation reaction has been hypothesized to be spontaneous. The BBE-like proteins appear to perform one of the final steps of nicotine biosynthesis as knocking down expression resulted in the production of dihydrometanicotine whereas knocking out all BBE-like genes resulted in no nicotine production [219,220]. Specifically, they are hypothesized to oxidize the pyridine ring resulting in the final product. If a similar method of condensation is involved for the 3-APs, it could also be that similar oxidizing and reducing enzymes specific to pyridine rings may play a role. Potential evidence of such a process can be inferred by the fact that reduction of the pyridine ring is a commonly observed feature in isolated 3-APs such as cyclostellettamine P, haliclamine C, and dihydrohaliclamine C which are otherwise identical in structure (Figure 1.14B) [122,170,172].

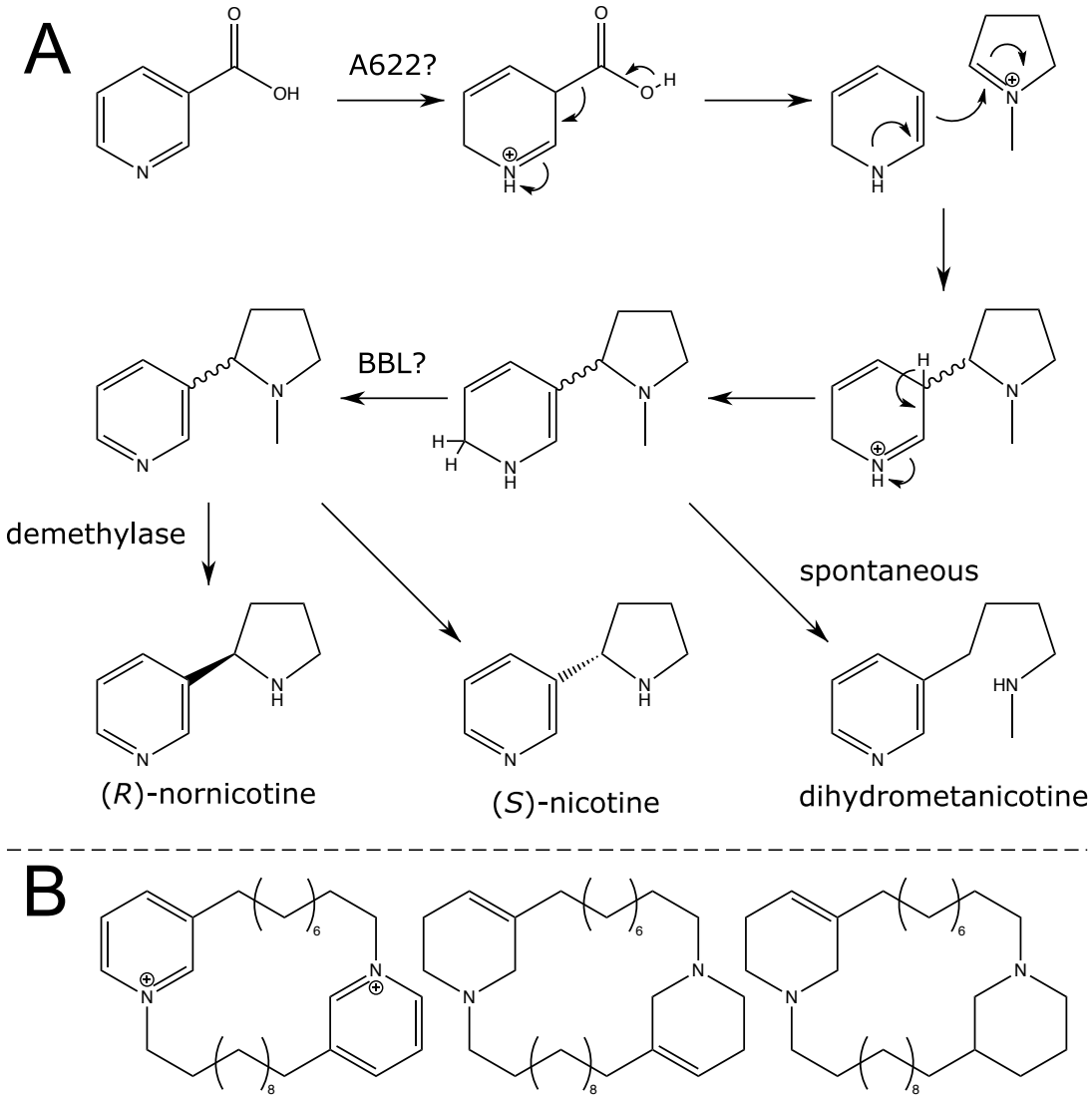


Figure 1.14 (**A**) Hypothetical biosynthesis of nicotine [216]. (**B**) Cyclostellettamine P, haliclamine C, and dihydrohaliclamine C respectively [122,170,172].

While the above proposals explain how the common monomeric building block may be produced, what is more difficult to envision is how these building blocks are oligomerized into larger 3-APs. Some synthetic studies have relied on utilizing halogen atoms at the terminal ends of the alkyl chain to function as a leaving group [221,222]. In this way the nitrogen atom of the pyridine ring of a second monomer acts as a nucleophile towards the now electron-deficient carbon atom of the first monomer resulting in oligomerization. No biosynthetic enzymes have been described which perform such a reaction; the closest is the pathway in which the dimeric cyanobacterial natural product cylindrocyclophane F is formed [223]. This pathway involves chlorination of the fatty acid moiety of one precursor monomer by the halogenase *CylC*. Following this, the enzyme *CylK* catalyses an SN2-like nucleophilic substitution with the second monomer in which the chlorine atom functions as a leaving group (Figure 1.15A). That said, halogenated 3-APs are exceptionally uncommon and have only been seen in those which are already dimerized [144,180]. An alternative to a halogen leaving group could be a quaternary ammonium moiety akin to the Hofmann elimination (Figure 1.15B). When observing the structure of isolated linear monomeric 3-APs, they almost always exhibit a nitrogenous moiety at the terminal end of their alkyl chain. In particular, several exhibit methylation of this nitrogen atom; it could be that additional methylation may result in a quaternary ammonium necessary for such a reaction [55,60,167].

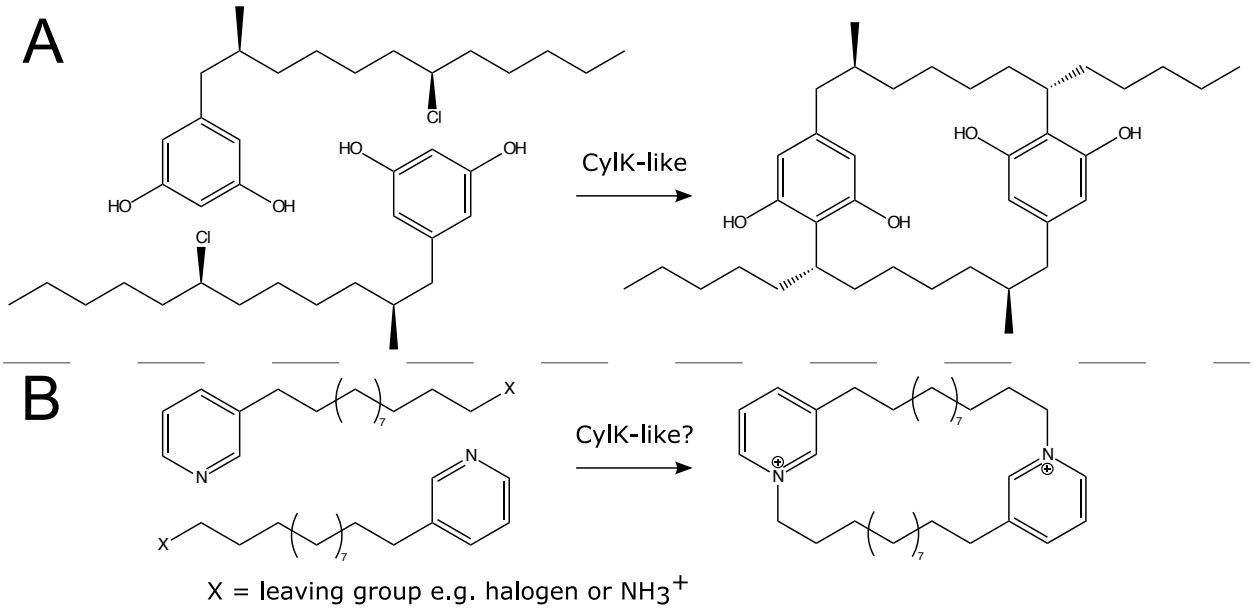


Figure 1.15 (**A**) Biosynthesis of cylindrocyclophane F [223]. (**B**) Hypothetical biosynthesis of 3-AP oligomers as inspired by the biosynthesis of cylindrocyclophane F.

Overall, the biosynthetic mechanisms from which 3-AP monomers are created and oligomerized has remained unknown since the initial discovery of halitoxin [42]. While a number of theories have been proposed, most assume reactions which have not been

observed to be catalysed by known enzymes [136,142,191]. In contrast, PKSs are a class of enzyme which have been heavily studied for decades and many of their catalytic functions are well characterized. In turn, numerous bioinformatics resources have been designed to not only identify PKS genes in genomic information, but also predict what types of substrates the PKS would utilize [6]. As such, the biosynthesis of 3-AP monomers *via* a PKS represents the most testable hypothesis in the current genomics era.

Poriferan versus Microbial Origin of 3-Alkylpyridine Alkaloids

An age-old question is whether the host organism, its associated microbiota, or a combination of both are responsible for the biosynthesis of MNPs. Sponges are particularly at the centre of this debate due to being a lucrative source of MNPs while often harbouring complex microbial communities. To date, the closest evidence of the sponge itself being the source of an MNP is the bioactive peptides known as barrettides from the sponge *Geodia barretti* as supported by genomic and transcriptomic evidence [224]. In contrast, many sponge-derived MNPs have been putatively shown to be produced by microbial symbionts [225]. With this said, regarding the majority of 3-APs it is not truly known whether the sponge or the associated microorganisms are responsible. The proposed status of 3-APs functioning as chemotaxonomic markers for the Order Haplosclerida gives initial strength to the concept of the sponge itself having a contribution in the biosynthesis of these compounds [41,226]. Furthermore, a pattern is observed in which haplosclerid sponges that are known sources of 3-APs consistently group together in a distinct clade separate from those which do not based on 28S rRNA phylogenetics (Grace McCormack, personal communication) [227]. However, previously described chemotaxonomic markers for the haplosclerids, particularly the renieramycin alkaloids, have been shown to be produced by intracellular bacterial symbionts with reduced genomes [228]. Thus, the proposed chemotaxonomic status of the 3-APs does not necessarily translate to them being a haplosclerid-specific marker unless the microbial producer is specific to the Order.

One study attempted to address the question of true origin by using Percoll density gradient centrifugation to separate and enrich different sponge cell types, as well as an intracellular dinoflagellate symbiont, of a Great Barrier Reef *Haliclona* sp. known to contain the cyclic dimeric haliclonacyclamines A-B [229]. By using thin-layer chromatography and proton NMR it was determined that the alkaloids were predominantly in fractions enriched with smaller sponge cells such as spongocytes and choanocytes. On the other hand, the fraction containing dinoflagellates was devoid of alkaloids. Initially, this may imply that the sponge cells produce these compounds or sequester them from a symbiont producer. However, caution should be taken in such an interpretation of these results as intracellular microbes were also detected in the cells of this specimen. In turn, intracellular symbionts of sponges have also been shown to be producers of some MNPs [228].

Assuming that the biosynthesis of 3-APs involves a Claisen condensation characteristic of PKS and FAS biosynthesis, some insights can also be gained from studies which surveyed the ketosynthase region through PCR amplification. In particular the low microbial abundance sponge *A. compressa*, which is known to be the source of several cationic alkylpyridinium salts, produced a single amplicon which was found to belong to the sponge ubiquitous type 1 PKS group *via* phylogenetic analysis [230]. However, this observation was not repeated *via* PCR amplification of the primers specific for this PKS group using metagenomic DNA from *A. compressa* collected in the same area [231]. In contrast, a diverse array of ketosynthase sequences were derived from the sponge *Arenosclera brasiliensis* which is known to produce the cyclic, dimeric arenosclerins A-C; a novel group of KS sequences found only from *A. brasiliensis* were shown to cluster in their own monophyletic clade representing PKS specialized metabolism possibly unique to this sponge [232].

The greatest support for a microbial role in the production of 3-APs is the reported isolation of a *Micromonospora* sp. from *Acanthostrongylophora ingens* which produced manzamine A *via* fermentation [233]. Specifically, fermentation resulted in the production of manzamine A without any sort of added precursor monomeric or dimeric 3-AP, although the addition of ircinol A, a likely 3-AP precursor to manzamine A, to the media did enhance production. Assuming that the Baldwin-Whitehead hypothesis on manzamine biosynthesis is correct, this discovery implies that the isolated *Micromonospora* sp. is also capable of *de novo* biosynthesis of a dimeric, cyclostellettamine-like 3-AP. Thus, it would possess all necessary genes for the biosynthesis of 3-APs. However, there is an issue of reproducibility as the authors reported that subsequent fermentation of this *Micromonospora* isolate resulted in a second strain incapable of producing manzamines. Yet the original glycerol stock returns the manzamine-producing strain and 16S analysis showed both strains to be identical in this regard. These observations led the authors to conclude that the biosynthetic genes responsible for manzamines may be found on a plasmid which could be lost over multiple generations of subcultures. However, despite the advances in genomics and the relative ease of sequencing and assembling a bacterial genome in the present day, no genome, plasmid, or BGCs from this isolate have been published.

The notion of a plasmid origin of a 3-AP presents an interesting possible explanation as to the diversity of 3-APs which can be found in a single sponge [233]. It has been postulated that cationic 3-AP salts may allow for the transfer of genetic information across the sponge holobiont based on their pore-forming abilities [234]. Furthermore, microbes isolated from *Amphimedon chloros* (previously *Amphimedon viridis* (Keller)) have been shown to be resistant to the lethality of the cationic halitoxin and amphitoxin present in their host unlike those isolated from surrounding seawater [235,236]. The ability of sponge-associated microbial symbionts to survive the exchange of genetic information, such as plasmids encoding the biosynthesis of 3-APs, may be mediated by these cationic 3-AP salts. In other words, this may represent the method by which different symbionts acquire genes for the

biosynthesis of a common precursor 3-AP. In turn, the unique metabolic capabilities encoded on the nuclear chromosome of the differing symbionts may then be responsible for the numerous more structurally complex and diverse 3-APs. It should be noted that cationic 3-AP salts have yet to be reported from *A. ingens* which downplays the notion that they have a role in the transfer of a hypothetical plasmid to the *Micromonospora* strain.

Any potential microbial origin of 3-AP polymers from Irish *H. indistincta* or *H. viscosa* has particularly been questioned based on both transmission electron microscopy and 16*S* rRNA sequencing which showed little consistency in microbial communities between the two as well as a very low microbial abundance status [237]. Rather, the primary noted consistency between these two species, as well as the Mediterranean *H. sarai* from which 3-APs can also be isolated, was the presence of sponge cells with inclusion bodies which were hypothesized to be associated with the 3-AP polymers and the aforementioned detergent-like secretion (Marra et al., unpublished). While 16S sequences with homology to those from *Micromospora* sp. were identified from *H. indistincta* and *H. viscosa* data, they were not consistently found among all specimens sampled. Based on these qualities, a possible animal origin of these compounds should not be ignored. Curiously, none of the smaller, non-polymeric 3-APs have been isolated from Irish specimens of these species. This contrasts with observations from arctic specimens of *H. viscosa* which may imply an environmental or microbial influence on these smaller derivatives [238].

Alternative Natural Products from Irish *Haliclona* Species

As mentioned previously, polymeric 3-APs are likely responsible for the observed antitumoral activity of extracts from *H. indistincta* and *H. viscosa*, but this type of polymeric compound is associated with nonselective activity towards pharmaceutical targets [185,186]. This potentially renders them not suitable for development into new pharmaceuticals. Based on this, alternative natural products from these species should also be considered. While no other compounds have been isolated from *H. indistincta* and *H. viscosa* collected in Irish waters, inspiration can be drawn from other Irish *Haliclona* from which MNPs derived from terpenoid precursors have been consistently discovered. *Haliclona simulans* was found to be a rewarding source of bioactive sterols which exhibited antitrypanosomal and antimycobacterial activity [38]. Furthermore, *Haliclona oculata* was found to be the source of two novel sterols bearing uncommon hydroperoxide groups [173]. In the same work, numerous previously identified sterols were also discovered from *H. oculata, H. simulans* and *Haliclona urceolus*. While not collected in Ireland, a specimen of *H. oculata* from New Brunswick was shown to be a source of novel polar sterols; it may be that other specimens of Irish *H. oculata* collected in different locations or at different times could yield similar compounds [239]. Based on the prevalence of sterols and other terpenoid derivatives in Irish *Haliclona* species it is not unreasonable to consider these types of compounds as alternative natural products to be isolated from *H. indistincta* and *H. viscosa*.

In addition to MNPs, sponges are also a source of larger, bioactive proteins with potential pharmaceutical or biotechnological applications. One of the most well-studied examples from this phylum are lectins which have largely been derived from the Demospongiae and exhibit biological activities including cell agglutination, proinflammatory modulation, and antimicrobial effects [240]. Additional examples from sponges include an antibacterial pore-forming protein as well as the cytotoxic suberitine and chondrosin [241-243]. Furthermore, the genome of the sponge *Oscarella pearsei* was found to harbour genes encoding for actinoporin-like proteins [244-246]. As their name suggests, these proteins are structurally and biochemically similar to proteinaceous α-pore-forming toxins known as actinoporins which were originally derived from cnidarians of the order Actiniaria [247]. Outside of the Actiniaria, actinoporin-like proteins have been found in diverse taxonomic groups such as hydrozoans, molluscs, and fish [248-250]. Based on their biochemical properties actinoporins, and thus potentially actinoporin-like proteins such as those from sponges, are of interest for various applied uses such as the development of immunotoxins [251]. In turn, the reported cytotoxic activities of numerous proteinaceous toxins from sponges also makes them a target of interest for identifying new cytotoxic molecules from *H. indistincta* and *H. viscosa* [240,242,243].

Discovery and characterization of Novel Poriferan Biosynthetic Pathways *via* Next-Generation Sequencing

Due to the increased affordability and efficiency of next-generation sequencing, the genetic resources of biotechnologically promising organisms such as sponges have now become accessible. Considering this opportunity, the scope of this project has been to utilize both transcriptomics and genomics to identify new genes from Irish *Haliclona* species which could have applied uses in fields such as pharmaceuticals or biotechnology. Altogether, I primarily sought to answer two questions related to these Irish *Haliclona* species and their observed bioactivity. First, can evidence of a PKS which accepts NTA as a starter unit be identified in the genomes of said species or their associated microbiota? Second, are there alternative genes associated with new natural products to be discovered? To pursue these postulations, I first sequenced new transcriptomes of the sister species *H. indistincta* and *H. viscosa*, both of which are associated with the polymeric 3-APs, to investigate whether they possessed metabolic pathways which would enable the production of hypothetical 3-AP building blocks. Furthermore, I sequenced the genome of *H. indistincta*, along with those of its associated microbiota, using both long- and short-read sequencing technology. Genome mining software was then employed to identify megasynthase genes which would best fit the hypothesis of 3-AP biosynthesis. Cloning and expression of one megasynthase gene of interest in *Saccharomyces cerevisiae* with the intent of *in vitro* functional characterization was then attempted. Tangentially, I explored the presence and characteristics of ALPs in my data as well as in all publicly available sponge genomes and transcriptomes.

References

1. Newman, D.J., Cragg, G.M., 2016. Natural products as sources of new drugs from 1981 to 2014. Journal of Natural Products 79, 629–661. https://doi.org/10.1021/acs.jnatprod.5b01055
2. P. Gaudêncio, S., Pereira, F., 2015. Dereplication: racing to speed up the natural products discovery process. Natural Product Reports 32, 779–810. https://doi.org/10.1039/C4NP00134F
3. Zhang, X., Hindra, Elliot, M.A., 2019. Unlocking the trove of metabolic treasures: activating silent biosynthetic gene clusters in bacteria and fungi. Current Opinion in Microbiology, Antimicrobials 51, 9–15. https://doi.org/10.1016/j.mib.2019.03.003
4. Udwary, D.W., Zeigler, L., Asolkar, R.N., Singan, V., Lapidus, A., Fenical, W., Jensen, P.R., Moore, B.S., 2007. Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. Proceedings of the National Academy of Sciences 104, 10376–10381. https://doi.org/10.1073/pnas.0700962104
5. Medema, M.H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J.B., Blin, K., de Bruijn, I., Chooi, Y.H., Claesen, J., Coates, R.C., Cruz-Morales, P., Duddela, S., Düsterhus, S., Edwards, D.J., Fewer, D.P., Garg, N., Geiger, C., Gomez-Escribano, J.P., Greule, A., Hadjithomas, M., Haines, A.S., Helfrich, E.J.N., Hillwig, M.L., Ishida, K., Jones, A.C., Jones, C.S., Jungmann, K., Kegler, C., Kim, H.U., Kötter, P., Krug, D., Masschelein, J., Melnik, A.V., Mantovani, S.M., Monroe, E.A., Moore, M., Moss, N., Nützmann, H.-W., Pan, G., Pati, A., Petras, D., Reen, F.J., Rosconi, F., Rui, Z., Tian, Z., Tobias, N.J., Tsunematsu, Y., Wiemann, P., Wyckoff, E., Yan, X., Yim, G., Yu, F., Xie, Y., Aigle, B., Apel, A.K., Balibar, C.J., Balskus, E.P., Barona-Gómez, F., Bechthold, A., Bode, H.B., Borriss, R., Brady, S.F., Brakhage, A.A., Caffrey, P., Cheng, Y.-Q., Clardy, J., Cox, R.J., De Mot, R., Donadio, S., Donia, M.S., van der Donk, W.A., Dorrestein, P.C., Doyle, S., Driessen, A.J.M., Ehling-Schulz, M., Entian, K.-D., Fischbach, M.A., Gerwick, L., Gerwick, W.H., Gross, H., Gust, B., Hertweck, C., Höfte, M., Jensen, S.E., Ju, J., Katz, L., Kaysser, L., Klassen, J.L., Keller, N.P., Kormanec, J., Kuipers, O.P., Kuzuyama, T., Kyrpides, N.C., Kwon, H.-J., Lautru, S., Lavigne, R., Lee, C.Y., Linquan, B., Liu, X., Liu, W., Luzhetskyy, A., Mahmud, T., Mast, Y., Méndez, C., Metsä-Ketelä, M., Micklefield, J., Mitchell, D.A., Moore, B.S., Moreira, L.M., Müller, R., Neilan, B.A., Nett, M., Nielsen, J., O'Gara, F., Oikawa, H., Osbourn, A., Osburne, M.S., Ostash, B., Payne, S.M., Pernodet, J.-L., Petricek, M., Piel, J., Ploux, O., Raaijmakers, J.M., Salas, J.A., Schmitt, E.K., Scott, B., Seipke, R.F., Shen, B., Sherman, D.H., Sivonen, K., Smanski, M.J., Sosio, M., Stegmann, E., Süssmuth, R.D., Tahlan, K., Thomas, C.M., Tang, Y., Truman, A.W., Viaud, M., Walton, J.D., Walsh, C.T., Weber, T., van Wezel, G.P., Wilkinson, B., Willey, J.M., Wohlleben, W., Wright, G.D., Ziemert, N., Zhang, C., Zotchev, S.B., Breitling, R., Takano, E., Glöckner, F.O., 2015. Minimum information about a biosynthetic gene cluster. Nature Chemical Biology 11, 625–631. https://doi.org/10.1038/nchembio.1890
6. Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M.H., Weber, T., 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Research 47, W81–W87. https://doi.org/10.1093/nar/gkz310
7. Skinnider, M.A., Johnston, C.W., Gunabalasingam, M., Merwin, N.J., Kieliszek, A.M., MacLellan, R.J., Li, H., Ranieri, M.R.M., Webster, A.L.H., Cao, M.P.T., Pfeifle, A., Spencer, N., To, Q.H., Wallace, D.P., Dejong, C.A., Magarvey, N.A., 2020. Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. Nature Communications 11, 6058. https://doi.org/10.1038/s41467-020-19986-1
8. Mungan, M.D., Alanjary, M., Blin, K., Weber, T., Medema, M.H., Ziemert, N., 2020. ARTS 2.0: feature updates and expansion of the Antibiotic Resistant Target Seeker for comparative genome mining. Nucleic Acids Research 48, W546–W552. https://doi.org/10.1093/nar/gkaa374
9. Golicz, A.A., Bayer, P.E., Bhalla, P.L., Batley, J., Edwards, D., 2020. Pangenomics comes of age: from bacteria to plant and animal applications. Trends in Genetics 36, 132–145. https://doi.org/10.1016/j.tig.2019.11.006

10. Marcy, Y., Ouverney, C., Bik, E.M., Lösekann, T., Ivanova, N., Martin, H.G., Szeto, E., Platt, D., Hugenholtz, P., Relman, D.A., Quake, S.R., 2007. Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. Proceedings of the National Academy of Sciences 104, 11889–11894. https://doi.org/10.1073/pnas.0704662104
11. Lok, C., 2015. Mining the microbial dark matter. Nature 522, 270–273. https://doi.org/10.1038/522270a
12. Marchand, J.A., Neugebauer, M.E., Ing, M.C., Lin, C.-I., Pelton, J.G., Chang, M.C.Y., 2019. Discovery of a pathway for terminal-alkyne amino acid biosynthesis. Nature 567, 420–424. https://doi.org/10.1038/s41586-019-1020-y
13. Alagoz, Y., Gurkok, T., Zhang, B., Unver, T., 2016. Manipulating the biosynthesis of bioactive compound alkaloids for next-generation metabolic engineering in opium poppy using CRISPR-Cas 9 genome editing Technology. Scientific Reports 6, 30910. https://doi.org/10.1038/srep30910
14. Liu, D., Awazu, A., Sakuma, T., Yamamoto, T., Sakamoto, N., 2019. Establishment of knockout adult sea urchins by using a CRISPR-Cas9 system. Development, Growth & Differentiation 61, 378–388. https://doi.org/10.1111/dgd.12624
15. Hesp, K., Flores Alvarez, J.L., Alexandru, A.-M., van der Linden, J., Martens, D.E., Wijffels, R.H., Pomponi, S.A., 2020. CRISPR/Cas12a-mediated gene editing in *Geodia barretti* sponge cell culture. Frontiers in Marine Science 7. https://doi.org/10.3389/fmars.2020.599825
16. Jia Zhang, J., Tang, X., S. Moore, B., 2019. Genetic platforms for heterologous expression of microbial natural products. Natural Product Reports 36, 1313–1332. https://doi.org/10.1039/C9NP00025A
17. Torres, J.P., Lin, Z., Winter, J.M., Krug, P.J., Schmidt, E.W., 2020. Animal biosynthesis of complex polyketides in a photosynthetic partnership. Nature Communications 11, 2882. https://doi.org/10.1038/s41467-020-16376-5
18. Blockley, A., Elliott, D.R., Roberts, A.P., Sweet, M., 2017. Symbiotic microbes from marine invertebrates: driving a new era of natural product drug discovery. Diversity 9, 49. https://doi.org/10.3390/d9040049
19. Li, F., Lin, Z., Torres, J.P., Hill, E.A., Li, D., Townsend, C.A., Schmidt, E.W., 2022. Sea urchin polyketide synthase spPks1 produces the naphthalene precursor to echinoderm pigments. Journal of the American Chemical Society 144, 9363–9371. https://doi.org/10.1021/jacs.2c01416
20. Scesa, P.D., Lin, Z., Schmidt, E.W., 2022. Ancient defensive terpene biosynthetic gene clusters in the soft corals. Nature Chemical Biology 18, 659–663. https://doi.org/10.1038/s41589-022-01027-1
21. Burkhardt, I., de Rond, T., Chen, P.Y.-T., Moore, B.S., 2022. Ancient plant-like terpene biosynthesis in corals. Nature Chemical Biology 18, 664–669. https://doi.org/10.1038/s41589-022-01026-2
22. Chekan, J.R., McKinnie, S.M.K., Moore, M.L., Poplawski, S.G., Michael, T.P., Moore, B.S., 2019. Scalable biosynthesis of the seaweed neurochemical, kainic acid. Angewandte Chemie 131, 8542–8545. https://doi.org/10.1002/ange.201902910
23. Shang, J., Hu, B., Wang, J., Zhu, F., Kang, Y., Li, D., Sun, H., Kong, D.-X., Hou, T., 2018. Cheminformatic insight into the differences between terrestrial and marine originated natural products. Journal of Chemical Information Modeling 58, 1182–1193. https://doi.org/10.1021/acs.jcim.8b00125
24. Wu, A.C., Jelielek, K.K., Le, H.Q., Butt, M., Newman, D.J., Glaser, K.B., Pierce, M.L., Mayer, A.M., 2022. The 2021 marine pharmacology and pharmaceuticals pipeline. The FASEB Journal 36. https://doi.org/10.1096/fasebj.2022.36.S1.L7586
25. Rae, M., Folch, H., Moniz, M.B.J., Wolff, C.W., McCormack, G.P., Rindi, F., Johnson, M.P., 2013. Marine bioactivity in Irish waters. Phytochemistry Reviews 12, 555–565. https://doi.org/10.1007/s11101-012-9227-7
26. Munro, M.H.G., Blunt, J.W., Dumdei, E.J., Hickford, S.J.H., Lill, R.E., Li, S., Battershill, C.N., Duckworth, A.R., 1999. The discovery and development of marine compounds with pharmaceutical potential, in: Osinga, R., Tramper, J., Burgess, J.G., Wijffels, R.H. (Eds.), Progress in industrial microbiology, Marine Bioprocess Engineering

27. Cuevas, C., Francesch, A., 2009. Development of Yondelis ® (trabectedin , ET-743). A semisynthetic process solves the supply problem. Natural Product Reports 26, 322–337. https://doi.org/10.1039/B808331M
28. Potts, B.C., Lam, K.S., 2010. Generating a generation of proteasome inhibitors: from microbial fermentation to total synthesis of salinosporamide A (marizomib) and other salinosporamides. Marine Drugs 8, 835–880. https://doi.org/10.3390/md8040835
29. Renard, E., Gazave, E., Fierro-Constain, L., Schenkelaars, Q., Ereskovsky, A., Vacelet, J., Borchiellini, C., 2013. Porifera (sponges): recent knowledge and new perspectives, in: eLS. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470015902.a0001582.pub2
30. Centre, U.N.E.P.W.C.M., 2010. Deep-sea sponge grounds: reservoirs of biodiversity. UN environment programme.
31. Bart, M.C., de Kluijver, A., Hoetjes, S., Absalah, S., Mueller, B., Kenchington, E., Rapp, H.T., de Goeij, J.M., 2020. Differential processing of dissolved and particulate organic matter by deep-sea sponges and their microbial symbionts. Scientific Reports 10, 17515. https://doi.org/10.1038/s41598-020-74670-0
32. Erwin, P.M., Thacker, R.W., 2008. Phototrophic nutrition and symbiont diversity of two Caribbean sponge–cyanobacteria symbioses. Marine Ecology Progress Series 362, 139–147. https://doi.org/10.3354/meps07464
33. Vacelet, J., Duport, E., 2004. Prey capture and digestion in the carnivorous sponge *Asbestopluma hypogea* (Porifera: Demospongiae). Zoomorphology 123, 179–190. https://doi.org/10.1007/s00435-004-0100-0
34. Kijjoa, A., Sawangwong, P., 2004. Drugs and cosmetics from the sea. Marine Drugs 2, 73–82.
35. McBride, A., Butler, S.K., 2012. Eribulin mesylate: A novel halichondrin B analogue for the treatment of metastatic breast cancer. American Journal of Health-System Pharmacy 69, 745–755. https://doi.org/10.2146/ajhp110237
36. R. Carroll, A., R. Copp, B., A. Davis, R., A. Keyzers, R., R. Prinsep, M., 2022. Marine natural products. Natural Product Reports. https://doi.org/10.1039/D1NP00076D
37. Kennedy, J., Baker, P., Piper, C., Cotter, P.D., Walsh, M., Mooij, M.J., Bourke, M.B., Rea, M.C., O'Connor, P.M., Ross, R.P., Hill, C., O'Gara, F., Marchesi, J.R., Dobson, A.D.W., 2009. Isolation and analysis of bacteria with antimicrobial activities from the marine sponge *Haliclona simulans* collected from Irish waters. Marine Biotechnology 11, 384–396. https://doi.org/10.1007/s10126-008-9154-1
38. Viegelmann, C., Parker, J., Ooi, T., Clements, C., Abbott, G., Young, L., Kennedy, J., Dobson, A.D.W., Edrada-Ebel, R., 2014. Isolation and identification of antitrypanosomal and antimycobacterial active steroids from the sponge *Haliclona simulans*. Marine Drugs 12, 2937–2952. https://doi.org/10.3390/md12052937
39. Afoullouss, S., Calabro, K., Genta-Jouve, G., Gegunde, S., Alfonso, A., Nesbitt, R., Morrow, C., Alonso, E., Botana, L.M., Allcock, A.L., Thomas, O.P., 2019. Treasures from the deep: characellides as anti-inflammatory lipoglycotripeptides from the sponge *Characella pachastrelloides*. Organic Letters 21, 246–251. https://doi.org/10.1021/acs.orglett.8b03684
40. Jennings, L.K., Khan, N.M.D., Kaur, N., Rodrigues, D., Morrow, C., Boyd, A., Thomas, O.P., 2019. Brominated bisindole alkaloids from the Celtic Sea sponge *Spongosorites calcicola*. Molecules 24, 3890. https://doi.org/10.3390/molecules24213890
41. Galitz, A., Nakao, Y., Schupp, P.J., Wörheide, G., Erpenbeck, D., 2021. A soft spot for chemistry–current taxonomic and evolutionary implications of sponge secondary metabolite distribution. Marine Drugs 19, 448. https://doi.org/10.3390/md19080448
42. Schmitz, F.J., Hollenbeak, K.H., Campbell, D.C., 1978. Marine natural products: halitoxin, toxic complex of several marine sponges of the genus *Haliclona*. The Journal of Organic Chemistry 43, 3916–3922. https://doi.org/10.1021/jo00414a026
43. Braekman, J.C., Daloze, D., de Abreu, P.M., Piccinni-Leopardi, C., Germain, G., van Meerssche, M., 1982. A novel type of bis-quinolizidine alkaloid from the sponge: *Petrosia seriata*. Tetrahedron Letters 23, 4277–4280. https://doi.org/10.1016/S0040-4039(00)88724-9

44. Braekman, J.C., Daloze, D., Defay, N., Zimmermann, D., 1984. Petrosin-A and -B, two new bis-quinolizidine alkaloids from the sponge *Petrosia Seriata*(1). Bulletin des Sociétés Chimiques Belges 93, 941–944. https://doi.org/10.1002/bscb.19840931102
45. Nakagawa, M., Endo, M., Tanaka, N., Gen-Pei, L., 1984. Structures of xestospongin A,B,C and D, novel vasodilativecompounds from marine sponge, *xestospongia exigua*. Tetrahedron Letters 25, 3227–3230. https://doi.org/10.1016/S0040-4039(01)91016-0
46. Cimino, G., Rosa, S.D., Stefano, S.D., Sodano, G., 1986. Marine natural products: new results from Mediterranean invertebrates. Pure and Applied Chemistry 58, 375–386. https://doi.org/10.1351/pac198658030375
47. Sakai, Ryuichi., Higa, Tatsuo., Jefford, C.W., Bernardinelli, Gerald., 1986. Manzamine A, a novel antitumor alkaloid from a sponge. The Journal of the American Chemical Society 108, 6404–6405. https://doi.org/10.1021/ja00280a055
48. Nakamura, H., Deng, S., Kobayashi, J., Ohizumi, Y., Tomotake, Y., Matsuzaki, T., Hirata, Y., 1987. Keramamine-A and -B, novel antimicrobial alkaloids from the Okinawan marine sponge *Pellina* sp. Tetrahedron Letters 28, 621–624. https://doi.org/10.1016/S0040-4039(00)95796-4
49. Quiñoà, E., Crews, P., 1987. Niphatynes, methoxylamine pyridines from the marine sponge, *Niphates* sp. Tetrahedron Letters 28, 2467–2468. https://doi.org/10.1016/S0040-4039(00)95442-X
50. Sakai, R., Kohmoto, S., Higa, T., Jefford, C.W., Bernardinelli, G., 1987. Manzamine B and C, two novel alkaloids from the sponge *Haliclona* sp. Tetrahedron Letters 28, 5493–5496. https://doi.org/10.1016/S0040-4039(00)96762-5
51. Ichiba, T., Sakai, R., Kohmoto, S., Saucy, G., Higa, T., 1988. New manzamine alkaloids from a sponge of the genus *Xestospongia*. Tetrahedron Letters 29, 3083–3086. https://doi.org/10.1016/0040-4039(88)85091-3
52. Cimino, G., Mattia, C.A., Mazzarella, L., Puliti, R., Scognamiglio, G., Spinella, A., Trivellone, E., 1989a. Unprecedented alkaloid skeleton from the Mediterranean sponge *Reniera sarai*: X-ray structure of an acetate derivative of sarain-a. Tetrahedron 45, 3863–3872. https://doi.org/10.1016/S0040-4020(01)89245-0
53. Cimino, G., Spinella, A., Trivellone, E., 1989b. Isosarain-1: A new alkaloid from the Mediterranean sponge *Reniera sarai*. Tetrahedron Letters 30, 133–136. https://doi.org/10.1016/S0040-4039(01)80344-0
54. Fusetani, N., Yasumuro, K., Matsunaga, S., Hirota, H., 1989. Haliclamines A and B, cytotoxic macrocyclic alkaloids from a sponge of the genus *Haliclona*. Tetrahedron Letters 30, 6891–6894. https://doi.org/10.1016/S0040-4039(01)93381-7
55. Kobayashi, J., Murayama, T., Ohizumi, Y., Sasaki, T., Ohta, T., Nozoe, S., 1989. Theonelladins A ~ D, novel antineoplastic pyridine alkaloids from the Okinawan marine sponge *Theonella swinhoei*. Tetrahedron Letters 30, 4833–4836. https://doi.org/10.1016/S0040-4039(01)80521-9
56. Kobayashi, M., Kawazoe, K., Kitagawa, I., 1989a. Araguspongines B, C, D, E, F, G, H, and J, new vasodilative bis-1-oxaquinolizidine alkaloids from an Okinawan marine sponge, *Xestospongia* sp. Chemical & Pharmaceutical Bulletin 37, 1676–1678. https://doi.org/10.1248/cpb.37.1676
57. Kobayashi, M., Kawazoe, K., Kitagawa, I., 1989b. Aragupetrosine a, a new vasodilative macrocyclic quinolizidine alkaloid from an Okinawan marine sponge *Xestospongia* sp. Tetrahedron Letters 30, 4149–4152. https://doi.org/10.1016/S0040-4039(00)99345-6
58. Carroll, A.R., Scheuer, P.J., 1990. Four beta-alkylpyridines from a sponge. Tetrahedron 46, 6637–6644. https://doi.org/10.1016/S0040-4020(01)87855-8
59. Kobayashi, J., Murayama, T., Kosuge, S., Kanda, F., Ishibashi, M., Kobayashi, H., Ohizumi, Y., Ohta, T., Nozoe, S., Sasaki, T., 1990. Niphatesines A–D, new antineoplastic pyridine alkaloids from the Okinawan marine sponge *Niphates* sp. Journal of the Chemical Society, Perkin Transactions 1 3301–3303. https://doi.org/10.1039/P19900003301

60. Sakemi, S., Totton, L.E., Sun, H.H., 1990. Xestamines A, B, and C, three new long-chain methoxylamine pyridines from the sponge *Xestospongia wiedenmayeri*. Journal of Natural Products 53, 995–999. https://doi.org/10.1021/np50070a038
61. Cimino, G., Fontana, A., Madaio, A., Scognamiglio, G., Trivellone, E., 1991. Application of two-dimensional shift correlated NMR techniques to the structure determination of an unusual marine alkaloid, isosaraine-2. Magnetic Resonance in Chemistry 29, 327–332. https://doi.org/10.1002/mrc.1260290408
62. Stierle, D.B., Faulkner, D.J., 1991. Antimicrobial *N*-methylpyridinium salts related to the xestamines from the Caribbean sponge *Calyx podatypa*. Journal of Natural Products 54, 1134–1136. https://doi.org/10.1021/np50076a039
63. Kobayashi, J., Zeng, C., Ishibashi, M., Shigemori, H., Sasaki, T., Mikami, Y., 1992. Niphatesines E–H, new pyridine alkaloids from the Okinawan marine sponge *Niphates* sp. Journal of the Chemical Society, Perkin Transactions 1 1291–1294. https://doi.org/10.1039/P19920001291
64. Kondo, K., Shigemori, H., Kikuchi, Y., Ishibashi, M., Sasaki, T., Kobayashi, J., 1992. Ircinals A and B from the Okinawan marine sponge *Ircinia* sp.: plausible biogenetic precursors of manzamine alkaloids. The Journal of Organic Chemistry 57, 2480–2483. https://doi.org/10.1021/jo00034a052
65. Quirion, J.-C., Sevenet, T., Husson, H.-P., Weniger, B., Debitus, C., 1992. Two new alkaloids from *Xestospongia* sp., a New Caledonian sponge. Journal of Natural Products 55, 1505–1508. https://doi.org/10.1021/np50088a017
66. Talpir, R., Rudi, A., Ilan, M., Kashman, Y., 1992. Niphatoxin A and B; two new ichthyo- and cytotoxic tripyridine alkaloids from a marine sponge. Tetrahedron Letters 33, 3033–3034. https://doi.org/10.1016/S0040-4039(00)79592-X
67. Davies-Coleman, M.T., Faulkner, D.J., Dubowchik, G.M., Roth, G.P., Polson, C., Fairchild, C., 1993. A new EGF-active polymeric pyridinium alkaloid from the sponge *Callyspongia fibrosa*. The Journal of Organic Chemistry 58, 5925–5930. https://doi.org/10.1021/jo00074a017
68. Matsunaga, S., Shinoda, K., Fusetani, N., 1993. Cribrochalinamine oxides A and B, antifungal β-substituted pyridines with an azomethine *N*-oxide from a marine sponge *Cribrochalina* sp. Tetrahedron Letters 34, 5953–5954. https://doi.org/10.1016/S0040-4039(00)73823-8
69. Rodriguez, J., Peters, B.M., Kurz, L., Schatzman, R.C., McCarley, D., Lou, L., Crews, P., 1993. An alkaloid protein kinase C inhibitor, xestocyclamine A, from the marine sponge *Xestospongia* sp. Journal of the American Chemical Society 115, 10436–10437. https://doi.org/10.1021/ja00075a100
70. Crews, P., Cheng, X.-C., Adamczeski, M., Rodríguez, J., Jaspars, M., Schmitz, F.J., Traeger, S.C., Pordesimo, E.O., 1994. 1,2,3,4-tetrahydro-8-hydroxymanzamines, alkaloids from two different haplosclerid sponges. Tetrahedron 50, 13567–13574. https://doi.org/10.1016/S0040-4020(01)85671-4
71. Fusetani, N., Asai, N., Matsunaga, S., Honda, K., Yasumuro, K., 1994. Cyclostellettamines A-F, pyridine alkaloids which inhibit binding of methyl quinuclidinyl benzilate (QNB) to muscarinic acetylcholine receptors, from the marine sponge, *Stelletta maxima*. Tetrahedron Letters 35, 3967–3970. https://doi.org/10.1016/S0040-4039(00)76715-3
72. Ichiba, T., Corgiat, J.M., Scheuer, P.J., Kelly-Borges, M., 1994. 8-Hydroxymanzamine A, a β-carboline alkaloid from a sponge, *Pachypellina* sp. Journal of Natural Products 57, 168–170. https://doi.org/10.1021/np50103a027
73. Jaspars, M., Pasupathy, V., Crews, P., 1994. A tetracyclic diamine alkaloid, halicyclamine A, from the marine sponge *Haliclona* sp. The Journal of Organic Chemistry 59, 3253–3255. https://doi.org/10.1021/jo00091a005
74. Kobayashi, J., Tsuda, M., Kawasaki, N., Matsumoto, K., Adachi, T., 1994a. Keramaphidin B, a novel pentacyclic alkaloid from a marine sponge *Amphimedon* sp.: A plausible biogenetic precursor of manzamine alkaloids. Tetrahedron Letters 35, 4383–4386. https://doi.org/10.1016/S0040-4039(00)73362-4

75. Kobayashi, J., Tsuda, M., Kawasaki, N., Sasaki, T., Mikami, Y., 1994b. 6-Hydroxymanzamine A and 3,4-dihydromanzamine A, new alkaloids from the Okinawan marine sponge *Amphimedon* sp. Journal of Natural Products 57, 1737–1740. https://doi.org/10.1021/np50114a021
76. Kong, F., Andersen, R.J., Allen, T.M., 1994a. Ingamines A and B, new cytotoxic alkaloids from the marine sponge *Xestospongia ingens*. Tetrahedron 50, 6137–6144. https://doi.org/10.1016/S0040-4020(01)80635-9
77. Kong, F., Andersen, R.J., Allen, T.M., 1994b. Ingenamine, a novel pentacyclic alkaloid from the marine sponge *Xestospongia ingens*. Tetrahedron Letters 35, 1643–1646. https://doi.org/10.1016/0040-4039(94)88308-4
78. Kong, F., Andersen, R.J., Allen, T.M., 1994c. Madangamine A, a novel cytotoxic alkaloid from the marine sponge *Xestospongia ingens*. The Journal of the American Chemical Society 116, 6007–6008. https://doi.org/10.1021/ja00092a077
79. Rodríguez, J., Crews, P., 1994. Revised structure of xestocyclamine A and description of a new analogue. Tetrahedron Letters 35, 4719–4722. https://doi.org/10.1016/S0040-4039(00)76950-4
80. Tsuda, M., Kawasaki, N., Kobayashi, J., 1994. Ircinols A and B, first antipodes of manzamine-related alkaloids from an Okinawan marine sponge. Tetrahedron 50, 7957–7960. https://doi.org/10.1016/S0040-4020(01)85280-7
81. Venkateswarlu, Y., Reddy, M.V.R., Rao, J.V., 1994. Bis-1-oxaquinolizidines from the sponge *Haliclona exigua*. Journal of Natural Products 57, 1283–1285. https://doi.org/10.1021/np50111a017
82. Albrizio, S., Ciminiello, P., Fattorusso, E., Magno, S., Pawlik, J.R., 1995. Amphitoxin, a new high molecular weight antifeedant pyridinium salt from the Caribbean sponge *Amphimedon compressa*. Journal of Natural Products 58, 647–652. https://doi.org/10.1021/np50119a002
83. Kobayashi, M., Chen, Y.-J., Aoki, S., In, Y., Ishida, T., Kitagawa, I., 1995. Four new β-carboline alkaloids isolated from two Okinawan marine sponges of *Xestospongia* sp. and *Haliclona* sp. Tetrahedron 51, 3727–3736. https://doi.org/10.1016/0040-4020(95)95723-9
84. Kong, F., Andersen, R.J., 1995. Ingenamine alkaloids isolated from the sponge *Xestospongia ingens*: structures and absolute configurations. Tetrahedron 51, 2895–2906. https://doi.org/10.1016/0040-4020(95)00043-8
85. Ohtani, I.I., Ichiba, T., Isobe, M., Kelly-Borges, M., Scheuer, P.J., 1995. Kauluamine, an unprecedented manzamine dimer from an Indonesian marine sponge, *Prianos* sp. The Journal of the American Chemical Society 117, 10743–10744. https://doi.org/10.1021/ja00148a017
86. Charan, R.D., Garson, M.J., Brereton, I.M., Willis, A.C., Hooper, J.N.A., 1996. Haliclonacyclamines A and B, cytotoxic alkaloids from the tropical marine sponge *Haliclona* sp. Tetrahedron 52, 9111–9120. https://doi.org/10.1016/0040-4020(96)00436-X
87. Edrada, R.A., Proksch, P., Wray, V., Witte, L., Müller, W.E.G., Van Soest, R.W.M., 1996. Four new bioactive manzamine-type alkaloids from the Philippine marine sponge *Xestospongia ashmorica*. Journal of Natural Products 59, 1056–1060. https://doi.org/10.1021/np9604083
88. Guo, Y., Madaio, A., Trivellone, E., Scognamiglio, G., Cimino, G., 1996. Further studies of alkaloids from *Reniera sarai*: structures of saraine-3 and isosaraine-3; absolute stereochemistry of saraine-1 and saraine-2. Tetrahedron 52, 14961–14974. https://doi.org/10.1016/0040-4020(96)00908-8
89. Harrison, B., Talapatra, S., Lobkovsky, E., Clardy, J., Crews, P., 1996. The structure and biogenetic origin of (-) halicyclamine B from a *Xestospongia* sponge. Tetrahedron Letters 37, 9151–9154. https://doi.org/10.1016/S0040-4039(96)02165-X
90. Tsuda, M., Inaba, K., Kawasaki, N., Honma, K., Kobayashi, J., 1996. Chiral resolution of (±)-keramaphidin B and isolation of manzamine L, a new β-carboline alkaloid from a sponge *Amphimedon* sp. Tetrahedron 52, 2319–2324. https://doi.org/10.1016/0040-4020(95)01057-2

91. Wang, G.-Y.-S., Kuramoto, M., Uemura, D., Yamada, A., Yamaguchi, K., Yazawa, K., 1996. Three novel anti-microfouling nitroalkyl pyridine alkaloids from the Okinawan marine sponge *Callyspongia* sp. Tetrahedron Letters 37, 1813–1816. https://doi.org/10.1016/0040-4039(96)00125-6
92. Kobayashi, J., Watanabe, D., Kawasaki, N., Tsuda, M., 1997. Nakadomarin A, a novel hexacyclic manzamine-related alkaloid from *Amphimedon* sponge. The Journal of Organic Chemistry 62, 9236–9239. https://doi.org/10.1021/jo9715377
93. Reddy, M.V.R., Faulkner, D.J., 1997. 3β,3′β-Dimethylxestospongin C, a new nis-1-oxaquinolizidine alkaloid from the Palauan sponge *Xestospongia* sp. Natural Product Letters 11, 53–59. https://doi.org/10.1080/10575639708043757
94. Sepčić, K., Guella, G., Mancini, I., Pietra, F., Serra, M.D., Menestrina, G., Tubbs, K., Maček, P., Turk, T., 1997. Characterization of anticholinesterase-active 3-alkylpyridinium polymers from the marine sponge *Reniera sarai* in aqueous solutions. Journal of Natural Products 60, 991–996. https://doi.org/10.1021/np970292q
95. Clark, R.J., Field, K.L., Charan, R.D., Garson, M.J., Brereton, M., Willis, A.C., 1998. The haliclonacyclamines, cytotoxic tertiary alkaloids from the tropical marine sponge *Haliclona* sp. Tetrahedron 54, 8811–8826. https://doi.org/10.1016/S0040-4020(98)00473-6
96. Guo, Y., Trivellone, E., Scognamiglio, G., Cimino, G., 1998. Misenine, a novel macrocyclic alkaloid with an unusual skeleton from the Mediterranean sponge *Reniera* sp. Tetrahedron 54, 541–550. https://doi.org/10.1016/S0040-4020(97)10314-3
97. Kong, F., Graziani, E.I., Andersen, R.J., 1998. Madangamines B–E, pentacyclic alkaloids from the marine sponge *Xestospongia ingens*. Journal of Natural Products 61, 267–271. https://doi.org/10.1021/np970377r
98. Tsuda, M., Watanabe, D., Kobayashi, J., 1998. Ma'eganedin A, a new manzamine alkaloid from *Amphimedon* sponge. Tetrahedron Letters 39, 1207–1210. https://doi.org/10.1016/S0040-4039(97)10842-5
99. Watanabe, D., Tsuda, M., Kobayashi, J., 1998. Three new manzamine congeners from *Amphimedon* sponge. Journal of Natural Products 61, 689–692. https://doi.org/10.1021/np970564p
100. Tsuda, M., Hirano, K., Kubota, T., Kobayashi, J., 1999. Pyrinodemin A, a cytotoxic pyridine alkaloid with an isoxazolidine moiety from sponge *Amphimedon* sp. Tetrahedron Letters 40, 4819–4820. https://doi.org/10.1016/S0040-4039(99)00852-7
101. Hirano, K., Kubota, T., Tsuda, M., Mikami, Y., Kobayashi, J., 2000. Pyrinodemins B-D, potent cytotoxic bis-pyridine alkaloids from marine sponge *Amphimedon* sp. Chemical & Pharmaceutical Bulletin 48, 974–977. https://doi.org/10.1248/cpb.48.974
102. Iwagawa, T., Kaneko, M., Okamura, H., Nakatani, M., van Soest, R.W.M., Shiro, M., 2000. A new quinolizidine alkaloid from the Papua New Guinean sponge *Xestospongia exigua*. Journal of Natural Products 63, 1310–1311. https://doi.org/10.1021/np000111b
103. Jiménez, J.I., Goetz, G., Mau, C.M.S., Yoshida, W.Y., Scheuer, P.J., Williamson, R.T., Kelly, M., 2000. 'Upenamide: an unprecedented macrocyclic alkaloid from the Indonesian sponge *Echinochalina* sp. The Journal of Organic Chemistry 65, 8465–8469. https://doi.org/10.1021/jo000789w
104. Nicholas, G.M., Molinski, T.F., 2000. Structures of cribochalines A and B, branched-chain methoxylaminoalkyl pyridines from the Micronesian sponge, *cribochalina* sp. Absolute Configuration and Enantiomeric Purity of Related O-Methyl Oximes. Tetrahedron 56, 2921–2927. https://doi.org/10.1016/S0040-4020(00)00189-7
105. Torres, Y.R., Berlinck, R.G.S., Magalhães, A., Schefer, A.B., Ferreira, A.G., Hajdu, E., Muricy, G., 2000. Arenosclerins A–C and haliclonacyclamine E, new tetracyclic alkaloids from a Brazilian endemic haplosclerid sponge *Arenosclera brasiliensis*. Journal of Natural Products 63, 1098–1105. https://doi.org/10.1021/np9905618
106. Tsukamoto, S., Takahashi, M., Matsunaga, S., Fusetani, N., van Soest, R.W.M., 2000. Hachijodines A–G: seven new cytotoxic 3-alkylpyridine alkaloids from two marine sponges of the genera *Xestospongia* and *Amphimedon*. Journal of Natural Products 63, 682–684. https://doi.org/10.1021/np9905766

107. Zhou, B.-N., Slebodnick, C., Johnson, R.K., Mattern, M.R., Kingston, D.G.I., 2000. New cytotoxic manzamine alkaloids from a Palaun sponge. Tetrahedron 56, 5781–5784. https://doi.org/10.1016/S0040-4020(00)00534-2
108. El Sayed, K.A., Kelly, M., Kara, U.A.K., Ang, K.K.H., Katsuyama, I., Dunbar, D.C., Khan, A.A., Hamann, M.T., 2001. New manzamine alkaloids with potent activity against infectious diseases. The Journal of the American Chemical Society 123, 1804–1808. https://doi.org/10.1021/ja002073o
109. Chill, L., Yosief, T., Kashman, Y., 2002. Halichondramine, a new tetracyclic bipiperidine alkaloid from the marine sponge *Halichondria* sp. Journal of Natural Products 65, 1738–1741. https://doi.org/10.1021/np0200663
110. Moon, S.-S., MacMillan, J.B., Olmstead, M.M., Ta, T.A., Pessah, I.N., Molinski, T.F., 2002. (+)-7S-hydroxyxestospongin A from the marine sponge *Xestospongia* sp. and absolute configuration of (+)-xestospongin D. Journal of Natural Products 65, 249–254. https://doi.org/10.1021/np010427z
111. Orabi, K.Y., El Sayed, K.A., Hamann, M.T., Dunbar, D.C., Al-Said, M.S., Higa, T., Kelly, M., 2002. Araguspongines K and L, new bioactive bis-1-oxaquinolizidine *N*-oxide alkaloids from Red Sea specimens of *Xestospongia exigua*. Journal of Natural Products 65, 1782–1785. https://doi.org/10.1021/np0202226
112. Yousaf, M., El Sayed, K.A., Rao, K.V., Lim, C.W., Hu, J.-F., Kelly, M., Franzblau, S.G., Zhang, F., Peraud, O., Hill, R.T., Hamann, M.T., 2002. 12,34-oxamanzamines, novel biocatalytic and natural products from manzamine producing Indo-Pacific sponges. Tetrahedron 58, 7397–7402. https://doi.org/10.1016/S0040-4020(02)00825-6
113. Peng, J., Hu, J.-F., Kazi, A.B., Li, Z., Avery, M., Peraud, O., Hill, R.T., Franzblau, S.G., Zhang, F., Schinazi, R.F., Wirtz, S.S., Tharnish, P., Kelly, M., Wahyuono, S., Hamann, M.T., 2003. Manadomanzamines A and B: a novel alkaloid ring system with potent activity against mycobacteria and HIV-1. The Journal of the American Chemical Society 125, 13382–13386. https://doi.org/10.1021/ja030087z
114. Rao, K.V., Santarsiero, B.D., Mesecar, A.D., Schinazi, R.F., Tekwani, B.L., Hamann, M.T., 2003. New manzamine alkaloids with activity against infectious and tropical parasitic diseases from an Indonesian sponge. Journal of Natural Products 66, 823–828. https://doi.org/10.1021/np020592u
115. Volk, C.A., Köck, M., 2003. Viscosamine: The first naturally occurring trimeric 3-alkyl pyridinium alkaloid. Organic Letters 5, 3567–3569. https://doi.org/10.1021/ol035006i
116. de Oliveira, J.H.H.L., Grube, A., Köck, M., Berlinck, R.G.S., Macedo, M.L., Ferreira, A.G., Hajdu, E., 2004. Ingenamine G and Cyclostellettamines G–I, K, and L from the new Brazilian species of marine sponge *Pachychalina* sp. Journal of Natural Products 67, 1685–1689. https://doi.org/10.1021/np0498713
117. Liu, H., Mishima, Y., Fujiwara, T., Nagai, H., Kitazawa, A., Mine, Y., Kobayashi, H., Yao, X., Yamada, J., Oda, T., Namikoshi, M., 2004. Isolation of araguspongine M, a new stereoisomer of an araguspongine/xestospongin alkaloid, and dopamine from the marine sponge *Neopetrosia exigua* collected in Palau. Marine Drugs 2, 154–163. https://doi.org/10.3390/md204154
118. Matsunaga, S., Miyata, Y., van Soest, R.W.M., Fusetani, N., 2004. Tetradehydrohalicyclamine A and 22-hydroxyhalicyclamine A, new cytotoxic bis-piperidine alkaloids from a marine sponge *Amphimedon* sp. Journal of Natural Products 67, 1758–1760. https://doi.org/10.1021/np049824a
119. Oku, N., Nagai, K., Shindoh, N., Terada, Y., van Soest, R.W.M., Matsunaga, S., Fusetani, N., 2004. Three new cyclostellettamines, which inhibit histone deacetylase, from a marine sponge of the genus *Xestospongia*. Bioorganic & Medicinal Chemistry Letters 14, 2617–2620. https://doi.org/10.1016/j.bmcl.2004.02.062
120. Rao, K.V., Kasanah, N., Wahyuono, S., Tekwani, B.L., Schinazi, R.F., Hamann, M.T., 2004. Three new manzamine alkaloids from a common Indonesian sponge and their activity against infectious and tropical parasitic diseases. Journal of Natural Products 67, 1314–1318. https://doi.org/10.1021/np0400095
121. Volk, C.A., Köck, M., 2004. Viscosaline: new 3-alkyl pyridinium alkaloid from the Arctic sponge *Haliclona viscosa*. Organic & Biomolecular Chemistry 2, 1827–1830. https://doi.org/10.1039/B403413A

122. Volk, C.A., Lippert, H., Lichte, E., Köck, M., 2004. Two new haliclamines from the Arctic sponge *Haliclona viscosa*. European Journal of Organic Chemistry 2004, 3154–3158. https://doi.org/10.1002/ejoc.200400026

123. Yousaf, M., Hammond, N.L., Peng, J., Wahyuono, S., McIntosh, K.A., Charman, W.N., Mayer, A.M.S., Hamann, M.T., 2004. New manzamine alkaloids from an Indo-Pacific sponge. Pharmacokinetics, Oral Availability, and the Significant Activity of Several Manzamines against HIV-I, AIDS Opportunistic Infections, and Inflammatory Diseases. Journal of Medicinal Chemistry 47, 3512–3517. https://doi.org/10.1021/jm030475b

124. Ondeyka, J.G., Herath, K. b., Jayasuriya, H., Polishook, J.D., Bills, G.F., Dombrowski, A.W., Mojena, M., Koch, G., DiSalvo, J., DeMartino, J., Guan, Z., Nanakorn, W., Morenberg, C.M., Balick, M.J., Stevenson, D.W., Slattery, M., Borris, R.P., Singh, S.B., 2005. Discovery of structurally diverse natural product antagonists of chemokine receptor CXCR3. Molecular Diversity 9, 123–129. https://doi.org/10.1007/s11030-005-1296-8

125. Kariya, Y., Kubota, T., Fromont, J., Kobayashi, J., 2006a. Pyrinadine A, a novel pyridine alkaloid with an azoxy moiety from sponge *Cribrochalina* sp. Tetrahedron Letters 47, 997–998. https://doi.org/10.1016/j.tetlet.2005.11.163

126. Kariya, Y., Kubota, T., Fromont, J., Kobayashi, J., 2006b. Pyrinadines B–G, new bis-pyridine alkaloids with an azoxy moiety from sponge *Cribrochalina* sp. Bioorganic & Medicinal Chemistry 14, 8415–8419. https://doi.org/10.1016/j.bmc.2006.08.045

127. Rao, K.V., Donia, M.S., Peng, J., Garcia-Palomero, E., Alonso, D., Martinez, A., Medina, M., Franzblau, S.G., Tekwani, B.L., Khan, S.I., Wahyuono, S., Willett, K.L., Hamann, M.T., 2006. Manzamine B and E and ircinal A related alkaloids from an Indonesian *Acanthostrongylophora* sponge and their activity against infectious, tropical parasitic, and Alzheimer's diseases. Journal of Natural Products 69, 1034–1040. https://doi.org/10.1021/np0601399

128. Takekawa, Y., Matsunaga, S., van Soest, R.W.M., Fusetani, N., 2006. Amphimedosides, 3-alkylpyridine glycosides from a marine sponge *Amphimedon* sp. Journal of Natural Products 69, 1503–1505. https://doi.org/10.1021/np060122q

129. Teruya, T., Kobayashi, K., Suenaga, K., Kigoshi, H., 2006. Cyclohaliclonamines A–E: dimeric, trimeric, tetrameric, pentameric, and hexameric 3-alkyl pyridinium alkaloids from a marine sponge *Haliclona* sp. Journal of Natural Products 69, 135–137. https://doi.org/10.1021/np050308+

130. Buchanan, M.S., Carroll, A.R., Addepalli, R., Avery, V.M., Hooper, J.N.A., Quinn, R.J., 2007. Niphatoxin C, a cytotoxic tripyridine alkaloid from *Callyspongia* sp. Journal of Natural Products 70, 2040–2041. https://doi.org/10.1021/np070366q

131. de Oliveira, J.H.H.L., Nascimento, A.M., Kossuga, M.H., Cavalcanti, B.C., Pessoa, C.O., Moraes, M.O., Macedo, M.L., Ferreira, A.G., Hajdu, E., Pinheiro, U.S., Berlinck, R.G.S., 2007. Cytotoxic alkylpiperidine alkaloids from the Brazilian marine sponge *Pachychalina alcaloidifera*. Journal of Natural Products 70, 538–543. https://doi.org/10.1021/np060450q

132. Kubota, T., Nishi, T., Fukushi, E., Kawabata, J., Fromont, J., Kobayashi, J., 2007. Nakinadine A, a novel bis-pyridine alkaloid with a β-amino acid moiety from sponge *Amphimedon* sp. Tetrahedron Letters 48, 4983–4985. https://doi.org/10.1016/j.tetlet.2007.05.121

133. Reyes, F., Fernández, R., Urda, C., Francesch, A., Bueno, S., de Eguilior, C., Cuevas, C., 2007. Njaoamines A–F, new cytotoxic polycyclic alkaloids from the haplosclerid sponge *Reniera* sp. Tetrahedron 63, 2432–2438. https://doi.org/10.1016/j.tet.2007.01.013

134. Sorek, H., Rudi, A., Benayahu, Y., Kashman, Y., 2007. Njaoamines G and H, two new cytotoxic polycyclic alkaloids and a tetrahydroquinolone from the marine sponge *Neopetrosia* sp. Tetrahedron Letters 48, 7691–7694. https://doi.org/10.1016/j.tetlet.2007.08.079

135. Xu, N.J., Sun, X., Yan, X.J., 2007. A new cyclostellettamine from sponge *Amphimedon compressa*. Chinese Chemical Letters 18, 947–950. https://doi.org/10.1016/j.cclet.2007.06.006

136. Laville, R., Thomas, O.P., Berrue, F., Reyes, F., Amade, P., 2008. Pachychalines A–C: novel 3-alkylpyridinium salts from the marine sponge *Pachychalina* sp. European Journal of Organic Chemistry 2008, 121–125. https://doi.org/10.1002/ejoc.200700741
137. Nishi, T., Kubota, T., Fromont, J., Sasaki, T., Kobayashi, J., 2008. Nakinadines B–F: new pyridine alkaloids with a β-amino acid moiety from sponge *Amphimedon* sp. Tetrahedron 64, 3127–3132. https://doi.org/10.1016/j.tet.2008.01.111
138. Timm, C., Volk, C., Sasse, F., Köck, M., 2008. The first cyclic monomeric 3-alkylpyridinium alkaloid from natural sources: identification, synthesis, and biological activity. Organic & Biomolecular Chemistry 6, 4036–4040. https://doi.org/10.1039/B808647H
139. Arai, M., Ishida, S., Setiawan, A., Kobayashi, M., 2009. Haliclonacyclamines, tetracyclic alkylpiperidine alkaloids, as anti-dormant mycobacterial substances from a marine sponge of *Haliclona* sp. Chemical and Pharmaceutical Bulletin 57, 1136–1138. https://doi.org/10.1248/cpb.57.1136
140. Casapullo, A., Pinto, O.C., Marzocco, S., Autore, G., Riccio, R., 2009. 3-Alkylpyridinium alkaloids from the Pacific sponge *Haliclona* sp. Journal of Natural Products 72, 301–303. https://doi.org/10.1021/np800610p
141. Jang, K.H., Kang, G.W., Jeon, J., Lim, C., Lee, H.-S., Sim, C.J., Oh, K.-B., Shin, J., 2009. Haliclonin A, a new macrocyclic diamide from the sponge *Haliclona* sp. Organic Letters 11, 1713–1716. https://doi.org/10.1021/ol900282m
142. Laville, Rémi, Amade, P., Thomas, O.P., 2009. 3-Alkylpyridinium salts from Haplosclerida marine sponges: isolation, structure elucidations, and biosynthetic considerations. Pure and Applied Chemistry 81, 1033–1040. https://doi.org/10.1351/PAC-CON-08-11-14
143. Laville, Remi, Genta-Jouve, G., Urda, C., Fernández, R., Thomas, O.P., Reyes, F., Amade, P., 2009. Njaoaminiums A, B, and C: cyclic 3-alkylpyridinium salts from the marine sponge *Reniera* sp. Molecules 14, 4716–4724. https://doi.org/10.3390/molecules14114716
144. Samoylenko, V., Khan, S.I., Jacob, M.R., Tekwani, B.L., Walker, L.A., Hufford, C.D., Muhammad, I., 2009. Bioactive (+)-manzamine A and (+)-8-hydroxymanzamine A tertiary bases and salts from *Acanthostrongylophora Ingens* and their preparations. Natural Product Communications 4, 1934578X0900400204. https://doi.org/10.1177/1934578X0900400204
145. Schmidt, G., Timm, C., Köck, M., 2009. New haliclamines E and F from the Arctic sponge *Haliclona viscosa*. Organic & Biomolecular Chemistry 7, 3061–3064. https://doi.org/10.1039/B904157E
146. Takahashi, Y., Kubota, T., Fromont, J., Kobayashi, J., 2009. Zamamidines A and B, new manzamine alkaloids from the sponge *Amphimedon* species. Organic Letters 11, 21–24. https://doi.org/10.1021/ol802251q
147. Yamada, M., Takahashi, Y., Kubota, T., Fromont, J., Ishiyama, A., Otoguro, K., Yamada, H., Ōmura, S., Kobayashi, J., 2009. Zamamidine C, 3,4-dihydro-6-hydroxy-10,11-epoxymanzamine A, and 3,4-dihydromanzamine J N-oxide, new manzamine alkaloids from sponge *Amphimedon* sp. Tetrahedron 65, 2313–2317. https://doi.org/10.1016/j.tet.2009.01.032
148. Mudianta, I.W., Katavic, P.L., Lambert, L.K., Hayes, P.Y., Banwell, M.G., Munro, M.H.G., Bernhardt, P.V., Garson, M.J., 2010. Structure and absolute configuration of 3-alkylpiperidine alkaloids from an Indonesian sponge of the genus *Halichondria*. Tetrahedron 66, 2752–2760. https://doi.org/10.1016/j.tet.2010.01.068
149. Wei, X., Nieves, K., Rodríguez, A.D., 2010. Neopetrosiamine A, biologically active bis-piperidine alkaloid from the Caribbean sea sponge *Neopetrosia proxima*. Bioorganic & Medicinal Chemistry Letters 20, 5905–5908. https://doi.org/10.1016/j.bmcl.2010.07.084
150. Kura, K., Kubota, T., Fromont, J., Kobayashi, J., 2011. Pyrinodemins E and F, new 3-alkylpyridine alkaloids from sponge *Amphimedon* sp. Bioorganic & Medicinal Chemistry Letters 21, 267–270. https://doi.org/10.1016/j.bmcl.2010.11.020

151. Li, Y., Qin, S., Guo, Y.-W., Gu, Y.-C., Soest, R.W.M. van, 2011. 9′-Epi-3β,3′β–dimethylxestospongin C, a new macrocyclic diamine alkaloid from the *Hainan* sponge *Neopetrosia exigua*. Planta Medica 77, 179–181. https://doi.org/10.1055/s-0030-1250164
152. Morinaka, B.I., Molinski, T.F., 2011. Xestoproxamines A–C from *Neopetrosia proxima*. Assignment of absolute stereostructure of bis-piperidine alkaloids by integrated degradation-CD analysis. Journal of Natural Products 74, 430–440. https://doi.org/10.1021/np1008637
153. Schmidt, G., Timm, C., Köck, M., 2011. Haliclocyclin C, a new monomeric 3-alkyl pyridinium alkaloid from the Arctic marine sponge *Haliclona viscosa*. Zeitschrift für Naturforschung B 66, 745–748. https://doi.org/10.1515/znb-2011-0717
154. Singh, K.S., Das, B., Naik, C.G., 2011. Quinolizidines alkaloids: petrosin and xestospongins from the sponge *Oceanapia* sp. Journal of Chemical Sciences 123, 601–607. https://doi.org/10.1007/s12039-011-0124-1
155. Cychon, C., Schmidt, G., Mordhorst, T., Köck, M., 2012. Structure elucidation of submilligram auantities of natural products - application to *Haliclamines* G and H from the Arctic marine sponge *Haliclona viscosa*. Zeitschrift für Naturforschung B 67, 944–950. https://doi.org/10.5560/znb.2012-0039
156. Hwang, B.S., Oh, J.S., Jeong, E.J., Sim, C.J., Rho, J.-R., 2012. Densanins A and B, new macrocyclic pyrrole alkaloids isolated from the marine sponge *Haliclona densaspicula*. Organic Letters 14, 6154–6157. https://doi.org/10.1021/ol3028303
157. Ilias, M., Ibrahim, M.A., Khan, S.I., Jacob, M.R., Tekwani, B.L., Walker, L.A., Samoylenko, V., 2012. Pentacyclic ingamine alkaloids, a new antiplasmodial pharmacophore from the marine sponge Petrosid Ng5 Sp5. Planta Medica 78, 1690–1697. https://doi.org/10.1055/s-0032-1315213
158. Lee, Y., Jang, K.H., Jeon, J., Yang, W.-Y., Sim, C.J., Oh, K.-B., Shin, J., 2012. Cyclic bis-1,3-dialkylpyridiniums from the sponge *Haliclona* sp. Marine Drugs 10, 2126–2137. https://doi.org/10.3390/md10092126
159. Wahba, A.E., Fromentin, Y., Zou, Y., Hamann, M.T., 2012. Acantholactone, a new manzamine related alkaloid with an unprecedented δ-lactone and ε-lactam ring system. Tetrahedron Letters 53, 6329–6331. https://doi.org/10.1016/j.tetlet.2012.08.140
160. Damodaran, V., Ryan, J.L., Keyzers, R.A., 2013. Cyclic 3-alkyl pyridinium alkaloid monomers from a New Zealand *Haliclona* sp. Marine Sponge. Journal of Natural Products 76, 1997–2001. https://doi.org/10.1021/np400661c
161. Kubota, T., Kamijyo, Y., Takahashi-Nakaguchi, A., Fromont, J., Gonoi, T., Kobayashi, J., 2013a. Zamamiphidin A, a New manzamine related alkaloid from an Okinawan marine sponge *Amphimedon* sp. Organic Letters 15, 610–612. https://doi.org/10.1021/ol3034274
162. Kubota, T., Kura, K., Fromont, J., Kobayashi, J., 2013b. Pyrinodemins G–I, new bis-3-alkylpyridine alkaloids from a marine sponge *Amphimedon* sp. Tetrahedron 69, 96–100. https://doi.org/10.1016/j.tet.2012.10.062
163. Zhao, Z.-B., Sun, J.-Z., Mao, S.-C., Guo, Y.-W., 2013. Fasciospyrinadine, a novel sesquiterpene pyridine alkaloid from a Guangxi sponge *Fasciospongia* sp. Journal of Asian Natural Products Research 15, 198–202. https://doi.org/10.1080/10286020.2012.751098
164. Dewi, A.S., Hadi, T.A., Fajarningsih, N.D., Blanchfield, J.T., Bernhardt, P.V., Garson, M.J., Dewi, A.S., Hadi, T.A., Fajarningsih, N.D., Blanchfield, J.T., Bernhardt, P.V., Garson, M.J., 2014. Acanthocyclamine A from the Indonesian marine sponge *Acanthostrongylophora ingens*. Australian Journal of Chemistry 67, 1205–1210. https://doi.org/10.1071/CH14107
165. El-Desoky, A.H., Kato, H., Eguchi, K., Kawabata, T., Fujiwara, Y., Losung, F., Mangindaan, R.E.P., de Voogd, N.J., Takeya, M., Yokosawa, H., Tsukamoto, S., 2014. Acantholactam and pre-*neo*-kauluamine, manzamine-related alkaloids from the Indonesian marine sponge *Acanthostrongylophora ingens*. Journal of Natural Products 77, 1536–1540. https://doi.org/10.1021/np500290a
166. Furusato, A., Kato, H., Nehira, T., Eguchi, K., Kawabata, T., Fujiwara, Y., Losung, F., Mangindaan, R.E.P., de Voogd, N.J., Takeya, M., Yokosawa, H., Tsukamoto, S., 2014. Acanthomanzamines A–E with new

manzamine frameworks from the marine sponge *Acanthostrongylophora ingens*. Organic Letters 16, 3888–3891. https://doi.org/10.1021/ol5015569

167. Sun, J.-Z., Jiang, C.-S., Chen, X.-Q., Chen, K.-S., Zhen, X.-C., van Soest, R.W.M., Guo, Y.-W., 2014. Topsendines A–F, new 3-alkylpyridine alkaloids from a Hainan sponge *Topsentia* sp. Tetrahedron 70, 3166–3171. https://doi.org/10.1016/j.tet.2014.03.051
168. AlTarabeen, M., Daletos, G., Ebrahim, W., Müller, W.E.G., Hartmann, R., Lin, W., Proksch, P., 2015. Ircinal E, a new manzamine derivative from the Indonesian marine sponge *Acanthostrongylophora ingens*. Natural Product Communications 10, 1934578X1501001136. https://doi.org/10.1177/1934578X1501001136
169. Arai, M., Kamiya, K., Shin, D., Matsumoto, H., Hisa, T., Setiawan, A., Kotoku, N., Kobayashi, M., 2016. *N*-Methylniphatyne A, a new 3-alkylpyridine alkaloid as an inhibitor of the cancer cells adapted to nutrient starvation, from an Indonesian marine sponge of *Xestospongia* sp. Chemical and Pharmaceutical Bulletin 64, 766–771. https://doi.org/10.1248/cpb.c16-00118
170. Tribalat, M.-A., 2016. Métabolismes spécialisés d'éponges méditerranéennes du genre *Haliclona* Grant, 1836 205.
171. Zhang, H., Loveridge, S.T., Tenney, K., Crews, P., 2016. A new 3-alkylpyridine alkaloid from the marine sponge *Haliclona* sp. and its cytotoxic activity. Natural Product Research 30, 1262–1265. https://doi.org/10.1080/14786419.2015.1054826
172. Einarsdottir, E., Magnusdottir, M., Astarita, G., Köck, M., Ögmundsdottir, H.M., Thorsteinsdottir, M., Rapp, H.T., Omarsdottir, S., Paglia, G., 2017. Metabolic profiling as a screening tool for cytotoxic compounds: identification of 3-alkyl pyridine alkaloids from sponges collected at a shallow water hydrothermal vent site north of Iceland. Marine Drugs 15, 52. https://doi.org/10.3390/md15020052
173. Firsova, D., 2017. Combining biological and chemical approaches in marine biodiscovery (Thesis).
174. Kim, C.-K., Riswanto, R., Won, T.H., Kim, H., Elya, B., Sim, C.J., Oh, D.-C., Oh, K.-B., Shin, J., 2017. Manzamine alkaloids from an *Acanthostrongylophora* sp. sponge. Journal of Natural Products 80, 1575–1583. https://doi.org/10.1021/acs.jnatprod.7b00121
175. Kubota, T., Nakamura, K., Kurimoto, S., Sakai, K., Fromont, J., Gonoi, T., Kobayashi, J., 2017. Zamamidine D, a manzamine alkaloid from an Okinawan *Amphimedon* sp. marine sponge. Journal of Natural Products 80, 1196–1199. https://doi.org/10.1021/acs.jnatprod.6b01110
176. Lyakhova, E.G., Kolesnikova, S.A., Kalinovsky, A.I., Berdyshev, D.V., Pislyagin, E.A., Kuzmich, A.S., Popov, R.S., Dmitrenok, P.S., Makarieva, T.N., Stonik, V.A., 2017. Lissodendoric acids A and B, manzamine-related alkaloids from the Far Eastern sponge *Lissodendoryx florida*. Organic Letters 19, 5320–5323. https://doi.org/10.1021/acs.orglett.7b02608
177. Maarisit, W., Abdjul, D.B., Yamazaki, H., Kato, H., Rotinsulu, H., Wewengkang, D.S., Sumilat, D.A., Kapojos, M.M., Ukai, K., Namikoshi, M., 2017. Anti-mycobacterial alkaloids, cyclic 3-alkyl pyridinium dimers, from the Indonesian marine sponge *Haliclona* sp. Bioorganic & Medicinal Chemistry Letters 27, 3503–3506. https://doi.org/10.1016/j.bmcl.2017.05.067
178. Urda, C., Pérez, M., Rodríguez, J., Fernández, R., Jiménez, C., Cuevas, C., 2018. Njaoamine I, a cytotoxic polycyclic alkaloid from the Haplosclerida sponge *Haliclona* (*Reniera*) sp. Tetrahedron Letters 59, 2577–2580. https://doi.org/10.1016/j.tetlet.2018.05.059
179. Dung, D.T., Hang, D.T.T., Yen, P.H., Quang, T.H., Nhiem, N.X., Tai, B.H., Minh, C.V., Kim, Y.-C., Kim, D.C., Oh, H., Kiem, P.V., 2019. Macrocyclic bis-quinolizidine alkaloids from *Xestospongia muta*. Natural Product Research 33, 400–406. https://doi.org/10.1080/14786419.2018.1455043
180. Esposito, G., Mai, L.H., Longeon, A., Mangoni, A., Durieu, E., Meijer, L., Van Soest, R., Costantino, V., Bourguet-Kondracki, M.-L., 2019. A Collection of bioactive nitrogen-containing molecules from the marine sponge *Acanthostrongylophora ingens*. Marine Drugs 17, 472. https://doi.org/10.3390/md17080472
181. Kato, H., El-Desoky, A.H., Takeishi, Y., Nehira, T., Angkouw, E.D., Mangindaan, R.E.P., de Voogd, N.J., Tsukamoto, S., 2019. Tetradehydrohalicyclamine B, a new proteasome inhibitor from the marine sponge

Acanthostrongylophora ingens. Bioorganic & Medicinal Chemistry Letters 29, 8–10. https://doi.org/10.1016/j.bmcl.2018.11.028

182. Zhang, X., Li, P.-L., Qin, G.-F., Li, S., De Voogd, N.J., Tang, X.-L., Li, G.-Q., 2019. Isolation and absolute configurations of diversiform C17, C21 and C25 terpenoids from the marine sponge *Cacospongia* sp. Marine Drugs 17, 14. https://doi.org/10.3390/md17010014
183. Chen, B., Huan, X.-J., Miao, Z.-H., de Voogd, N.J., Gu, Y.-C., Wang, C.-Y., Guo, Y.-W., Li, X.-W., 2021. Uncommon bis-quinolizidine alkaloids from the Hainan sponge *Neopetrosia chaliniformis*. Chinese Journal of Chemistry 39, 1838–1842. https://doi.org/10.1002/cjoc.202100091
184. Hwang, B.S., Jeong, Y.T., Lee, S., Jeong, E.J., Rho, J.-R., 2021. Densazalin, a new cytotoxic diazatricyclic alkaloid from the marine sponge *Haliclona densaspicula*. Molecules 26, 3164. https://doi.org/10.3390/molecules26113164
185. Patil, A.D., Freyer, A.J., Carte, B., Taylor, P.B., Johnson, R.K., Faulkner, D.J., 2002. Haploscleridamine, a novel tryptamine-derived alkaloid from a sponge of the order Haplosclerida: an inhibitor of cathepsin K. Journal of Natural Products 65, 628–629. https://doi.org/10.1021/np010500l
186. Bugni, T.S., Harper, M.K., McCulloch, M.W.B., Reppart, J., Ireland, C.M., 2008. Fractionated marine invertebrate extract libraries for drug discovery. Molecules 13, 1372–1383. https://doi.org/10.3390/molecules13061372
187. Kubota, T., Kurimoto, S., Kobayashi, J., 2020. Chapter One - The manzamine alkaloids, in: Knölker, H.-J. (Ed.), The Alkaloids: Chemistry and Biology, The Alkaloids. Academic Press, pp. 1–124. https://doi.org/10.1016/bs.alkal.2020.03.001
188. Smet, P.D., Parys, J.B., Callewaert, G., Weidema, A.F., Hill, E., Smedt, H.D., Erneux, C., Sorrentino, V., Missiaen, L., 1999. Xestospongin C is an equally potent inhibitor of the inositol 1,4,5-trisphosphate receptor and the endoplasmic-reticulum Ca^{2+} pumps. Cell Calcium 26, 9–13. https://doi.org/10.1054/ceca.1999.0047
189. Shorthouse, David P. 2010. SimpleMappr, an online tool to produce publication-quality point maps.
190. Cimino, G., Stefano, S.D., Scognamiglio, G., Sodano, G., Trivellone, E., 1986. Sarains: a new class of alkaloids from the marine sponge *Reniera sarai*. Bulletin des Sociétés Chimiques Belges 95, 783–800. https://doi.org/10.1002/bscb.19860950907
191. Baldwin, J.E.; Whitehead, R.C. On the biosynthesis of manzamines. Tetrahedron Letters 1992, 33, 2059–2062, doi:10.1016/0040-4039(92)88141-Q.
192. Andersen, R.J.; Van Soest, R.W.M.; Kong, F. Chapter three 3-alkylpiperidine alkaloids isolated from marine sponges in the order Haplosclerida. Alkaloids: Chemical and Biological Perspectives; Pelletier, S.W., Ed.; Pergamon, 1996; Vol. 10, pp. 301–355, https://doi.org/10.1016/S0735-8210(96)80027-6.
193. Fontana, A., 2006. Biogenetic proposals and biosynthetic studies on secondary metabolites of opisthobranch molluscs, in: Cimino, G., Gavagnin, M. (Eds.), Molluscs: from chemo-ecological study to biotechnological application, Progress in molecular and subcellular biology. Springer, Berlin, Heidelberg, pp. 303–332. https://doi.org/10.1007/978-3-540-30880-5_14
194. Krassowski, M., 2022. ComplexUpset.
195. Tsuda, M., Kawasaki, N., Kobayashi, J., 1994. Keramaphidin C and keramamine C plausible biogenetic precursors of manzamine C from an Okinawan marine sponge. Tetrahedron Letters 35, 4387–4388. https://doi.org/10.1016/S0040-4039(00)73363-6
196. Baker, B.J., Scheuer, P.J., Shoolery, J.N., 1988. Papuamine, an antifungal pentacyclic alkaloid from a marine sponge, *Haliclona* sp. The Journal of the American Chemical Society 110, 965–966. https://doi.org/10.1021/ja00211a046
197. Koren-Goldshlager, G., Kashman, Y., Schleyer, M., 1998. Haliclorensin, a novel diamino alkaloid from the marine sponge *Haliclona tulearensis*. Journal of Natural Products 61, 282–284. https://doi.org/10.1021/np970442x

198. Williams, D.E., Lassota, P., Andersen, R.J., 1998. Motuporamines A–C, cytotoxic alkaloids isolated from the marine sponge *Xestospongia exigua* (Kirkpatrick). The Journal of Organic Chemistry 63, 4838–4841. https://doi.org/10.1021/jo980355p
199. Lin, H., Kwan, A.L., Dutcher, S.K., 2010. Synthesizing and salvaging NAD^+: lessons learned from *Chlamydomonas reinhardtii*. PLOS Genetics 6, e1001105. https://doi.org/10.1371/journal.pgen.1001105
200. Gossmann, T.I., Ziegler, M., Puntervoll, P., de Figueiredo, L.F., Schuster, S., Heiland, I., 2012. NAD^+ biosynthesis and salvage – a phylogenetic perspective. The Federation of European Biochemical Societies Journal 279, 3355–3363. https://doi.org/10.1111/j.1742-4658.2012.08559.x
201. Cutignano, A., Tramice, A., De Caro, S., Villani, G., Cimino, G., Fontana, A., 2003. Biogenesis of 3-alkylpyridine alkaloids in the marine mollusc *Haminoea Orbignyana*. Angewandte Chemie 115, 2737–2740. https://doi.org/10.1002/ange.200250642
202. Cutignano, A., Cimino, G., Giordano, A., d'Ippolito, G., Fontana, A., 2004. Polyketide origin of 3-alkylpyridines in the marine mollusc *Haminoea orbignyana*. Tetrahedron Letters 45, 2627–2629. https://doi.org/10.1016/j.tetlet.2004.01.138
203. Sleeper, H.L., Fenical, W., 1977. Navenones A-C: trail-breaking alarm pheromones from the marine opisthobranch *Navanax inermis*. The Journal of the American Chemical Society 99, 2367–2368. https://doi.org/10.1021/ja00449a072
204. Spinella, A., Alvarez, L.A., Passeggio, A., Cimino, G., 1993. New 3-alkylpyridines from three Mediterranean cephalaspidean molluscs: structure, ecological role and taxonomic relevance. Tetrahedron 49, 1307–1314. https://doi.org/10.1016/S0040-4020(01)85820-8
205. Fukuda, T., Miller, E.D., Clark, B.R., Alnauman, A., Murphy, C.D., Jensen, P.R., Fenical, W., 2011. Structures and biosynthesis of the pyridinopyrones, polyenepyrones from a marine-derived *Streptomyces* species. Journal of Natural Products 74, 1773–1778. https://doi.org/10.1021/np200323e
206. Myronovskyi, M., Rosenkränzer, B., Nadmid, S., Pujic, P., Normand, P., Luzhetskyy, A., 2018. Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters. Metabolic Engineering 49, 316–324. https://doi.org/10.1016/j.ymben.2018.09.004
207. Itoh, T., Tokunaga, K., Matsuda, Y., Fujii, I., Abe, I., Ebizuka, Y., Kushiro, T., 2010. Reconstitution of a fungal meroterpenoid biosynthesis reveals the involvement of a novel family of terpene cyclases. Nature Chemistry 2, 858–864. https://doi.org/10.1038/nchem.764
208. Yaegashi, J., Romsdahl, J., Chiang, Y.-M., Wang, C.C.C., 2015. Genome mining and molecular characterization of the biosynthetic gene cluster of a diterpenic meroterpenoid, 15-deoxyoxalicine B, in *Penicillium canescens*. Chemical Science 6, 6537–6544. https://doi.org/10.1039/C5SC01965F
209. Ma, H.-M., Zhou, Q., Tang, Y.-M., Zhang, Z., Chen, Y.-S., He, H.-Y., Pan, H.-X., Tang, M.-C., Gao, J.-F., Zhao, S.-Y., Igarashi, Y., Tang, G.-L., 2013. Unconventional origin and hybrid system for construction of pyrrolopyrrole moiety in kosinostatin biosynthesis. Chemistry & Biology 20, 796–805. https://doi.org/10.1016/j.chembiol.2013.04.013
210. Hu, Y., Zhou, Q., Zhang, Z., Pan, H.-X., Ilina, Y., Metsä-Ketelä, M., Igarashi, Y., Tang, G.-L., 2021. Deciphering the origin and formation of aminopyrrole moiety in kosinostatin biosynthesis. Chinese Journal of Chemistry 39, 3329–3333. https://doi.org/10.1002/cjoc.202100525
211. Kong, F., Andersen, R.J., Allen, T.M., 1994. Madangamine A, a novel cytotoxic alkaloid from the marine sponge *Xestospongia ingens*. The Journal of the American Chemical Society 116, 6007–6008. https://doi.org/10.1021/ja00092a077
212. Macklin, T.K., Micalizio, G.C., 2010. Convergent and stereospecific synthesis of complex skipped polyenes and polyunsaturated fatty acids. Nature Chemistry 2, 638–643. https://doi.org/10.1038/nchem.665
213. Korman, T.P., Crawford, J.M., Labonte, J.W., Newman, A.G., Wong, J., Townsend, C.A., Tsai, S.-C., 2010. Structure and function of an iterative polyketide synthase thioesterase domain catalyzing Claisen

cyclization in aflatoxin biosynthesis. Proceedings of the National Academy of Sciences 107, 6246–6251. https://doi.org/10.1073/pnas.0913531107

214. Martins, T., Glasser, N., Kountz, D., Oliveira, P., Balskus, E., Leão, P., 2022. Biosynthesis of the unusual carbon skeleton of nocuolin A. https://doi.org/10.26434/chemrxiv-2022-jv5n7
215. Silva, S.B.L., Oberhänsli, F., Tribalat, M.-A., Genta-Jouve, G., Teyssié, J.-L., Dechraoui-Bottein, M.-Y., Gallard, J.-F., Evanno, L., Poupon, E., Thomas, O.P., 2019. Insights into the biosynthesis of cyclic guanidine Alkaloids from Crambeidae marine sponges. Angewandte Chemie International Edition 58, 520–525. https://doi.org/10.1002/anie.201809539
216. R. Lichman, B., 2021. The scaffold-forming steps of plant alkaloid biosynthesis. Natural Product Reports 38, 103–129. https://doi.org/10.1039/D0NP00031K
217. Kajikawa, M., Hirai, N., Hashimoto, T., 2008. A PIP-family protein is required for biosynthesis of tobacco alkaloids. Plant Molecular Biology 69, 287. https://doi.org/10.1007/s11103-008-9424-3
218. DeBoer, K.D., Lye, J.C., Aitken, C.D., Su, A.K.-K., Hamill, J.D., 2008. The A622 gene in *Nicotiana glauca* (tree tobacco): evidence for a functional role in pyridine alkaloid synthesis. Plant Molecular Biology 69, 299. https://doi.org/10.1007/s11103-008-9425-2
219. Kajikawa, M., Shoji, T., Kato, A., Hashimoto, T., 2011. Vacuole-localized berberine bridge enzyme-like proteins are required for a late step of nicotine biosynthesis in Tobacco1. Plant Physiology 155, 2010–2022. https://doi.org/10.1104/pp.110.170878
220. Schachtsiek, J., Stehle, F., 2019. Nicotine-free, nontransgenic tobacco (*Nicotiana tabacum* L.) edited by CRISPR-Cas9. Plant Biotechnology Journal 17, 2228–2230. https://doi.org/10.1111/pbi.13193
221. Baldwin, J.E., Claridge, T.D.W., Culshaw, A.J., Heupel, F.A., Smrcková, S., Whitehead, R.C., 1996. A biomimetic approach to the manzamine alkaloids. Tetrahedron Letters 37, 6919–6922. https://doi.org/10.1016/0040-4039(96)01516-X
222. Baldwin, J.E., Spring, D.R., Atkinson, C.E., Lee, V., 1998. Efficient synthesis of the sponge alkaloids cyclostellettamines A-F. Tetrahedron 54, 13655–13680. https://doi.org/10.1016/S0040-4020(98)00842-4
223. Nakamura, H., Schultz, E.E., Balskus, E.P., 2017. A new strategy for aromatic ring alkylation in cylindrocyclophane biosynthesis. Nature Chemical Biology 13, 916–921. https://doi.org/10.1038/nchembio.2421
224. Steffen, K., Laborde, Q., Gunasekera, S., Payne, C.D., Rosengren, K.J., Riesgo, A., Göransson, U., Cárdenas, P., 2021. Barrettides: a peptide family specifically produced by the deep-sea sponge *Geodia barretti*. Journal of Natural Products 84, 3138–3146. https://doi.org/10.1021/acs.jnatprod.1c00938
225. Wilson, M.C., Mori, T., Rückert, C., Uria, A.R., Helf, M.J., Takada, K., Gernert, C., Steffens, U.A.E., Heycke, N., Schmitt, S., Rinke, C., Helfrich, E.J.N., Brachmann, A.O., Gurgui, C., Wakimoto, T., Kracht, M., Crüsemann, M., Hentschel, U., Abe, I., Matsunaga, S., Kalinowski, J., Takeyama, H., Piel, J., 2014. An environmental bacterial taxon with a large and distinct metabolic repertoire. Nature 506, 58–62. https://doi.org/10.1038/nature12959
226. Becking, L.E., Nakao, Y., de Voogd, N.J., van Soest, R.W., Fusetani, N. and Matsunaga, S., 2007. Perplexing distribution of 3-alkylpyridines in haplosclerid sponges. Porifera Research: Biodiversity, Innovation and Sustainability Série Livros. Museu Nacional: Rio de Janeiro, 173-178.
227. Redmond, N.E., Raleigh, J., Soest, R.W.M. van, Kelly, M., Travers, S.A.A., Bradshaw, B., Vartia, S., Stephens, K.M., McCormack, G.P., 2011. Phylogenetic relationships of the marine Haplosclerida (phylum Porifera) employing ribosomal (28S rRNA) and mitochondrial (cox1, nad1) gene sequence data. PLOS ONE 6, e24344. https://doi.org/10.1371/journal.pone.0024344
228. Tianero, M.D., Balaich, J.N., Donia, M.S., 2019. Localized production of defence chemicals by intracellular symbionts of *Haliclona* sponges. Nature Microbiology 4, 1149–1159. https://doi.org/10.1038/s41564-019-0415-8

229. Garson, M.J., Flowers, A.E., Webb, R.I., Charan, R.D., McCaffrey, E.J., 1998. A sponge/dinoflagellate association in the haplosclerid sponge *Haliclona* sp.: cellular origin of cytotoxic alkaloids by Percoll density gradient fractionation. Cell Tissue Res 293, 365–373. https://doi.org/10.1007/s004410051128
230. Fieseler, L., Hentschel, U., Grozdanov, L., Schirmer, A., Wen, G., Platzer, M., Hrvatin, S., Butzke, D., Zimmermann, K., Piel, J., 2007. Widespread occurrence and genomic context of unusually small polyketide synthase genes in microbial consortia associated with marine sponges. Applied and Environmental Microbiology 73, 2144–2155. https://doi.org/10.1128/AEM.02260-06
231. Hochmuth, T., Niederkrüger, H., Gernert, C., Siegl, A., Taudien, S., Platzer, M., Crews, P., Hentschel, U., Piel, J., 2010. Linking chemical and microbial diversity in marine sponges: possible role for poribacteria as producers of methyl-branched fatty acids. ChemBioChem 11, 2572–2578. https://doi.org/10.1002/cbic.201000510
232. Trindade-Silva, A.E., Rua, C.P.J., Andrade, B.G.N., Vicente, A.C.P., Silva, G.G.Z., Berlinck, R.G.S., Thompson, F.L., 2013. Polyketide synthase gene diversity within the microbiome of the sponge *Arenosclera brasiliensis*, endemic to the southern Atlantic Ocean. Applied and Environmental Microbiology 79, 1598–1605. https://doi.org/10.1128/AEM.03354-12
233. Waters, A.L., Peraud, O., Kasanah, N., Sims, J.W., Kothalawala, N., Anderson, M.A., Abbas, S.H., Rao, K.V., Jupally, V.R., Kelly, M., Dass, A., Hill, R.T., Hamann, M.T., 2014. An analysis of the sponge *Acanthostrongylophora igens*' microbiome yields an actinomycete that produces the natural product manzamine A. Frontiers in Marine Science 1. https://doi.org/10.3389/fmars.2014.00054
234. Tucker, S.J., McClelland, D., Jaspars, M., Sepčić, K., MacEwan, D.J., Scott, R.H., 2003. The influence of alkyl pyridinium sponge toxins on membrane properties, cytotoxicity, transfection and protein expression in mammalian cells. Biochimica et Biophysica Acta (BBA) - Biomembranes 1614, 171–181. https://doi.org/10.1016/S0005-2736(03)00175-5
235. Kelman, D., Kashman, Y., Rosenberg, E., Ilan, M., Ifrach, I., Loya, Y., 2001. Antimicrobial activity of the reef sponge *Amphimedon viridis* from the Red Sea: evidence for selective toxicity. Aquatic Microbial Ecology 24, 9–16. https://doi.org/10.3354/ame024009
236. Ilan, M., Gugel, J., Van Soest, R., 2004. Taxonomy, reproduction and ecology of new and known Red Sea sponges. Sarsia 89, 388–410. https://doi.org/10.1080/00364820410002659
237. Marra, M.V., 2019. Investigation of biological factorsethat may contribute to bioactivity in *Haliclona* (Porifera, Haplosclerida).
238. Köck, M., Muñoz, J., Cychon, C., Timm, C., Schmidt, G., 2013. The Arctic sponge *Haliclona viscosa* as a source of a wide array of 3-alkyl pyridine alkaloids. Phytochemistry Reviews 12, 391–406. https://doi.org/10.1007/s11101-012-9249-1
239. Findlay, J.A., Patil, A.D., 1985. Novel sterols from the finger sponge *Haliclona oculata*. Canadian Journal of Chemistry 63, 2406–2410. https://doi.org/10.1139/v85-398
240. Gardères, J., Bourguet-Kondracki, M.-L., Hamer, B., Batel, R., Schröder, H.C., Müller, W.E.G., 2015. Porifera lectins: diversity, physiological roles and biotechnological potential. Marine Drugs 13, 5059–5101. https://doi.org/10.3390/md13085059
241. Wiens, M., Korzhev, M., Krasko, A., Thakur, N.L., Perović-Ottstadt, S., Breter, H.J., Ushijima, H., Diehl-Seifert, B., Müller, I.M., Müller, W.E.G., 2005. Innate immune defense of the sponge *Suberites domuncula* against bacteria involves a MyD88-dependent signaling pathway: induction of a perforin-like molecule. Journal of Biological Chemistry 280, 27949–27959. https://doi.org/10.1074/jbc.M504049200
242. Müller, W.E.G., Wang, X., Binder, M., Lintig, J. von, Wiens, M., Schröder, H.C., 2012. Differential expression of the demosponge (*Suberites domuncula*) carotenoid oxygenases in response to light: protection mechanism against the self-produced toxic protein (suberitine). Marine Drugs 10, 177–199. https://doi.org/10.3390/md10010177
243. Scarfì, S., Pozzolini, M., Oliveri, C., Mirata, S., Salis, A., Damonte, G., Fenoglio, D., Altosole, T., Ilan, M., Bertolino, M., Giovine, M., 2020. Identification, purification and molecular characterization of chondrosin,

a new protein with anti-tumoral activity from the marine sponge *Chondrosia Reniformis* Nardo 1847. Marine Drugs 18, 409. https://doi.org/10.3390/md18080409

244. Nichols, S.A., Roberts, B.W., Richter, D.J., Fairclough, S.R., King, N., 2012. Origin of metazoan cadherin diversity and the antiquity of the classical cadherin/β-catenin complex. Proceedings of the National Academy of Sciences 109, 13046–13051. https://doi.org/10.1073/pnas.1120685109
245. Ereskovsky, A.V., Richter, D.J., Lavrov, D.V., Schippers, K.J., Nichols, S.A., 2017. Transcriptome sequencing and delimitation of sympatric *Oscarella* species (*O. carmela* and *O. pearsei* sp. nov) from California, USA. PLOS ONE 12, e0183002. https://doi.org/10.1371/journal.pone.0183002
246. Ben-Ari, H., Paz, M., Sher, D., 2018. The chemical armament of reef-building corals: inter- and intra-specific variation and the identification of an unusual actinoporin in *Stylophora pistilata*. Scientific Reports 8, 251. https://doi.org/10.1038/s41598-017-18355-1
247. Kem, W.R., 1988. Sea anemone toxins: structure and action. The Biology of Nematocysts 375–405.
248. Kawashima, Y., Nagai, H., Ishida, M., Nagashima, Y., Shiomi, K., 2003. Primary structure of echotoxin 2, an actinoporin-like hemolytic toxin from the salivary gland of the marine gastropod Monoplex echo. Toxicon 42, 491–497. https://doi.org/10.1016/S0041-0101(03)00226-5
249. Gutiérrez-Aguirre, I., Trontelj, P., Maček, P., Lakey, J.H., Anderluh, G., 2006. Membrane binding of zebrafish actinoporin-like protein: AF domains, a novel superfamily of cell membrane binding domains. Biochemical Journal 398, 381–392. https://doi.org/10.1042/BJ20060206
250. Glasser, E., Rachamim, T., Aharonovich, D., Sher, D., 2014. Hydra actinoporin-like toxin-1, an unusual hemolysin from the nematocyst venom of *Hydra magnipapillata* which belongs to an extended gene family. Toxicon, special issue: Freshwater and marine toxins 91, 103–113. https://doi.org/10.1016/j.toxicon.2014.04.004
251. Ramírez-Carreto, S., Miranda-Zaragoza, B., Rodríguez-Almazán, C., 2020. Actinoporins: from the structure and function to the generation of biotechnological and therapeutic tools. Biomolecules 10, 539. https://doi.org/10.3390/biom10040539

Chapter 2

Transcriptomics and Metabolic Pathway Mapping of *Haliclona* Species

Part of this chapter contributed to the peer-reviewed publication:
Sandoval, K., McCormack, G.P., 2022. Actinoporin-like Proteins Are Widely Distributed in the Phylum Porifera. Marine Drugs 20, 74.

Introduction

A transcriptome constitutes a sequenced set of RNA transcripts derived from one or more specimens [1]. Because gene expression is influenced by a number of factors, such as environmental conditions and the type of tissue sampled, transcriptomes represent a snapshot of coding and non-coding RNA which was being expressed at the time of sampling. This means that, unlike a complete genome, a transcriptome does not represent a comprehensive picture of the total coding genes in an organism. However, for the purpose of studying protein-coding genes transcriptomics has several advantages over genome sequencing which make it a valuable tool for studying the protein-coding genes of an organism. For one, because the sequencing of non-coding nucleotides can be avoided the cost efficiency of sequencing a library prepared from mRNA is higher. This is especially so when massively parallel, high-throughput technologies such as short-read Illumina sequencing are utilized for RNA-seq [2]. Furthermore, library preparation from polyadenylated mRNA allows for the selective sequencing and analysis of protein-coding genes from a eukaryotic host as opposed to any associated bacteria [3]. As a result, transcriptomics can be an affordable alternative to genomics when the project is focused on identifying and characterizing what protein-coding genes an organism possesses, especially in the case of eukaryotes with complex microbiomes. Innovations in sequencing technology, particularly third generation long-read sequencing, have further pushed the capabilities of RNA-seq by improving upon read length allowing for better detection of variant isoforms [4]. Due to its affordability, transcriptomics has been applied to a wide range of non-model organisms from which reference quality genomes may be too difficult to achieve. Sponges are no exception, and a wealth of transcriptomic information has been generated for numerous species [5]. Such information has been used to gain insight into topics such as animal complexity [6], symbiosis with microorganisms [7], early diversification of animal lineages [8], response to pathogenic microorganisms [9], response to climate change [10], and species delimitation [11].

Furthermore, transcriptomics has been employed to understand specialized metabolism in different organisms. For example, the sequencing of long mRNA molecules with Pacific Biosciences technology has allowed the identification of isoforms alternate to those associated with the production of specialized metabolites [12]. In addition, differential expression of genes associated with specialized metabolism is a powerful approach for characterizing the biosynthetic origin of compounds not readily detected by genome mining algorithms [13]. Gene annotation and mapping to known metabolic pathways associated with specialized metabolism represents another method of gaining insight into whether the organism of study possesses the necessary genes to produce a compound of interest. Numerous databases of metabolic pathways exist such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) or MetaCyc [14,15]. KEGG contains pathways associated with primary and specialized metabolites such as the biosynthesis of alkaloids, terpenoids, steroids,

polyketides and fatty acids. This makes it a powerful tool to predict and visualize the completeness of an organism's metabolic pathways. For example, KEGG has been utilized to determine whether the sponge *Amphimedon queenslandica* possesses complete pathways for the biosynthesis of sterols and fatty acids [16].

The 3-alkylpyridine alkaloids (3-APs) are found throughout sponges of the Order Haplosclerida and their biosynthetic origin has largely remained a mystery [17]. This includes questions such as which enzymes are responsible and whether the host sponge or microbial symbionts encode the genes for said enzymes. The two major moieties of the 3-APs, the pyridine ring and the aliphatic chain, provide clues as to how these molecules may be synthesized. In particular, the pyridine ring has been hypothesized to originate from nicotinic acid (NTA) (Figure 1.11B) [18-20]. NTA is intrinsically linked to the cycle of the cofactor nicotinamide adenine dinucleotide (NAD^+) and as a result there are several ways it could be synthesized in the haplosclerid holobiome (Figure 2.1) [21]. First, the precursor quinolinic acid (QA) can be synthesized *de novo* in various organisms. Animals and fungi are known to utilize tryptophan, plants utilize aspartate, and bacteria can utilize both. QA is then converted to nicotinic acid mononucleotide (NTAM), nicotinic acid adenine dinucleotide ($NTAAD^+$) and then NAD^+ in what is known as the NAD^+ cycle. After being utilized, it is then converted to nicotinamide (NAM). From there, the two-step or four-step salvage pathways can recycle this product back into the NAD^+ cycle. Therefore, the *de novo* biosynthesis of NTA likely requires either a complete tryptophan or aspartate pathway, a complete NAD^+ cycle pathway, and then a four-step salvage pathway. Alternatively, NTA can also be acquired *via* the diet of an organism [22]. When examining the four-step pathway, there are two critical enzymes necessary for the recycling of NTA. The first is nicotinamide amidohydrolase which is responsible for the conversion of NAM to NTA [23]. The second is nicotinate phosphoribosyltransferase which then converts NTA to NTAM, but a reversible reaction from this enzyme which results in the production of NTA has been reported [24]. In addition, there exists evidence of several enzymes such as 5' nucleotidase being involved in the conversion of NTAM to NTA [25,26].

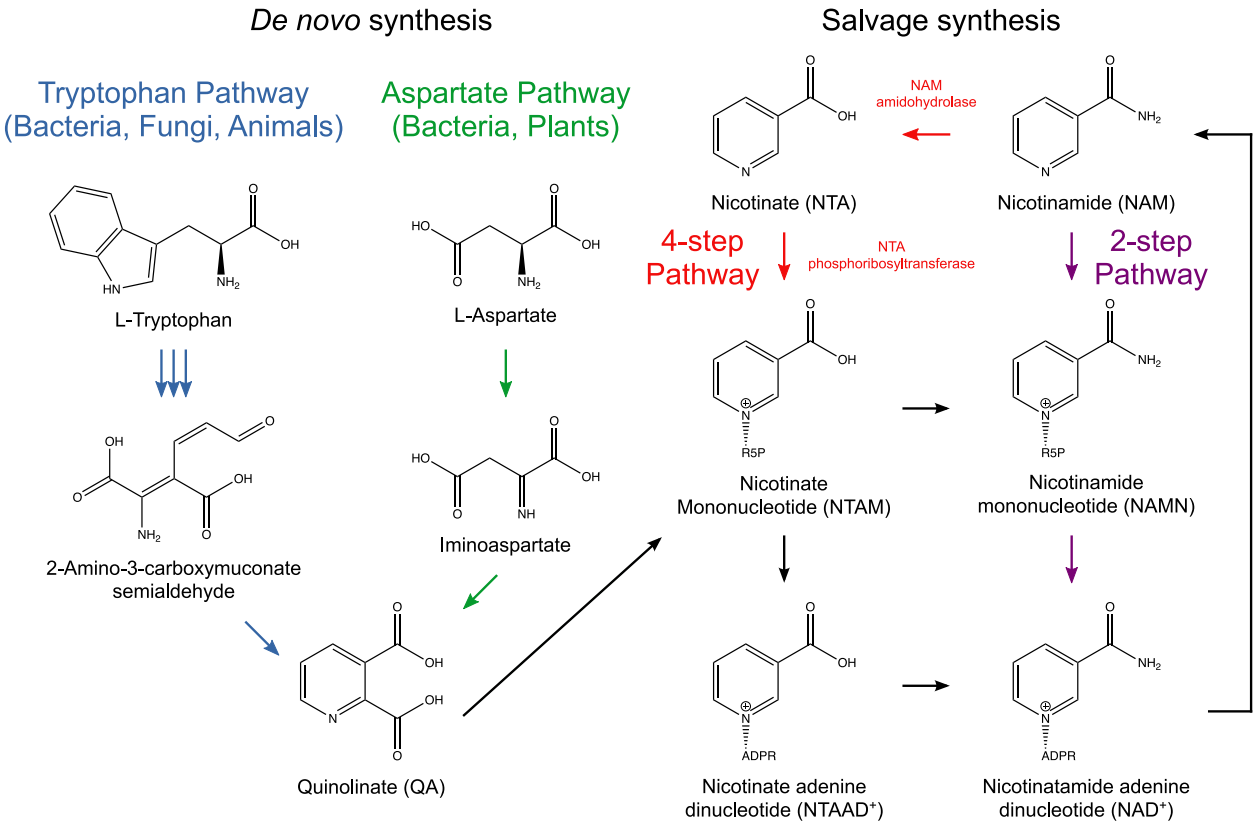


Figure 2.1 The NAD+ Cycle. Modified from Lin et al. 2010 [21]. Moiety abbreviations are as follows: R5P, ribose 5-phosphate; ADPR, adenosine diphosphate ribose.

The aliphatic chain of 3-APs could be derived from the condensation of malonyl-CoA or its derivatives [18-20]. Such a condensation is typically performed *via* the enzymes fatty acid synthase (FAS) associated with primary metabolism or polyketide synthase (PKS) associated with specialized metabolism. In particular, the highly reduced nature of 3-AP aliphatic chains supports the notion that this portion may be derived from a FAS or a highly reducing PKS. Two types of FAS exist. The type I system comprises a single enzyme containing all the necessary domains to synthesize fatty acids and is utilized by animals, fungi and some bacteria [27-29]. Alternatively, the type II system is found in most bacteria and plants and comprises distinct, individual enzymes each responsible for the different reactions necessary to synthesize fatty acids [30]. Both systems employ the same steps of initiation, elongation, and termination to reach the fully reduced product palmitic acid. Type I and II PKS are also divided in a similar manner, but often contain more exotic enzymatic domains allowing for more complex modifications and polyketide end products as compared to fatty acids. As a result, unlike fatty acids, polyketides often vary in methylation, present functional groups, and degrees of saturation [31]. Furthermore, Type III PKS also exist in which the system is independent of an acyl carrier protein. The polyketide hypothesis for 3-AP biosynthesis is typically paired with the idea of NTA being used as a starter unit. Genetic evidence of this reaction occurring, albeit with compounds different from 3-APs, has been

identified in three different microbial PKS biosynthetic gene clusters in which an AMP-binding enzyme ligates NTA to coenzyme A allowing it to be utilized by the PKS [32-34]. Furthermore, there exists a standalone adenylation domain involved in the biosynthesis of the polyketide kosinostatin which can promiscuously activate NTA and represents an alternative, hypothetical method to incorporate this substrate into a polyketide [35].

Plant alkaloids also offer an alternative source of inspiration for how the 3-APs could be synthesized. Rather than NTA being incorporated into polyketide biosynthesis, it may be that it is instead condensed with an already complete aliphatic chain. No publication indicates that such an enzyme exists to perform this reaction. However, the alkaloid nicotine does represent an instance in which an NTA derivative condensed with an *N*-methylpyrollinium at the same position that the aliphatic chain is located on a pyridine ring in 3-APs. Specifically, the plant *Nicotiana tabacum* appears to utilize two distinct enzymes to accomplish this process [36]. The first is an isoflavone reductase homolog and the second a berberine bridge enzyme-like (BBE-like) protein [37,38]. Based on their role in condensation, similar enzymes to these from *N. tabacum* could represent an alternate way in which a 3-AP monomer is created *via* the sponge holobiome. Furthermore 3-AP alkaloids are typically oligomerized resulting in the creation of 3-alkylpyridinium salts [39]. This process involves the bonding between the aliphatic chain terminal end of one monomer with the nitrogen atom of another. The pyridine alkaloid trigonelline, found in a wide variety of plants such as *Arabidopsis thaliana*, is produced by an *N*-methyltransferase [40]. Trigonelline has been found in several species of non-haplosclerid sponge and it may be that a similar enzyme could be involved in the oligomerization of monomeric units [41-44].

Because of the specific association of 3-APs with haplosclerid sponges it is often hypothesized that the host organism may be the true source of these compounds rather than its associated microbial consortia. Due to the nature of poly(A) selection, which enriches for eukaryotic mRNA, transcriptome sequencing with libraries prepared from this method represents a targeted insight into the metabolic capabilities of the host organism. Based on this reasoning, transcriptome sequencing and assembly of Irish *Haliclona* species was employed to assess the possession of genes hypothesized to be involved in the biosynthesis of 3-APs. A specific approach was the use of metabolic pathway mapping to determine the capability of these sponges to produce the hypothesized building blocks of 3-APs, namely NTA and an alkyl chain. Furthermore, the possession of metabolic pathways responsible to produce other bioactive compounds commonly associated with sponges, such as sterols, was a secondary, more general aim of this pursuit.

Materials and Methods

Sample Collection

Haliclona indistincta (MIIG1388) (Figure 2.2A) was collected at Corranroo on 17/05/2019 and *Haliclona viscosa* (MIIG1389 and MIIG1390) (Figure 2.2B) was collected at Bridges of Ross on 01/08/2019. Visible epibionts were removed. The sponges were rinsed in sterile artificial seawater. The sponges were then dissected into ~1 cm^3 pieces and flash-frozen with liquid nitrogen. Samples were stored at -70 °C until further use. Voucher specimens were stored in absolute ethanol.

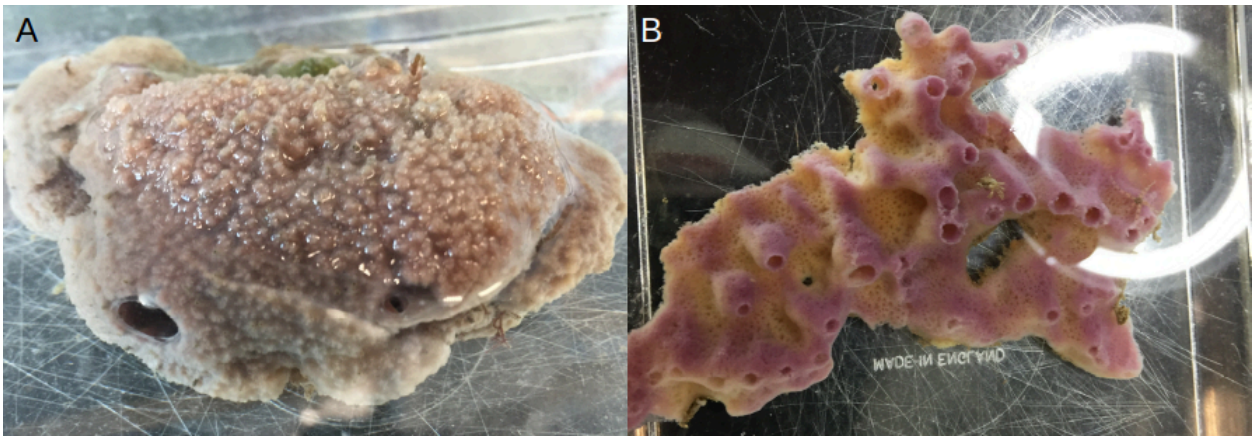


Figure 2.2 (**A**) MIIG1388 prior to dissection and flash freezing. (**B**) MIIG1389 prior to dissection and flash-freezing.

RNA Extraction

A ~1 cm^3 piece of flash-frozen sponge tissue was submerged in 500 µL of Trizol in a 2 mL microcentrifuge tube. The tissue was semi-homogenized by hand with a pestle. An additional 500 µL of Trizol was then added to the sample. The sample was mixed by gently inverting five times and allowed to incubate at room temperature for 5 min. The sample was then inverted and vortexed with a VWR Analogue mini vortex mixer at maximum speed for 2 min.

A volume of 100 µL 1–bromo–3–chloropropane was added to the sample. The sample was mixed by hand for 20 sec and then vortexed for 10 sec with a VWR Analogue mini vortex mixer at maximum speed. The sample was incubated at room temperature for 5-10 min. The sample was then centrifuged at 16,000 × g for 15 min at 6 °C. The clear, aqueous layer at the top was transferred to a fresh microcentrifuge tube.

RNA was purified by adding 500 µL of 100% isopropanol to the aqueous phase sample. The sample was inverted and vortexed with a VWR Analogue mini vortex mixer at maximum speed for 2 min. The sample was left to incubate for 10 min at room temperature. The sample was centrifuged at 16,000 × g for 15 min at 4 °C. The supernatant was discarded and 1 mL of 75% ethanol was added to the RNA pellet. The pellet was disrupted by vortexing. The RNA

sample was then centrifuged for 5 min at 4 °C at 7,500 × g. The ethanol was carefully removed without disturbing the pellet. The washing with 75% ethanol was repeated once. The RNA sample was allowed to air dry until the edges of the pellet were visible. Finally, the pellet was resuspended in 100 µL molecular-grade water. The RNA sample was kept frozen at -70 °C until further use. A subsample of each RNA extraction was used for quality and quantity assessment on a 2100 Bioanalyzer RNA Eukaryotic Chip.

Transcriptome Sequencing

Samples were sent to Macrogen, Inc. for the preparation of Illumina TruSeq Stranded mRNA libraries from poly(A) selection with insert sizes of 150 bp. The libraries were sequenced on a Novaseq 6000 with a targeted 40 million reads per sample. Raw RNA-seq reads are available at the NCBI BioProjeect PRJNA795170.

Transcriptome Assembly

Previously sequenced raw cDNA Illumina reads of *Haliclona cinerea*, *H. indistincta* (MIIG1093, MIIG1094 and MIIG1095), *Haliclona oculata* (MIIG1250 and MIIG1251) and *Haliclona simulans* (MIIG1248 and MIIG1249) were acquired from Dr Grace P. McCormack and Dr Jose Maria Aguilar-Camacho for use in this study (personal communication).

Raw cDNA reads of *H. cinerea*, *H. indistincta*, *H. oculata*, *H. simulans* and *H. viscosa* were processed with fastp version 0.2 on default settings to remove adapters and low-quality regions [45]. The processed reads were then assembled with Trinity version 2.8.5 on default settings [46]. Reads were pooled so that one transcriptome per species was assembled. Isoforms and lowly-expressed transcripts were retained in the final assembly. The longest translated open reading frames per transcript were extracted using TransDecoder version 5.5.0 [47]. Homology searches using these open reading frames as a query against the SwissProt database (accessed 14/01/2020) with BLASTp version 2.9.0 [48,49] as well as the Pfam database (accessed 14/01/2020) with HMMER version 3.2.1 [50,51] were performed. Significant homologous alignments were used to guide TransDecoder in identifying additional open reading frames. The completeness of the transcriptomes was then assessed using BUSCO version 5.1.2; specifically, the assemblies were queried against the latest version of the eukaryota_odb10 dataset which comprises near universal single-copy orthologs expected in all members of the kingdom (downloaded 13/04/2021) [52]. Transcriptome assembly, open reading frame extraction and BUSCO analysis were performed with an account at the Leibniz Supercomputing Centre.

Screening for Enzymes Hypothesized to be Involved in 3-AP Biosynthesis

Several enzymes hypothesized to be involved in 3-AP biosynthesis, as described previously (Chapter 1), were used as tblastn queries against the assembled transcriptomes of *H. indistincta* and *H. viscosa* which are known to produce 3-APs. If significant hits were identified (E-value less than 1e-4), homologs were searched for in the transcriptomes of *H.*

cinerea, *H. oculata*, and *H. simulans* as these species are not known to produce 3-APs and should thus lack similar enzymes. These queries include the AMP-binding CoA ligases involved in preparing NTA for the polyketide biosynthesis of pyripyropene A, 5-deoxyoxalicine and pyridinopyrone A (XP_751267.1; no accession; AYD88521.1) [32,33,34]; the standalone adenylation domain involved in the biosynthesis of kosinostatin which can promiscuously activate NTA (AFJ52660.1) [35]; the nicotinate *N*-methyltransferase of *A. thaliana* involved in the biosynthesis of trigonelline (NP_190882.1) [40]; two isoflavone reductase homologs of A622 from *N. tabacum* involved in the biosynthesis of nicotine (NP_001312315.1; NP_001312742.1); and four individual BBE-like proteins from *N. tabacum* also involved in the biosynthesis of nicotine (NP_001312453.1, NP_001313171.1; NP_001313194.1; BAK18781.1). Specifically, the regions corresponding to the conserved domain Pfam00501, which represents AMP-binding domains, were identified and extracted from the aforementioned CoA ligases and adenylation domain using the NCBI conserved domain database [53]. Significant hits where then screened against the NCBI non-redundant database to assess whether they have a higher alignment with different proteins not of relevance [54].

Metabolic Pathway Mapping

The translated open reading frames of each assembly were uploaded to the KEGG blastKOALA service to annotate them with KEGG Orthology numbers [55]. Open reading frames were then mapped to metabolic pathways in the KEGG database [14]. Select pathways possibly linked to the biosynthesis of 3-APs were analysed to see whether *H. indistincta* and *H. viscosa* possessed the necessary enzymes to synthesize building blocks of these alkaloids. Furthermore, the ability of the *Halicona* sp. to synthesize other types of specialized metabolites was assessed by identifying pathways with a high degree of completeness within the comprehensive KEGG reference pathway for the biosynthesis of specialized metabolites. It should be noted that numerous KEGG Orthology numbers can correspond to different enzymes in different pathways. In particular, the pathways for animal and plant sterol biosynthesis are affected by this quality. In this case the enzymes identified as belonging to both these pathways were manually examined by running them as blastp queries against the nr database to estimate which pathway they belonged to. If key enzymes necessary for a complete metabolic pathway were not detected by KEGG, a query sequence from the UniProt database was used to manually screen the nucleotide sequences of the transcriptomes (Appendix 1) [56]. When possible, proteins from the sponge *A. queenslandica* were used as queries. If not, then proteins from model organisms were used.

Results

Transcriptome Assembly

After processing with fastp, approximately 472.68, 69.80, 257.97, 207.46, and 94.64 Mbp (base pair values rounded up to two decimal places here and for every instance hereafter) of

data were acquired for *H. cinerea*, *H. indistincta*, *H. oculata*, *H. simulans* and *H. viscosa* respectively. Five separate transcriptomes were then successfully assembled using the Trinity software (Table 2.1). More data were available for *H. cinerea*, *H. oculata* and *H. simulans* which appeared to be reflected in the larger assembly size and higher number of true genes when compared to *H. indistincta* and *H. viscosa*. Furthermore, this division was also apparent regarding GC content in which the first three exhibited a value around 39% while the latter were at 44.5%. The total amount of translated open reading frames was reflected by the size of the assemblies with *H. cinerea* yielding the most protein sequences while *H. viscosa* yielded the least (Table 2.2). All five *Haliclona* transcriptomes exhibited very high completeness when assessed with the BUSCO eukaryote data set (Table 2.3).

Table 2.1 Trinity assembly statistics of the five *Haliclona* transcriptomes.

Species	**Total (Mbp)**	**Number of contigs**	**Number of Trinity 'genes' excluding isoforms**	**Contig N50 (kbp)**	**GC (%)**
H. cinerea	156.12	123,111	64,261	2.81	39.59
H. indistincta	106.87	123,111	48,788	2.03	44.1
H. oculata	142.18	122,855	70,008	2.46	38.56
H. simulans	104.12	106,366	55,501	1.89	39.87
H. viscosa	94.09	105,831	59,949	1.73	44.5

Table 2.2 Transdecoder open reading frame statistics of the five *Haliclona* transcriptomes.

Species	**Total (aa)**	**Complete ORF**	**5' Partial ORF**	**3' Partial ORF**
H. cinerea	34,243,765	58,606	14,476	6,129
H. indistincta	26,553,534	29,232	15,794	7,290
H. oculata	30,530,592	49,528	16,191	5,687
H. simulans	24,643,377	31,081	16,112	7,561
H. viscosa	22,299,218	24,435	12,776	7,166

Table 2.3 BUSCO Eukaryotic score for the five *Haliclona* transcriptomes.

BUSCO	***H. cinerea***	***H. indistincta***	***H. oculata***	***H. simulans***	***H. viscosa***
Complete (%)	97.6	99.6	98.5	97.6	95.3
Single (%)	23.9	45.1	31.4	38.0	55.3
Duplicate (%)	73.7	54.5	67.1	59.6	40
Fragmented (%)	0.8	0.0	0.4	2.0	4.3
Missing (%)	1.6	0.4	1.1	0.4	0.4

Hypothesized Enzymes associated with 3-AP Biosynthesis

Only a single transcript in the *H. indistincta* assembly significantly aligned (bit score 45.4) with the second isoflavone reductase homolog A622-like protein from *N. tabacum.* Sequentially this protein most aligned with uncharacterized proteins from animals in the non-redundant database, but the conserved domains it possessed were not consistent with those of A622. All five *Haliclona* species were found to possess transcripts encoding proteins with both FAD binding 4 and BBE-like domains. All five appeared to share homologous genes which most aligned with those from cnidarians in the non-redundant database; these cnidarian sequences also encoded for both the FAD binding 4 and BBE-like domains. The closest named protein was a spectrin alpha chain (PFX14964.1) from the stony coral *Stylophora pistillata* with percent identities of 35.16%, 66.96, 47.15%, 47.22%, and 64.09% for the first isoforms of these proteins from *H. cinerea*, *H. indistincta*, *H. oculata*, *H. simulans*, and *H. viscosa* respectively. However, *H. indistincta* and *H. viscosa* possessed additional transcripts encoding for these two domains which instead most aligned with FAD-binding oxidoreductases from bacteria of the genus *Streptomyces*; these bacterial proteins also contained the BBE-like domain. Particularly, that of *H. indistincta* most aligned with a sequence from *Streptomyces chartreusis* (WP_150501450.1) with a percent identity of 28.54% whereas that of *H. viscosa* was with a sequence from *Streptomyces tubercidicus* (WP_159743968.1) with a percent identity of 30.47%. Significant local alignments were detected in all five *Haliclona* species when using the AMP-binding domains which act upon NTA as queries. When screening the top corresponding *Haliclona* sequences against the non-redundant database no top matches towards AMP-binding enzymes which act upon NTA were detected. Significant matches towards a predicted 4-coumarate—CoA ligase from *A. queenslandica* (XP_003386175.1) and a luciferase polypeptide from the sponge *Suberites domuncula* (CAR31336.1) were instead the consistent op hits with percent identities higher than 50%. The exception was a sequence from *H. oculata* which most aligned with a predicted acyl-CoA synthetase family member 4-like protein from *A. queenslandica* (XP_019859051.1) and a beta-alanine-activating enzyme from the fish *Megalops cyprinoides* (XP_036376699.1). None of the five *Haliclona* species possessed nicotinate *N*-methyltransferase to convert NTA to trigonelline, or N1-methylnicotinic acid (N1-MNTA), as possibly expected for 3-alkylpyridinium salts.

KEGG Orthology Functional Annotation

A total of 12,945, 11,161, 13,228, 10,413 and 12,600 extracted open reading frames from *H. cinerea*, *H. indistincta*, *H. oculata*, *H. simulans* and *H. viscosa* were respectively assigned KEGG Orthology numbers. This accounts for 23.29%, 26.63%, 25.31%, 23.21% and 26.35% of the total extracted open reading frames respectively for each of the aforementioned species. Regarding the different classifications of KO numbers, each of the five species generally exhibited similar proportions (Figure 2.3). The most significant exception is that *H. simulans* exhibited far fewer numbers of assigned KO identifiers associated with metabolism. Furthermore, *H. indistincta and H. simulans* also exhibited fewer assigned KO identifiers associated with human disease than the other three species. When focusing on different pathways which fall under the metabolism classification, the five species again tended to exhibit similar numbers for each category (Figure 2.4). The primary exception is again with *H. simulans* which received few amounts of KO identifiers associated with energy metabolism and nucleotide metabolism.

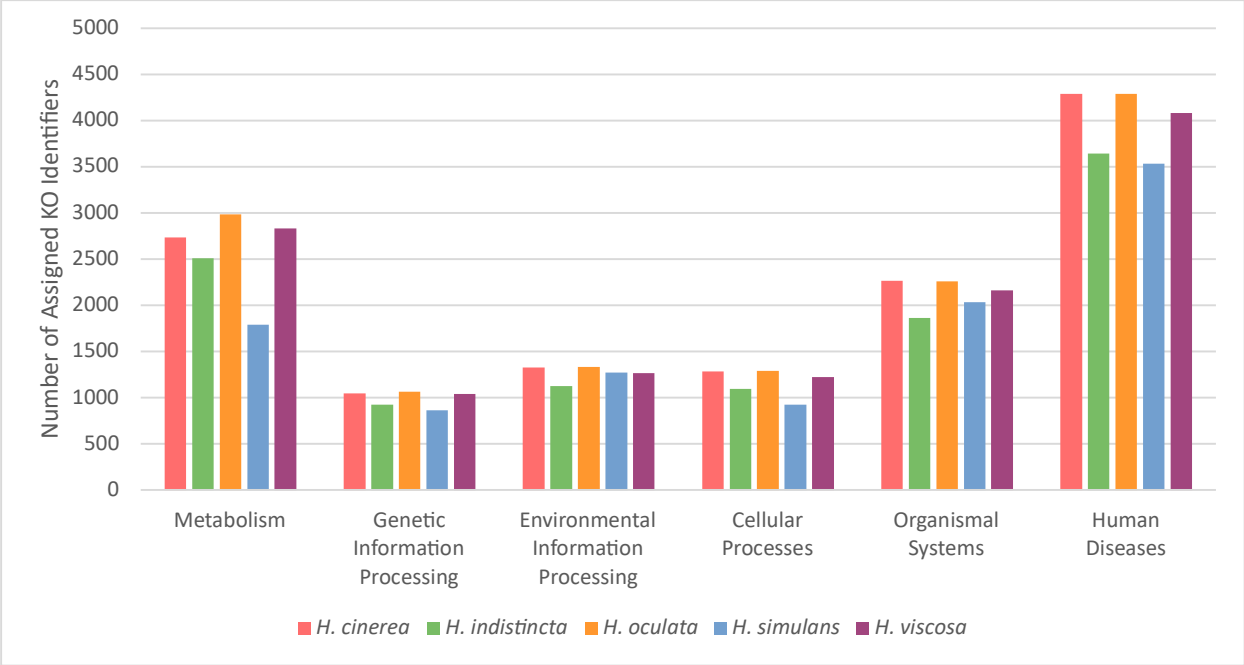


Figure 2.3 KEGG Orthology identifiers from six broad categories assigned to each of the five *Haliclona* species.

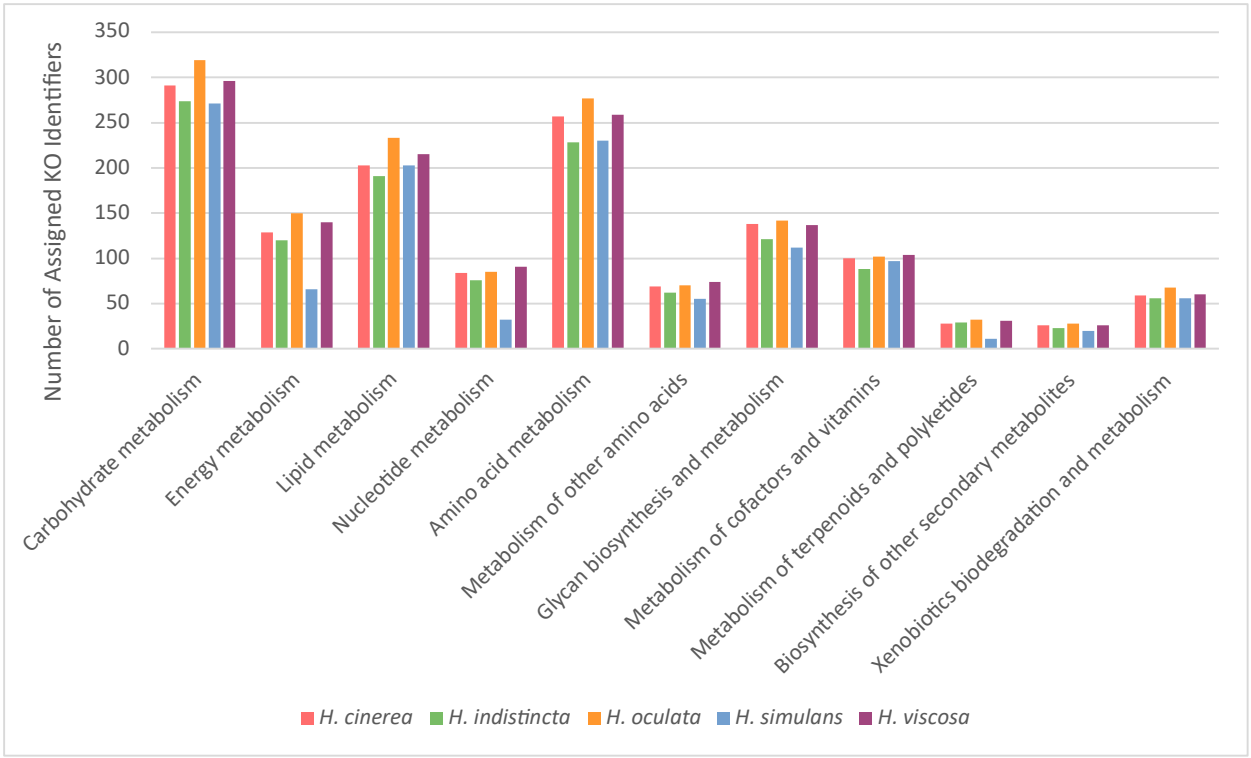


Figure 2.4 KEGG Orthology identifiers associated with metabolism that were assigned to each of the five *Haliclona* species.

Primary Metabolism of Nicotinic Acid in *Haliclona* Species

All five *Haliclona* species possess the necessary genes to convert tryptophan to QA (Figure 2.5). The only difference is that *H. cinerea* possesses an additional gene for tryptophan 2,3-dioxygenase (EC:1.13.11.11) which converts tryptophan to N-formylkynurenine. Furthermore, the enzyme nicotinate-nucleotide pyrophosphorylase (EC:2.4.2.19) was present in all species indicating a connection between tryptophan metabolism and the NAD^+ cycle. None of the five species possessed the necessary genes to convert L-aspartate to QA (data not shown).

All five *Haliclona* species possessed the necessary genes for the biosynthesis of NAD^+/$NADP^+$ (Figure 2.6). Furthermore, enzymes for the recycling of nicotinamide back into the NAD^+/$NADP^+$ cycle as well as converting it into N1-methylnicotinamide (N1-MNAM) for waste excretion were present. The enzyme nicotinate phosphoribosyltransferase (EC: 6.3.4.21), which is capable of converting NTAM to NTA, was not present in any of the five species. Only *H. viscosa* possessed the nicotinamidase gene for the conversion of NAM to NTA, but it most aligned with nicotinamidase sequences from other metazoans and bacteria rather than any sequence from *A. queenslandica*. No similar transcripts were found in the transcriptomes of the other four species using this nicotinamidase from *H. viscosa* as a query. Only one clear path towards the biosynthesis of NTA from QA was identified *via* pathway mapping. All five *Haliclona* species possess the genes for 5'-nucleotidase (EC:3.1.3.5) and

purine-nucleoside phosphorylase (EC:2.4.2.1) which can convert NTAM to nicotinic acid riboside (NTAR) and then NTA respectively according to KEGG.

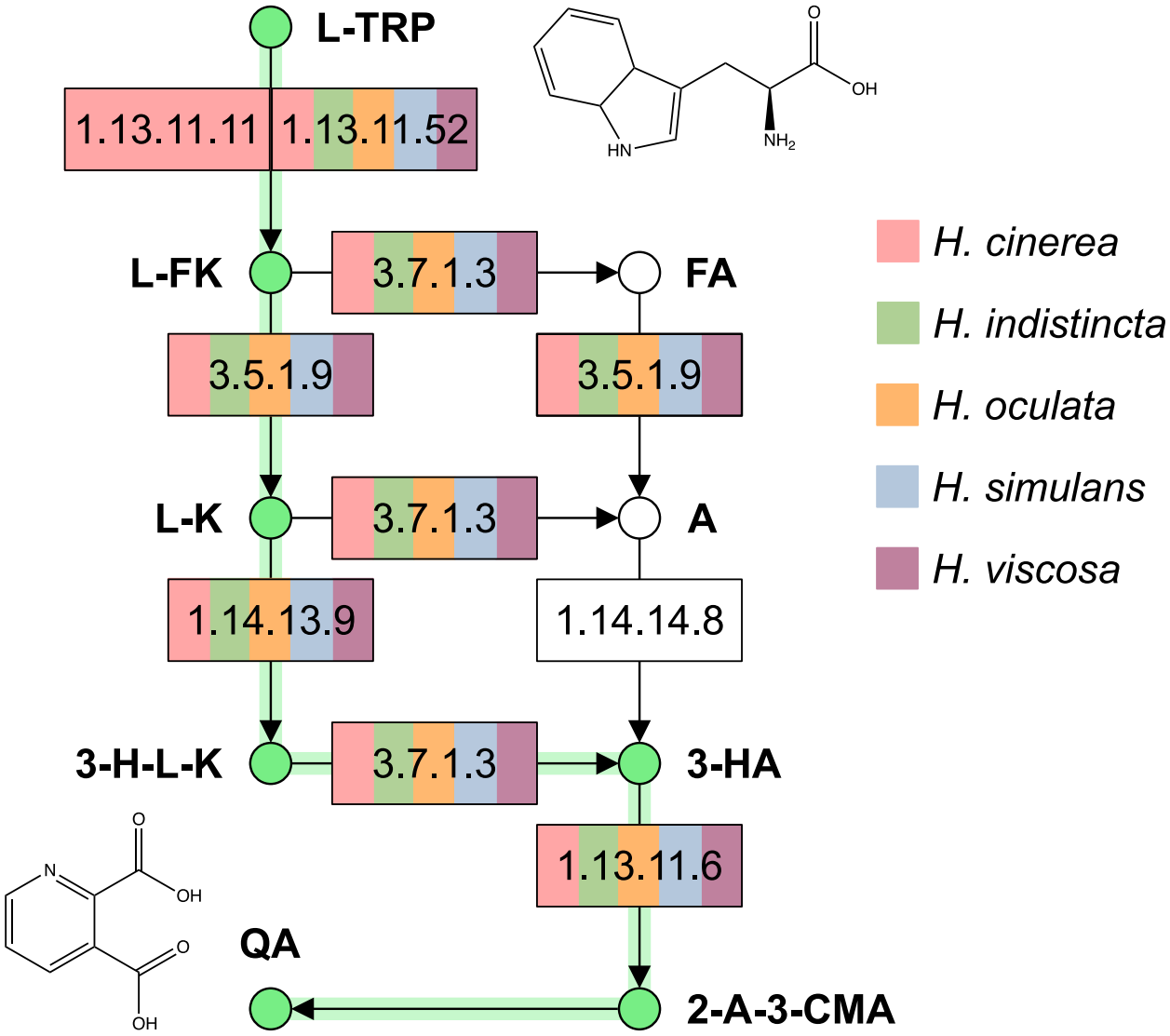


Figure 2.5 Tryptophan metabolism of five *Haliclona* species. Abbreviations are as follows: L-TRP, L-tryptophan; L-FK, L-Formylkynurenine; L-K, L-Kynurenine; 3-H-L-K, 3-Hydroxy-L-kynurenine; FA, Formyl-anthranilate; A, Anthranilate; 3-HA, 3-Hydroxyanthranilate; 2-A-3-CMA, 2-Amino-3-carboxymuconate semialdehyde; QA, Quinolinic acid. A possible pathway from L-TRP to QA is highlighted in green.

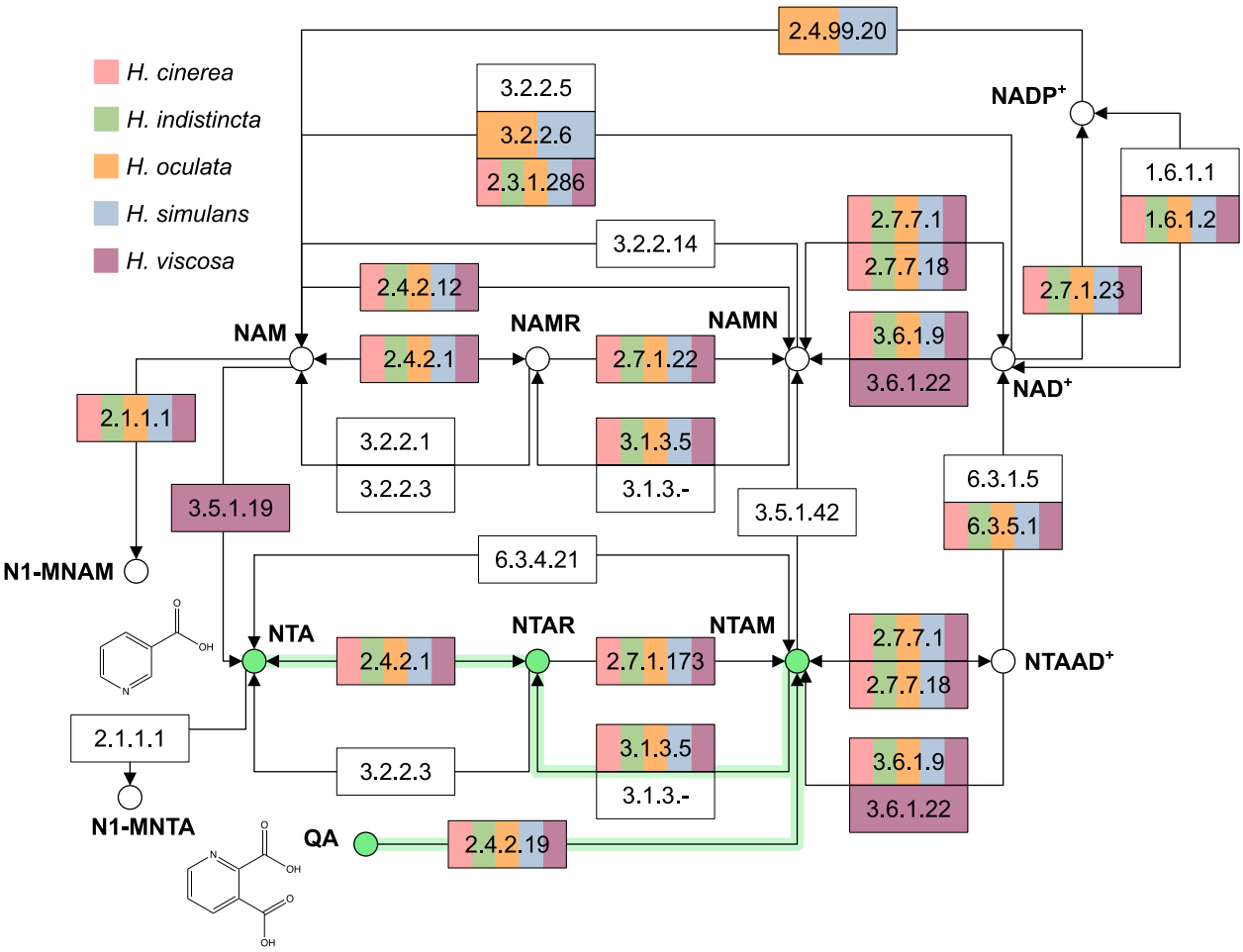


Figure 2.6 Nicotinate and Nicotinamide metabolism of five *Haliclona* species. Abbreviations are as follows: QA, Quinolinic acid; NTAM, Nicotinic acid mononucleotide; NTAAD+, Nicotinic acid adenine dinucleotide; NAMN, Nicotinamide mononucleotide; NAD+, Nicotinamide adenine dinucleotide; NADP+, Nicotinamide adenine dinucleotide phosphate; NAM, Nicotinamide; NAMR, Nicotinamide riboside; N1-MNAM, N1-Methylnicotinamide; NTA, Nicotinic acid; NTAR, Nicotinic acid riboside; N1-MNTA, N1-Methylnicotinic acid. The pathway from QA to NTA is highlighted in green.

Fatty Acid Biosynthesis in *Haliclona* Species

None of the *Haliclona* species possessed the necessary enzymes to synthesize fatty acids in the cytosol (Figure 2.7). No modular fatty acid synthase of type I biosynthesis similar to animals (FASN; EC:2.3.1.85), fungi (FAS1/FAS2; EC:2.3.1.86) or bacteria (fas; EC:2.3.1.-) were detected in any of the transcriptomes. Evidence of type II biosynthesis was detected in all species in that the enzymes [acyl-carrier-protein] S-malonyltransferase (fabD; EC:2.3.1.39) and 3-oxoacyl-[acyl-carrier-protein] synthase II (fabF; EC:2.3.1.179), were detected; a blastp search against the NCBI non-redundant database indicated poriferan origin as their closest hits were of similar proteins from *A. queenslandica*. Furthermore, genes encoding for 3-oxoacyl-[acyl-carrier protein] reductase (fabG; EC:1.1.1.100) were identified in each species with high sequence similarity to one from *A. queenslandica*. In addition, *H. indistincta* and *H. viscosa* possess the gene for HSD17B8 which has been reported to function as a 3-oxoacyl-[acyl-carrier protein] reductase alpha subunit (EC:1.1.1.62, 1.1.1.239). Finally all species possessed mitochondrial enzymes associated with type II fatty acid biosynthesis, namely mitochondrial enoyl-[acyl-carrier protein] reductase / trans-2-enoyl-CoA reductase (MECR; EC:1.3.1.-). However, no genes coding for enzymes with dehydratase activity, such as FabA, were detected indicating that the type II fatty acid biosynthesis pathway is incomplete in these organisms. The five species also possessed complete pathways for the elongation of hexadeconoate and longer fatty acids in the mitochondria (Figure 2.8A) and endoplasmic reticulum (Figure 2.8B). A β-oxidation pathway for the degradation of hexadecanoate and smaller fatty acids to acetyl-CoA was detected in all five *Haliclona* species (Figure 2.9).

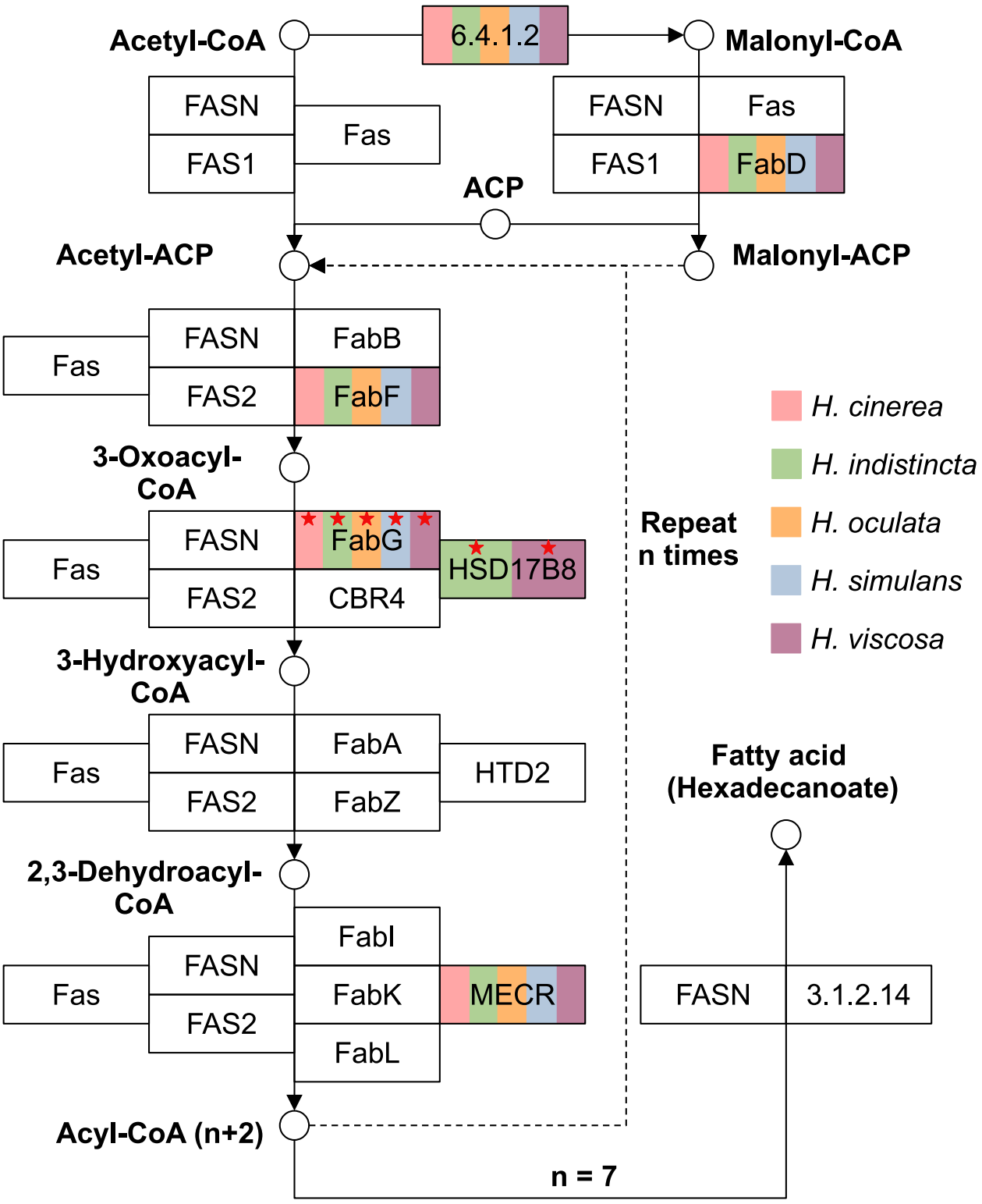


Figure 2.7 Fatty acid biosynthesis in the cytosol of five *Haliclona* species. Red stars indicate a gene was not identified by KEGG but instead by manual analysis.

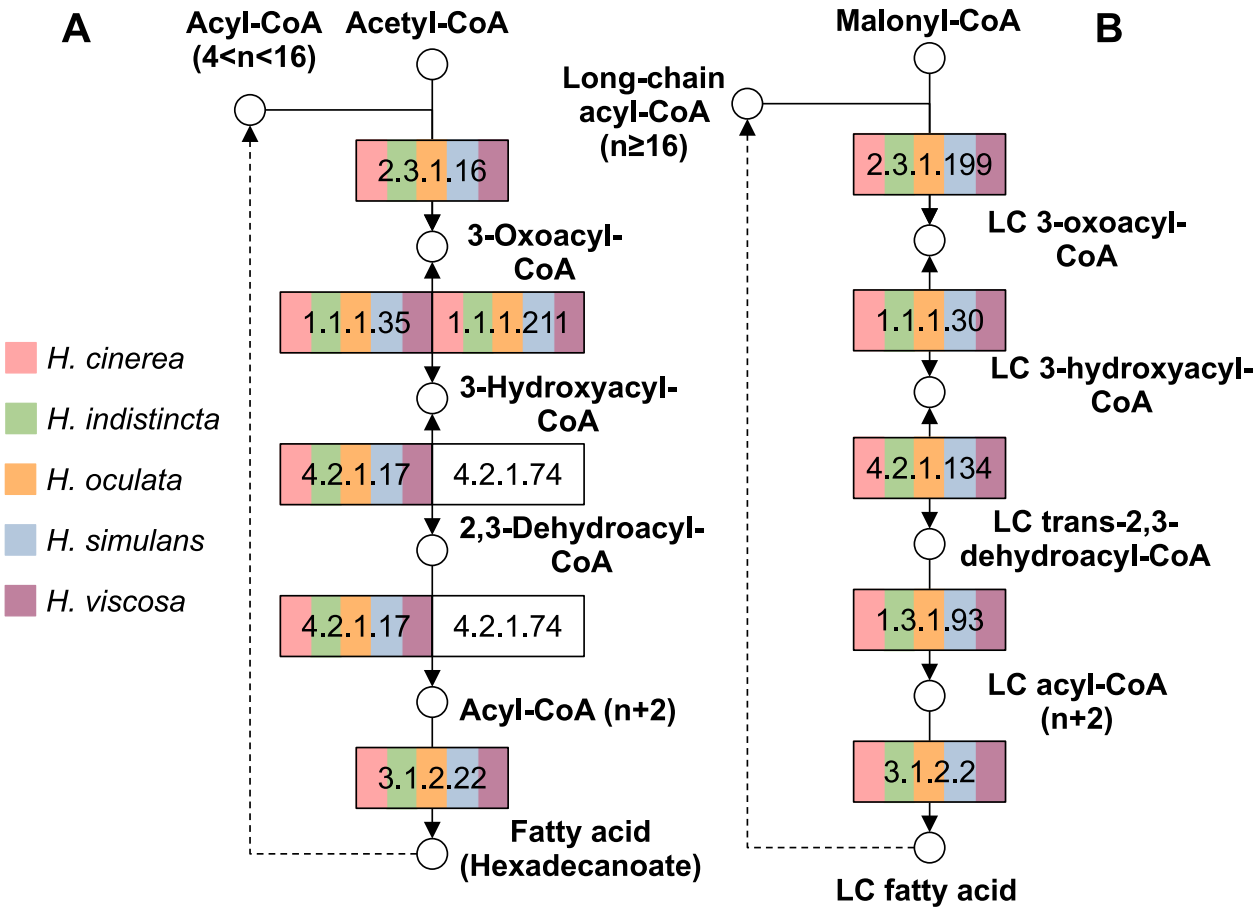


Figure 2.8 (**A**) Fatty acid elongation in the mitochondria of five *Haliclona* species. (**B**) Fatty acid elongation in the endoplasmic reticulum of five *Haliclona* species. LC stands for long-chain.

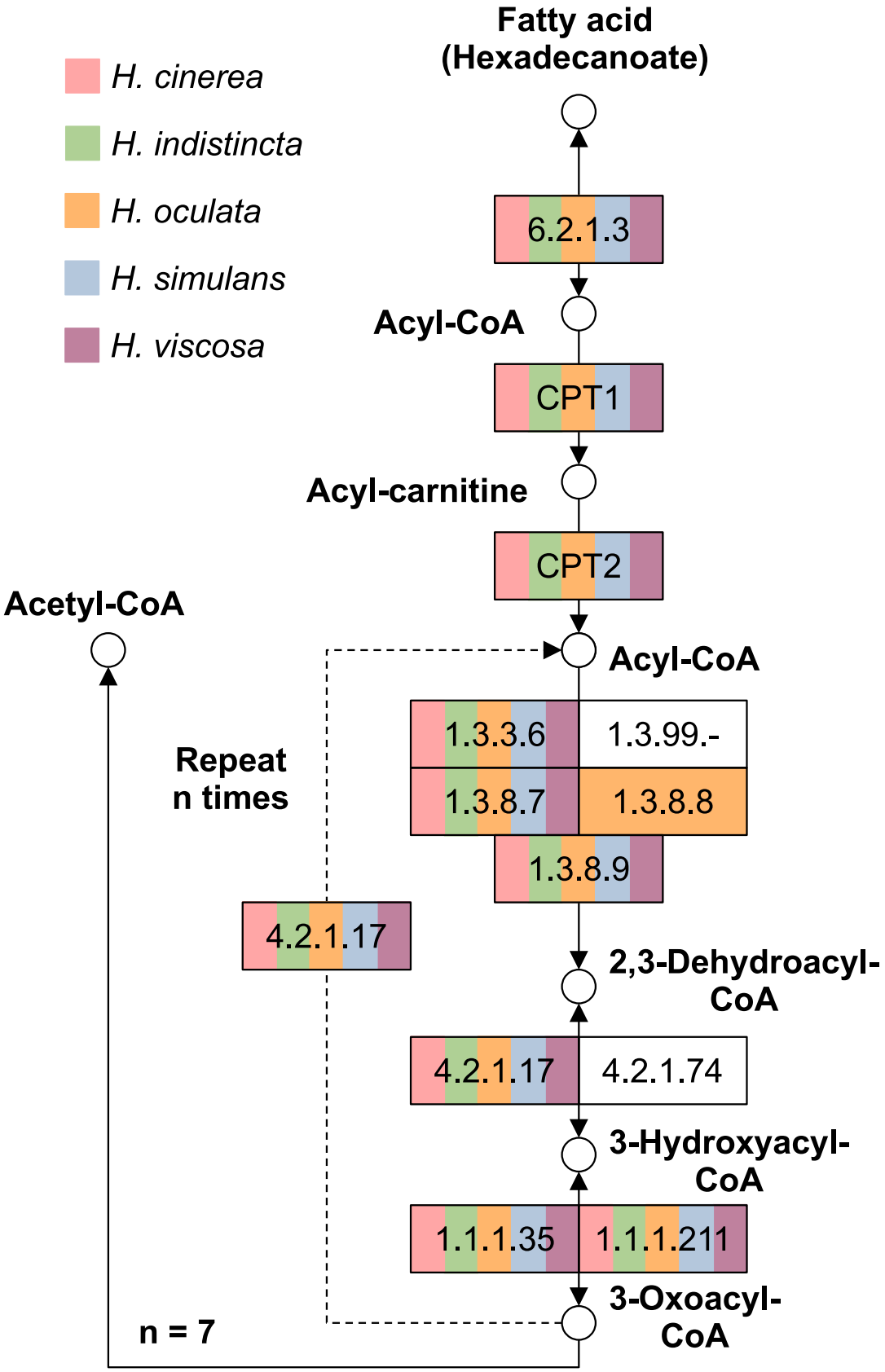


Figure 2.9 Fatty acid degradation of five *Haliclona* species.

Steroid Biosynthesis in *Haliclona* Species

The only KEGG biosynthetic pathway associated with specialized metabolites which possessed a degree of completion was that for steroid biosynthesis. Four of the five *Haliclona* species lacked at least one gene necessary for the complete animal steroid biosynthesis pathway in which farnesyl-PP is converted to cholesterol (Figure 2.10). *Haliclona cinerea*, *H. indistincta*, *H. simulans* and *H. viscosa* solely lack delta(24)-sterol reductase (EC: 1.3.1.72) which is one of the final enzyme necessary for formation of cholesterol. However, KEGG did identify this gene as being present in the *H. oculata* transcriptome. A manual blastx analysis against the nr database revealed hits towards animal proteins with the correct functional annotation. However, no similar proteins could be found from *A. queenslandica* or in the other four *Haliclona* transcriptomes. Originally, KEGG did not detect *H. indistincta* as possessing a lanosterol synthase gene (EC: 5.4.99.7), but manual analysis revealed the presence of this gene in the transcriptome with homology to that of *A. queenslandica*. In contrast, *H. cinerea* lacked several additional genes despite its transcriptome assembly being the largest. These missing genes include squalene monooxygenase (EC:1.14.14.17) and sterol 14α -demethylase (EC:1.14.14.154 1.14.15.36). While KEGG did not detect this species as expressing the gene for lanosterol synthase, manual analysis also identified a small transcript which aligned to that of *A. queenslandica*, but from which an open reading frame was not extracted.

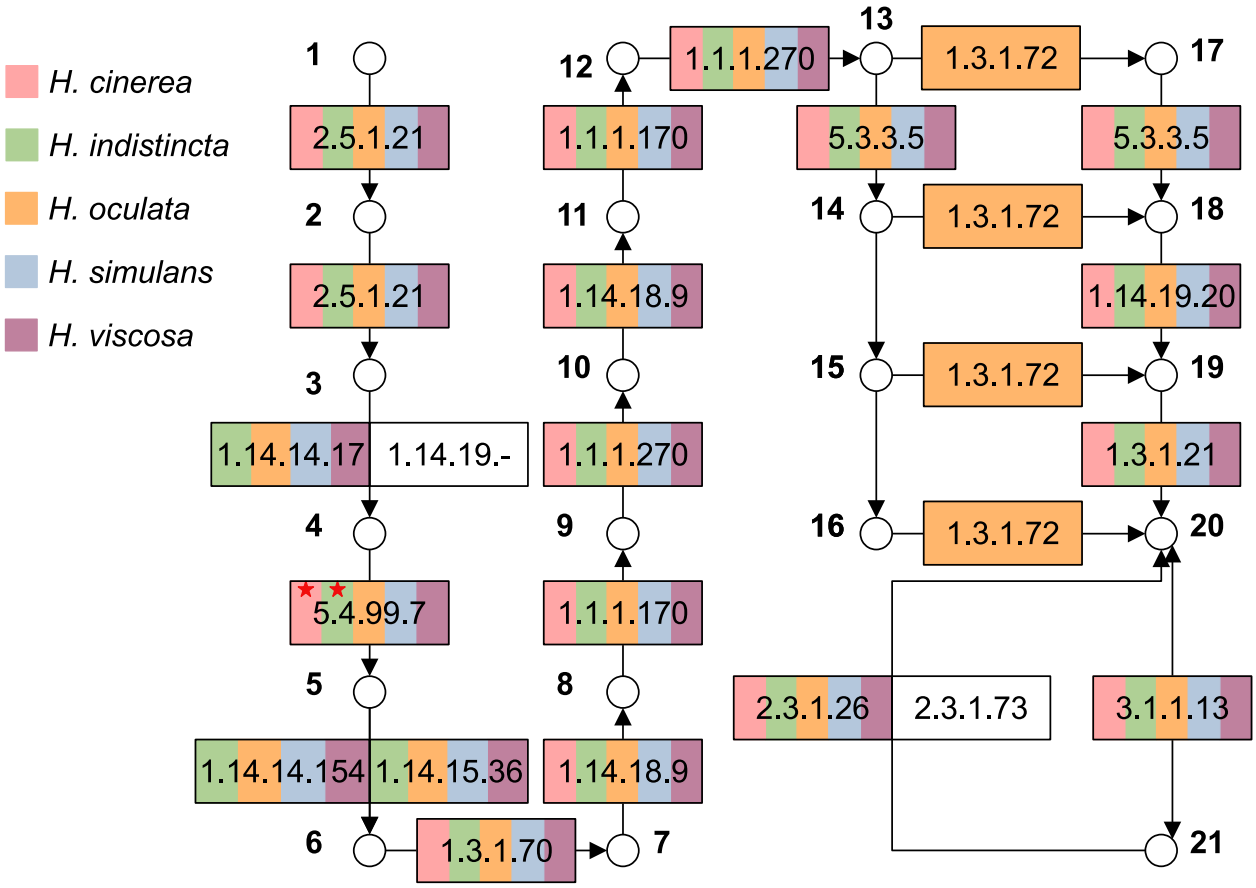


Figure 2.10 Animal steroid biosynthesis in five *Haliclona* species. Abbreviations are as follows: 1, Farnesyl-PP; 2, Pre-Squalene-PP; 3, Squalene; 4, (S)-Squalene-2,3-epoxide; 5, Lanosterol; 6, 4,4-Dimethyl-cholesta-4,14,24-trienol; 7, 14-Demethyllanosterol; 8, 4-Methylzymosterol-carboxylate; 9, 3-Keto-4-methylzymosterol; 10, 4-Methylzymosterol; 11, 4α-Carboxy-5α-cholesta-8,24-dien-3β-ol; 12, Zymosterone; 13, Zymosterol; 14, Cholesta-7,24-dien-3β-ol; 15, 7-Dehydrodesmosterol; 16, Desmosterol; 17, Cholesta-8-en-3β-ol; 18, Lathosterol; 19, 7-Dehydrocholesterol; 20, Cholesterol; 21, Cholesterol ester. Red stars indicate a gene was not identified by KEGG but instead by manual analysis.

Genes associated with plant sterol biosynthesis were also detected by KEGG (Figure 2.11). When screened against the nr database several displayed the pattern of being most homologous to a protein from *A. queenslandica*, but the rest of the alignments were of plant origin. Such a signature was interpreted as the enzymes being capable of performing the enzymatic reactions necessary for the biosynthesis of sterols associated with plants. First, all five species possessed sterol methyltransferase (SMT1; EC: 2.1.1.41). Second, cycloeucalenol cycloisomerase (EC:5.5.1.9) was found in all but *H. cinerea*. Third, 7-dehydrocholesterol reductase (DWF5; EC:1.3.1.21) was found in *H. cinerea*, *H. oculata* and *H. simulans*. Several additional genes which may perform enzymatic reactions found in the plant pathway were also detected. In particular, transcripts from *H. simulans* and *H. viscosa* were identified which had top hits towards a predicted methylsterol monooxygenase 1-like protein from *A. queenslandica* (XP_011404818.2), but also with proteins from other animals, fungi, and plants. Furthermore, a single transcript in *H. cinerea* which has a higher sequence similarity to the 24-methylenesterol C-methyltransferase (SMT2; EC:2.1.1.143) gene from *A. thaliana* rather than its SMT1 gene was detected. This gene most significantly aligned with sequences from several bacteria and numerous eukaryotes included the segmented worm *Capitella telata*, the chonaoflagellate *Capsaspora owczarzaki*, algae and plants. Because there was no homologous sequence to *A. queenslandica*, it is difficult to conclude if this SMT2 gene identified in the *H. cinerea* transcriptome is of sponge origin or contamination. All other enzymes associated with plant sterol biosynthesis were either absent or incorrectly identified as present by KEGG when in reality they represented their animal homologs.

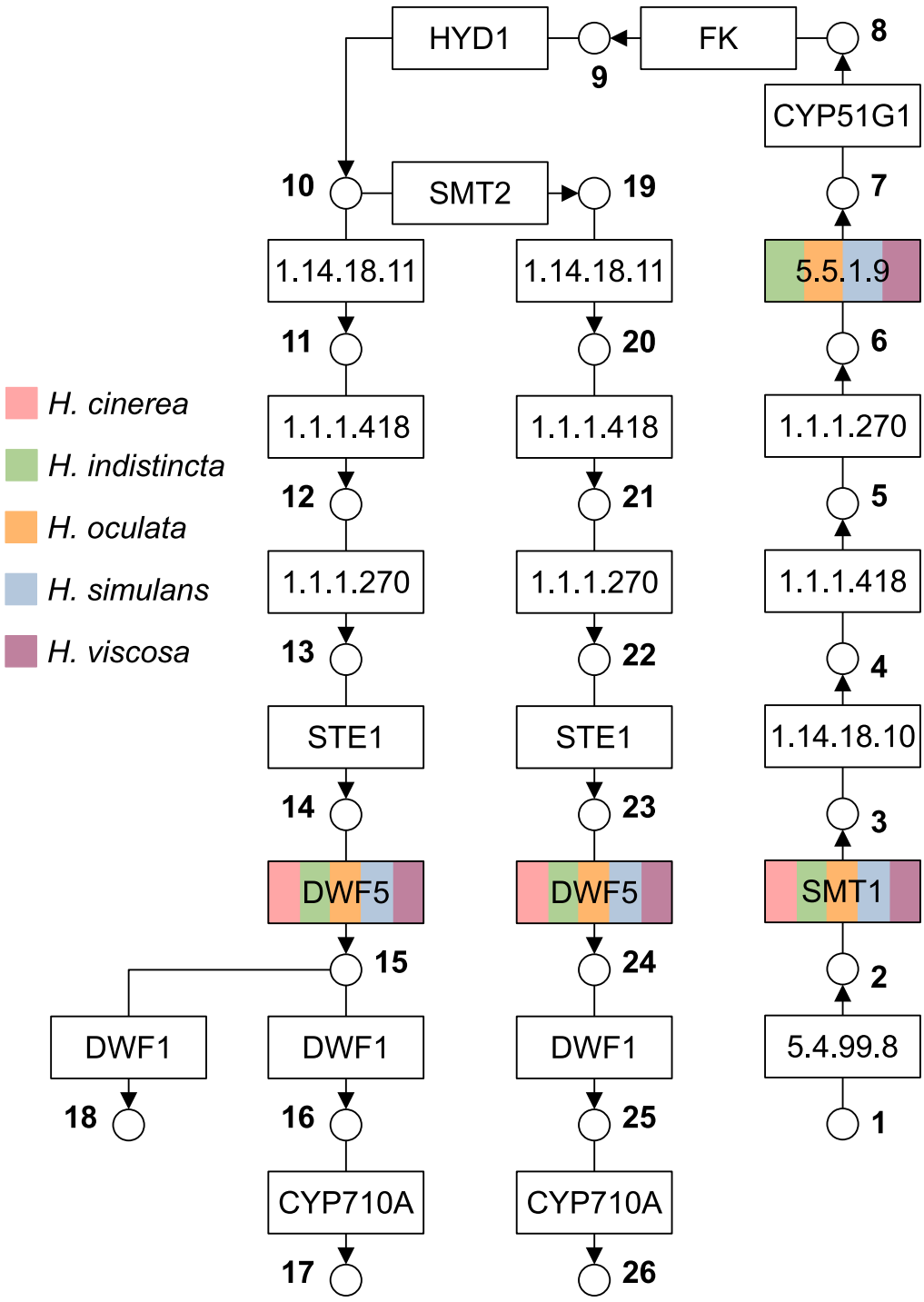


Figure 2.11 Plant steroid biosynthesis in five *Haliclona* species. Abbreviations are as follows: 1, (S)-Squalene-2,3-epoxide; 2, Cycloartenol; 3, 24-Methylenecycloartenol; 4, 3β-Hydroxy-4β,14α-dimethyl-9β,19-cyclo-5α-ergost-24(24(1))-en-4α-carboxylate; 5, Cycloeucalenone; 6, Cycloeucalenol; 7, Obtusifoliol; 8, δ8,14-Sterol; 9, 4α-Methylfecosterol; 10, 24-Methylenelophenol; 11, 3β-Hydroxyergosta-7,24(24(1))-dien-4α-carboxylate; 12, Episterone; 13, Episterol; 14, 5-Dehydroepisterol; 15, 24-Methylenecholesterol; 16, 24-epicampesterol; 17, Brassicasterol; 18, Campesterol; 19, 24-Ethylidenelophenol; 20, 4α-Carboxy-stigmasta-7,24(24(1))-dien-3β-ol; 21, Avenastenone; 22, δ7-Avenasterol; 23, 5-Dehydroavenasterol; 24, Isofucosterol; 25, Sitosterol; 26, Stigmasterol.

Discussion

In this chapter I present a larger and more comprehensive transcriptome assembly of *H. indistincta* by combining new RNAseq data from MIIG1388 with previously sequenced specimens [57]. This assembly exhibits a higher total Mbp size, number of true genes excluding isoforms and N50 value. Furthermore, novel transcriptome assemblies for both *H. cinerea* and *H. viscosa* were generated the latter of which was also sequenced in this chapter. For consistent comparison, the transcriptomes of *H. oculata* and *H. simulans* were also re-assembled using the same read processing methods and assembly parameters as that for *H. cinerea*, *H. indistincta*, and *H. viscosa*. This resulted in larger assemblies, but a similar number of true genes excluding isoforms. In turn this likely indicates that the difference in size is due to a difference in isoform number. Because the version of Trinity and parameters used for the old transcriptomes are not available in their corresponding publication, it is difficult to assess why such a size difference was observed. In addition, the N50 value of the *H. oculata* and *H. simulans* assemblies were significantly improved upon by 81.58% and 47.05% respectively. The BUSCO evaluation of each transcriptome indicated a high level of duplication which is likely due to not filtering out isoforms prior to running this analysis (Table 2.3). Furthermore, a high degree of completeness was observed with each transcriptome based on the BUSCO evaluation. However, a degree of caution should be taken when interpreting this result. Polyadenylation is a process used by eukaryotes to stabilize mRNA whereas in prokaryotes it marks the molecule for degradation [58]. As such, the selection of mRNA with poly(A) tails for library preparation and sequencing was a means to enrich for sponge mRNA. However, this would process likely also captured the mRNA of contaminating eukaryotic organisms such as crustaceans and nematodes. Furthermore, it is often assumed that selecting for mRNA with a poly(A) tail excludes bacterial mRNA. It should be emphasized that this is only a method to enrich for eukaryotic mRNA rather than exclude that of bacteria. Thus, the complete absence of bacterial sequences in the produced transcriptomes cannot be assumed. The folly of making such an assumption can be illustrated in analyses which have identified bacterial contamination in RNA-seq data derived from enriched eukaryotic, polyadenylated mRNA [59].

The BBE-like proteins of *N. tabacum*, which are involved in the condensation of NTA with an *N*-methylpyrrolinium, displayed significant matches in each of the *Haliclona* transcriptomes when used as query sequences [38]. Based on this information, it appears that each of the *Haliclona* species share homologous BBE-like proteins with a common evolutionary origin as their most significant alignments in the non-redundant database are with cnidarian sequences. It is especially promising that *H. indistincta* and *H. viscosa*, the two species which produce 3-APs, both appear to have an additional gene for BBE-like proteins with higher sequence similarity to those from Actinobacteria in the NCBI nr database. While this discovery provides a promising and alternative lead on identifying the biosynthetic origin of 3-APs, several factors remain which dispute this concept. The primary one is that no

significantly similar sequences to the isoflavone reductase homolog A622 of *N. tabacum* were identified in the *Haliclona* transcriptomes. This enzyme is hypothesized to convert NTA into a 3,6-dihydropyridine which can function as nucleophile towards an *N*-methylpyrollinium [60]. Afterwards, it is then hypothesized that the BBE-like protein oxidizes the 3,6-dihydropyridine ring back into a pyridine ring. As such the BBE-like proteins in *H. indistincta* may not necessarily be responsible for the condensation of NTA with another moiety, but rather the oxidation of a reduced pyridine ring. BBE-like proteins catalyse a variety of reactions involved in the biosynthesis of specialized metabolites, but are largely associated with archaea, bacteria, fungi and plants [61]. Whether or not these enzymes are involved in 3-AP biosynthesis, it is promising that sponges also appear to encode for BBE-like proteins in their genomes as these enzymes are also associated with the biosynthesis of complex specialized metabolites.

In addition, no strong evidence was found that any *Haliclona* species possessed genes encoding for AMP-binding enzymes which activate NTA. This conclusion is based on the significant local hits with the aforementioned AMP-binding enzyme queries most aligning with 4-coumarate—CoA ligase from *A. queenslandica* (XP_003386175.1) and the luciferase polypeptide from *S. domuncula* (CAR31336.1). The latter has been functionally characterized and shown to not be involved in specialized metabolism, but rather a photoreception system [62]. For this reason, the similar sequences detected in the transcriptomes of *H. indistincta* and *H. viscosa* are likely not responsible for 3-AP biosynthesis. Furthermore, it is likely that the method of conjugation between the alkyl chain of one monomer and the nitrogen atom of another is not similar to the reaction catalysed by trigonelline synthase as no significant alignments were detected when using this protein as a local query.

As expected of animals, all five *Haliclona* species possessed the necessary genes allowing for the *de novo* biosynthesis of NTAM, a fundamental metabolite in the NAD+ cycle, from tryptophan [21]. However, from that point it is difficult to devise a pathway towards the biosynthesis of NTA. This is because the fundamental enzyme, nicotinamide amidohydrolase, was not clearly found in any of the five *Haliclona* transcriptomes. While KEGG identified one protein sequence from *H. viscosa* as having this function, it is difficult to assess whether this sequence is eukaryotic contamination. This enzyme also appears to be lacking in the KEGG pathway for the model sponge *A. queenslandica* (data not shown) which may indicate that sponges fundamentally lack a 4-step salvage pathway. While nicotinamide amidohydrolase has been reported from the proteome of *A. queenslandica*, a lack of sharing the sequence makes it difficult to identify similar homologous proteins in the *Haliclona* transcriptomes [63]. Similarly, nicotinate phosphoribosyltransferase which is known to convert NTA to NTAM with KEGG reporting the reverse action as being possible, is lacking in all *Haliclona* transcriptomes [24].

The only identifiable pathway towards NTA biosynthesis in the *Haliclona* transcriptomes is *via* two detected enzymes, 5'-nucleotidease and purine-nucleoside phosphorylase, which all

species possessed. 5'-nucleotidase is known to act upon a broad range of substrates [64]. Specifically, the conversion of NTAM to NTAR has been reported by this type of enzyme from both mammals and yeast [25,65]. In addition, this enzyme from humans has been reported to catalyse the conversion of NTAM to NTAR, albeit at a lower efficiency than upon the NAM analogues [25]. In turn, purine-nucleoside phosphorylase has been reported to further convert NTAM to NTA in mammals and fungi [66]. Unless other pathways towards the biosynthesis of NTA exist outside of the KEGG database, the use of these two aforementioned enzymes is likely the only way in which *H. indistincta* and *H. viscosa* synthesize the pyridine ring of the 3-APs *de novo* from tryptophan assuming a host origin other than a dietary source.

The lack of a complete pathway for the biosynthesis of palmitic acid in all five *Haliclona* species is puzzling as fatty acids are essential components for both triglycerides and phospholipids. In turn these two types of compounds are widely utilized for energy storage and the construction of cell membranes respectively [67,68]. Furthermore, the majority of 3-APs possess an aliphatic chain that is no longer than that of palmitic acid (personal observation); if a fatty acid origin of 3-APs from the sponge were true it would be expected that the host organism possesses a complete biosynthetic pathway for this building block. Such results are consistent with a prior examination of the model sponge *A. queenslandica* and its inability to synthesize common short-chain fatty acids [16].

An additional pathway for fatty acid elongation using acetyl-CoA instead of malonyl-CoA as the elongation factor for the synthesis of palmitic acid also exists in the KEGG database. KEGG indicates that all *Haliclona* species possess the necessary enzymes to fulfil this pathway (data not shown). However, this pathway, known as the reverse β-oxidation pathway, is only known to exist in the mitochondria of the single-celled algae *Euglena gracilis* [69]. As such, its apparent presence in the *Haliclona* transcriptomes should be interpreted with caution because KEGG lists both the normal β-oxidation pathway, which is involved in the degradation of fatty acids, and the aforementioned reverse pathway as utilizing the same types of enzymes with the same KEGG orthology numbers. Indeed, all five *Haliclona* species were also shown by KEGG to possess a complete pathway for the degradation of fatty acids. This pathway is performed by the mitochondria of numerous animals and for this reason it is more likely that the *Haliclona* species also utilize the associated enzymes for degradation, rather than alternative biosynthesis, of fatty acids [70]. However, this conclusion cannot truly be absolute without biochemical analyses of the sponge enzymes involved in the β-oxidation pathway. Rather than complete *de novo* biosynthesis of fatty acids, it may be that sponges acquire these molecules *via* alternative means such as their diet or microbial symbionts [71,72]. If necessary, larger fatty acids derived from these sources could possibly be broken down into smaller derivatives by the sponge β-oxidation pathway and later utilized for 3-AP biosynthesis. Furthermore, if longer fatty acids are necessary then elongation of dietary fatty acids would be possible from the mitochondria as indicated by the complete pathway.

Sterols are essential compounds found in all animals with sponges being consistent sources of those with interesting structures and bioactivities [73]. Indeed, the species *H. cinerea*, *H. oculata* and *H. simulans* involved in this study have all been shown to be the source of unique sterols [74-76]. Ample evidence from isotope-feeding studies and genome sequencing indicate that sponges are capable of *de novo* biosynthesis of sterols in addition to acquiring these compounds from their diet [16,77]. Furthermore, sterol biosynthesis has been a trait largely found in eukaryotic organisms rather than bacteria such as those which may be associated with sponges [78]. For these reasons, it is to be expected that the Irish *Haliclona* species would possess pathways to synthesize common sterols, such as cholesterol, which could then be modified into more exotic derivatives. This is supported by an almost complete pathway for the biosynthesis of animal-associated sterols in each *Haliclona* species save a delta(24)-sterol reductase which is the final enzyme involved in cholesterol biosynthesis (Figure 2.10). The primary anomaly is the fact that KEGG indicated *H. oculata* solely possesses this missing gene. Considering that no homologous protein is present in the *A. queenslandica* genome, which was sequenced from cleaned larval tissue, it is possible that this sequence is instead the result of eukaryotic contamination [16,79]. This is supported by the fact that top hits for this sequence in the non-redundant database are from microscopic eukaryotic organisms such as the flatworm *Macrostomum lignano*. While the libraries for these transcriptomes were constructed from polyadenylated mRNA, which would largely exclude bacterial mRNA, this is no guaranteed avoidance of mRNA from smaller eukaryotes which may be associated with the sponge holobiome. Despite the lack of a likely delta(24)-sterol reductase in these *Haliclona* species, a number of explanations on how cholesterol is acquired other than diet can be inferred from prior publications on other organisms. For example microbial symbionts may be responsible for filling in this gap similar to the biosynthesis of cholesterol by *Bathymodiolus platifrons* and one of its endosymbionts [80].

At first glance, it is questionable why the five *Haliclona* species only possess several genes associated with phytosterol biosynthesis. However, such an observation is consistent with the reported isolation of phytosterols, such as 24-methylenecholesterol, from some of these species [74,76]. Furthermore, similar results were observed when analysing sterol biosynthetic genes of *A. queenslandica* [16]. Rather it is more likely that sterol biosynthetic pathways of sponges do not conform to the classic animal, fungal or plant pathways found on KEGG. Such a quality is not unheard of and can be observed in other non-model organisms such as diatoms [81]. Evidence of the connectivity between the lanosterol and cycloartenol pathway is evident by sterol dealkylation performed by sponges in which phytosterols, such as 24-methylenecholesterol, are dealkylated into desmosterol and then converted into cholesterol [82]. The gene responsible for sterol demethylation has been identified in the ciliate *Tetrahymena thermophila*, although no similar sequences could be identified in the *Haliclona* transcriptomes (data not shown) [83]. Overall, it is likely that biochemical analyses on individual enzymes will be required to elucidate a sterol biosynthetic pathway of sponges.

However, the presence of a nearly complete lanosterol pathway and several enzymes associated with the cycloartenol pathway confirm that the five assessed *Haliclona* species are likely a source of common sterols. Combining this observation with the history of the *H. cinerea*, *H. oculata* and *H. simulans* sources of novel sterols, it can be hypothesized that *H. indistincta* and *H. viscosa* may represent promising targets for the isolation of these compounds.

No other complete biosynthetic pathways for specialized metabolites, or primary metabolites of interest, where detected in the KEGG database. This includes the pathway for the biosynthesis of a polyketide backbone *via* a type II system. This is somewhat surprising because, while *H. cinerea*, *H. oculata* and *H. simulans* are not known to be the source of specialized metabolites with a hypothesized polyketide synthase origin, *H. indistincta* and *H. viscosa* are. Particularly, their associated 3-AP alkaloids were expected to involve polyketide biosynthesis in some manner based on the biosynthesis of haminol [19,20]. However, this possibly excludes only a type II PKS system as the type I PKS are not included in any KEGG metabolic pathway due to their modular nature. Combined with the observation that all five *Haliclona* species appear to lack a complete pathway for the biosynthesis of fatty acids, several conclusions can be drawn regarding the biosynthetic origin of 3-APs. That is, if a host origin is truly correct then a type I PKS is the remaining candidate system. Alternatively, fatty acids derived from the diet of the sponge and incorporated into 3-AP biosynthesis may represent a means to create these molecules without needing any type of FAS. Finally, a bacterial origin should highly be considered as genes from associated bacteria were largely ignored in this work based on the library construction of these transcriptomes.

In conclusion the *de novo* sequencing, assembly, and analysis of five *Haliclona* transcriptomes has provided insight into possible sources of known and novel bioactive compounds. Regarding the 3-APs, if a host origin is true then it appears a possible origin of the pyridine ring could be *via de novo* synthesis from tryptophan. In particular, a 5'-nucleotidase and purine-nucleoside phosphorylase may allow for the conversion of NTAM to NTA. However, investigations into a possible metazoan origin of the aliphatic chain of 3-APs indicates that *de novo* synthesis of fatty acids or type II polyketides is not a likely method. As an alternative to the polyketide hypothesis on 3-AP biosynthesis, BBE-like proteins were detected in all five *Haliclona* species with *H. indistincta* and *H. viscosa*, the two associated with 3-APs, possessing a second copy of these genes. In addition, a suite of biosynthetic genes associated with both plant and animal sterols were identified in all *Haliclona* species indicating that the biosynthetic pathway for sponge sterols may not neatly conform to established models. Overall, the aforementioned genes discovered from this work represent possible targets for functional characterization to identify and understand known and novel sources of bioactivity from these *Haliclona* species.

References

1. Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10, 57–63. https://doi.org/10.1038/nrg2484
2. Kumar, R., Ichihashi, Y., Kimura, S., Chitwood, D.H., Headland, L.R., Peng, J., Maloof, J.N., Sinha, N.R., 2012. A high-throughput method for Illumina RNA-seq library preparation. Frontiers in Plant Sciience 3. https://doi.org/10.3389/fpls.2012.00202
3. Kumar, N., Lin, M., Zhao, X., Ott, S., Santana-Cruz, I., Daugherty, S., Rikihisa, Y., Sadzewicz, L., Tallon, L.J., Fraser, C.M., Dunning Hotopp, J.C., 2016. Efficient enrichment of bacterial mRNA from host-bacteria total RNA samples. Scientific Reports 6, 34850. https://doi.org/10.1038/srep34850
4. Byrne, A., Cole, C., Volden, R., Vollmers, C., 2019. Realizing the potential of full-length transcriptome sequencing. Philosophical Transactions of the Royal Society B: Biological Sciences 374, 20190097. https://doi.org/10.1098/rstb.2019.0097
5. Aguilar-Camacho, J.M., McCormack, G.P., 2017. Molecular responses of sponges to climate change, in: Carballo, J.L., Bell, J.J. (Eds.), Climate change, ocean acidification and sponges: impacts across multiple levels of organization. Springer International Publishing, Cham, pp. 79–104. https://doi.org/10.1007/978-3-319-59008-0_4
6. Riesgo, A., Farrar, N., Windsor, P.J., Giribet, G., Leys, S.P., 2014. The analysis of eight transcriptomes from all poriferan classes reveals surprising genetic complexity in sponges. Molecular Biology and Evolution 31, 1102–1120. https://doi.org/10.1093/molbev/msu057
7. Riesgo, A., Peterson, K., Richardson, C., Heist, T., Strehlow, B., McCauley, M., Cotman, C., Hill, M., Hill, A., 2014. Transcriptomic analysis of differential host gene expression upon uptake of symbionts: a case study with *Symbiodinium* and the major bioeroding sponge *Cliona varians*. BioMed Central Genomics 15, 376. https://doi.org/10.1186/1471-2164-15-376
8. Simion, P., Philippe, H., Baurain, D., Jager, M., Richter, D.J., Di Franco, A., Roure, B., Satoh, N., Quéinnec, É., Ereskovsky, A., Lapébie, P., Corre, E., Delsuc, F., King, N., Wörheide, G., Manuel, M., 2017. A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. Current Biology 27, 958–967. https://doi.org/10.1016/j.cub.2017.02.031
9. Pita, L., Hoeppner, M.P., Ribes, M., Hentschel, U., 2018. Differential expression of immune receptors in two marine sponges upon exposure to microbial-associated molecular patterns. Scientific Reports 8, 16081. https://doi.org/10.1038/s41598-018-34330-w
10. González-Aravena, M., Kenny, N.J., Osorio, M., Font, A., Riesgo, A., Cárdenas, C.A., 2019. Warm temperatures, cool sponges: the effect of increased temperatures on the Antarctic sponge *Isodictya* sp. PeerJ 7, e8088. https://doi.org/10.7717/peerj.8088
11. Ereskovsky, A.V., Richter, D.J., Lavrov, D.V., Schippers, K.J., Nichols, S.A., 2017. Transcriptome sequencing and delimitation of sympatric *Oscarella* species (*O. carmela* and *O. pearsei* sp. nov) from California, USA. PLOS ONE 12, e0183002. https://doi.org/10.1371/journal.pone.0183002
12. Li, J., Harata-Lee, Y., Denton, M.D., Feng, Q., Rathjen, J.R., Qu, Z., Adelson, D.L., 2017. Long read reference genome-free reconstruction of a full-length transcriptome from *Astragalus membranaceus* reveals transcript variants involved in bioactive compound biosynthesis. Cell Discovery 3, 1–13. https://doi.org/10.1038/celldisc.2017.31
13. Brunson, J.K., McKinnie, S.M.K., Chekan, J.R., McCrow, J.P., Miles, Z.D., Bertrand, E.M., Bielinski, V.A., Luhavaya, H., Oborník, M., Smith, G.J., Hutchins, D.A., Allen, A.E., Moore, B.S., 2018. Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom. Science 361, 1356–1358. https://doi.org/10.1126/science.aau0382
14. Kanehisa, M., 2002. The KEGG database, in: '*In silico*' simulation of biological processes. John Wiley & Sons, Ltd, pp. 91–103. https://doi.org/10.1002/0470857897.ch8

15. Karp, P.D., Riley, M., Paley, S.M., Pellegrini-Toole, A., 2002. The MetaCyc database. Nucleic Acids Research 30, 59–61. https://doi.org/10.1093/nar/30.1.59
16. Gold, D.A., O'Reilly, S.S., Watson, J., Degnan, B.M., Degnan, S.M., Krömer, J.O., Summons, R.E., 2017. Lipidomics of the sponge *Amphimedon queenslandica* and implication for biomarker geochemistry. Geobiology 15, 836–843. https://doi.org/10.1111/gbi.12253
17. Tribalat, M.-A., Marra, M.V., McCormack, G., Thomas, O., 2016. Does the chemical diversity of the order Haplosclerida (phylum Porifera: class Demospongia) fit with current taxonomic classification? Planta Medica. https://doi.org/10.1055/s-0042-105879
18. Andersen, R.J., Van Soest, R.W.M., Kong, F., 1996. Chapter three 3-alkylpiperidine alkaloids isolated from marine sponges in the order haplosclerida, in: Pelletier, S.W. (Ed.), Alkaloids: Chemical and Biological Perspectives. Pergamon, pp. 301–355. https://doi.org/10.1016/S0735-8210(96)80027-6
19. Cutignano, A., Tramice, A., Caro, S.D., Villani, G., Cimino, G., Fontana, A., 2003. Biogenesis of 3-alkylpyridine alkaloids in the marine mollusc *Haminoea orbignyana*. Angewandte Chemie 115, 2737–2740. https://doi.org/10.1002/ange.200250642
20. Cutignano, A., Cimino, G., Giordano, A., d'Ippolito, G., Fontana, A., 2004. Polyketide origin of 3-alkylpyridines in the marine mollusc *Haminoea orbignyana*. Tetrahedron Letters 45, 2627–2629. https://doi.org/10.1016/j.tetlet.2004.01.138
21. Lin, H., Kwan, A.L., Dutcher, S.K., 2010. Synthesizing and salvaging NAD^+: lessons learned from *Chlamydomonas reinhardtii*. PLOS Genetics 6, e1001105. https://doi.org/10.1371/journal.pgen.1001105
22. Bogan, K.L., Brenner, C., 2008. Nicotinic acid, nicotinamide, and nicotinamide riboside: a molecular evaluation of NAD^+ precursor vitamins in human nutrition. Annual Review of Nutrition 28, 115–130. https://doi.org/10.1146/annurev.nutr.28.061807.155443
23. Hara, N., Yamada, K., Shibata, T., Osago, H., Hashimoto, T., Tsuchiya, M., 2007. Elevation of cellular NAD levels by nicotinic acid and involvement of nicotinic acid phosphoribosyltransferase in human cells. Journal of Biological Chemistry 282, 24574–24582. https://doi.org/10.1074/jbc.M610357200
24. Vinitsky, A., Grubmeyer, C., 1993. A new paradigm for biochemical energy coupling. *Salmonella typhimurium* nicotinate phosphoribosyltransferase. Journal of Biological Chemistry 268, 26004–26010. https://doi.org/10.1016/S0021-9258(19)74485-8
25. Imai, T., Anderson, B.M., 1987. Metabolism of nicotinamide mononucleotide in beef liver. Archives of Biochemistry and Biophysics 254, 241–252. https://doi.org/10.1016/0003-9861(87)90100-7
26. Zheng, X.-Q., Matsui, A., Ashihara, H., 2008. Biosynthesis of trigonelline from nicotinate mononucleotide in mungbean seedlings. Phytochemistry 69, 390–395. https://doi.org/10.1016/j.phytochem.2007.08.008
27. Smith, S., Witkowski, A., Joshi, A.K., 2003. Structural and functional organization of the animal fatty acid synthase. Progress in Lipid Research 42, 289–317. https://doi.org/10.1016/S0163-7827(02)00067-X
28. Jenni, S., Leibundgut, M., Maier, T., Ban, N., 2006. Architecture of a fungal fatty acid synthase at 5 Å resolution. Science 311, 1263–1267. https://doi.org/10.1126/science.1123251
29. Schweizer, E., Hofmann, J., 2004. Microbial Type I Fatty Acid Synthases (FAS): Major Players in a Network of Cellular FAS Systems. Microbiology and Molecular Biology Reviews 68, 501–517. https://doi.org/10.1128/MMBR.68.3.501-517.2004
30. White, S.W., Zheng, J., Zhang, Y.-M., Rock, C.O., 2005. The structural biology of type ii fatty acid biosynthesis. Annual Reviews of Biochemistry 74, 791–831. https://doi.org/10.1146/annurev.biochem.74.082803.133524
31. Shen, B., 2003. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. Current Opinion in Chemical Biology 7, 285–295. https://doi.org/10.1016/S1367-5931(03)00020-6
32. Itoh, T., Tokunaga, K., Matsuda, Y., Fujii, I., Abe, I., Ebizuka, Y., Kushiro, T., 2010. Reconstitution of a fungal meroterpenoid biosynthesis reveals the involvement of a novel family of terpene cyclases. Nature Chemistry 2, 858–864. https://doi.org/10.1038/nchem.764

33. Yaegashi, J., Romsdahl, J., Chiang, Y.-M., Wang, C.C.C., 2015. Genome mining and molecular characterization of the biosynthetic gene cluster of a diterpenic meroterpenoid, 15-deoxyoxalicine B, in *Penicillium canescens*. Chemical Science 6, 6537–6544. https://doi.org/10.1039/C5SC01965F
34. Myronovskyi, M., Rosenkränzer, B., Nadmid, S., Pujic, P., Normand, P., Luzhetskyy, A., 2018. Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters. Metabolic Engineering 49, 316–324. https://doi.org/10.1016/j.ymben.2018.09.004
35. Ma, H.-M., Zhou, Q., Tang, Y.-M., Zhang, Z., Chen, Y.-S., He, H.-Y., Pan, H.-X., Tang, M.-C., Gao, J.-F., Zhao, S.-Y., Igarashi, Y., Tang, G.-L., 2013. Unconventional origin and hybrid system for construction of pyrrolopyrrole moiety in kosinostatin biosynthesis. Chemistry & Biology 20, 796–805. https://doi.org/10.1016/j.chembiol.2013.04.013
36. Xu, S., Brockmöller, T., Navarro-Quezada, A., Kuhl, H., Gase, K., Ling, Z., Zhou, W., Kreitzer, C., Stanke, M., Tang, H., Lyons, E., Pandey, P., Pandey, S.P., Timmermann, B., Gaquerel, E., Baldwin, I.T., 2017. Wild tobacco genomes reveal the evolution of nicotine biosynthesis. The Proceedings of the National Academy of Sciences 114, 6133–6138. https://doi.org/10.1073/pnas.1700073114
37. Kajikawa, M., Hirai, N., Hashimoto, T., 2009. A PIP-family protein is required for biosynthesis of tobacco alkaloids. Plant Molecular Biology 69, 287–298. https://doi.org/10.1007/s11103-008-9424-3
38. Kajikawa, M., Shoji, T., Kato, A., Hashimoto, T., 2011. Vacuole-localized berberine bridge enzyme-like proteins are required for a late step of nicotine biosynthesis in tobacco. Plant Physiology 155, 2010–2022. https://doi.org/10.1104/pp.110.170878
39. Turk, T., Frangež, R., Sepčić, K., 2007. Mechanisms of toxicity of 3-alkylpyridinium polymers from marine sponge *Reniera sarai*. Marine Drugs 5, 157–167. https://doi.org/10.3390/MD504157
40. Li, W., Zhang, F., Wu, R., Jia, L., Li, G., Guo, Y., Liu, C., Wang, G., 2017. A novel *N*-methyltransferase in *Arabidopsis* appears to feed a conserved pathway for nicotinate detoxification among land plants and is associated with lignin biosynthesis. Plant Physiology 174, 1492–1504. https://doi.org/10.1104/pp.17.00259
41. Mourabit, A.A., Pusset, M., Chtourou, M., Gaigne, C., Ahond, A., Poupat, C., Potier, P., 1997. Pyraxinine, a Novel Nitrogenous compound from the marine sponge *Cymbastela cantharella*. Journal of Natural Products 60, 290–291. https://doi.org/10.1021/np960562p
42. Jahn, T., König, G.M., Wright, A.D., Wörheide, G., Reitner, J., 1997. Manzacidin D: an unprecedented secondary metabolite from the "living fossil" sponge *Astrosclera willeyana*. Tetrahedron Letters 38, 3883–3884. https://doi.org/10.1016/S0040-4039(97)00846-0
43. Cafieri, F., Fattorusso, E., Taglialatela-Scafati, O., 1998. Novel betaines from the marine sponge A*gelas dispar*. Journal of Natural Products 61, 1171–1173. https://doi.org/10.1021/np980157t
44. Mohanty, I., Moore, S.G., Yi, D., Biggs, J.S., Gaul, D.A., Garg, N., Agarwal, V., 2020. Precursor-guided mining of marine sponge metabolomes lends insight into biosynthesis of pyrrole–imidazole alkaloids. ACS Chem. Biol. 15, 2185–2194. https://doi.org/10.1021/acschembio.0c00375
45. Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890. https://doi.org/10.1093/bioinformatics/bty560
46. Grabherr, M.G., Haas, 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., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-seq data. Nature Biotechnology 29, 644–652. https://doi.org/10.1038/nbt.1883
47. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A., 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 8, 1494–1512. https://doi.org/10.1038/nprot.2013.084

48. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of Molecular Biology 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
49. O'Donovan, C., Martin, M.J., Gattiker, A., Gasteiger, E., Bairoch, A., Apweiler, R., 2002. High-quality protein knowledge resource: SWISS-PROT and TrEMBL. Briefings in Bioinformatics 3, 275–284. https://doi.org/10.1093/bib/3.3.275
50. Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Research 39, W29–W37. https://doi.org/10.1093/nar/gkr367
51. Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J., Punta, M., 2014. Pfam: the protein families database. Nucleic Acids Research 42, D222–D230. https://doi.org/10.1093/nar/gkt1223
52. 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, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351
53. Lu, S., Wang, J., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Marchler, G.H., Song, J.S., Thanki, N., Yamashita, R.A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C.J., Marchler-Bauer, A., 2020. CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Research 48, D265–D268. https://doi.org/10.1093/nar/gkz991
54. Pruitt, K.D., Tatusova, T., Maglott, D.R., 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Research 35, D61–D65. https://doi.org/10.1093/nar/gkl842
55. Kanehisa, M., Sato, Y., Morishima, K., 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. Journal of Molecular Biology, Computation resources for molecular biology 428, 726–731. https://doi.org/10.1016/j.jmb.2015.11.006
56. The UniProt Consortium, 2021. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Research 49, D480–D489. https://doi.org/10.1093/nar/gkaa1100
57. Aguilar-Camacho, J.M., Doonan, L., McCormack, G.P., 2019. Evolution of the main skeleton-forming genes in sponges (phylum Porifera) with special focus on the marine Haplosclerida (class Demospongiae). Molecular Phylogenetics and Evolution 131, 245–253. https://doi.org/10.1016/j.ympev.2018.11.015
58. Dreyfus, M., Régnier, P., 2002. The Poly(A) Tail of mRNAs: Bodyguard in Eukaryotes, Scavenger in Bacteria. Cell 111, 611–613. https://doi.org/10.1016/S0092-8674(02)01137-6
59. Olarerin-George, A.O., Hogenesch, J.B., 2015. Assessing the prevalence of *Mycoplasma* contamination in cell culture *via* a survey of NCBI's RNA-seq archive. Nucleic Acids Research 43, 2535–2542. https://doi.org/10.1093/nar/gkv136
60. R. Lichman, B., 2021. The scaffold-forming steps of plant alkaloid biosynthesis. Natural Product Reports 38, 103–129. https://doi.org/10.1039/D0NP00031K
61. Daniel, B., Konrad, B., Toplak, M., Lahham, M., Messenlehner, J., Winkler, A., Macheroux, P., 2017. The family of berberine bridge enzyme-like enzymes: a treasure-trove of oxidative reactions. Archives of Biochemistry and Biophysics, Flavoproteins: beyond the classical paradigms 632, 88–103. https://doi.org/10.1016/j.abb.2017.06.023
62. Müller, W.E.G., Kasueske, M., Wang, X., Schröder, H.C., Wang, Y., Pisignano, D., Wiens, M., 2008. Luciferase a light source for the silica-based optical waveguides (spicules) in the demosponge *Suberites domuncula*. Cellular and Molecular Life Sciences 66, 537. https://doi.org/10.1007/s00018-008-8492-5
63. Gossmann, T.I., Ziegler, M., Puntervoll, P., Figueiredo, L.F. de, Schuster, S., Heiland, I., 2012. NAD^+ biosynthesis and salvage – a phylogenetic perspective. The Federation of European Biochemical Societies Journal 279, 3355–3363. https://doi.org/10.1111/j.1742-4658.2012.08559.x
64. Hunsucker, S.A., Spychala, J., Mitchell, B.S., 2001. Human cytosolic 5′-nucleotidase I: characterization and role in nucleoside analog resistance. Journal of Biological Chemistry 276, 10498–10504. https://doi.org/10.1074/jbc.M011218200

65. Bogan, K.L., Evans, C., Belenky, P., Song, P., Burant, C.F., Kennedy, R., Brenner, C., 2009. Identification of Isn1 and Sdt1 as glucose- and vitamin-regulated nicotinamide mononucleotide and nicotinic acid mononucleotide 5′-nucleotidases responsible for production of nicotinamide riboside and nicotinic acid riboside. Journal of Biological Chemistry 284, 34861–34869. https://doi.org/10.1074/jbc.M109.056689
66. Belenky, P., Christensen, K.C., Gazzaniga, F., Pletnev, A.A., Brenner, C., 2009. Nicotinamide riboside and nicotinic acid riboside salvage in fungi and mammals: quantitative basis for Urh1 and purine nucleoside phosphorylase function in NAD+ metabolism. Journal of Biological Chemistry 284, 158–164. https://doi.org/10.1074/jbc.M807976200
67. Kahn, C.R., 2000. Triglycerides and toggling the tummy. Nat Genet 25, 6–7. https://doi.org/10.1038/75610
68. Lykidis, A., Jackowski, S., 2000. Regulation of mammalian cell membrane biosynthesis, in: Progress in nucleic acid research and molecular biology. Academic Press, pp. 361–393. https://doi.org/10.1016/S0079-6603(00)65010-9
69. Inui, H., Miyatake, K., Nakano, Y., Kitaoka, S., 1984. Fatty acid synthesis in mitochondria of *Euglena gracilis*. European Journal of Biochemistry 142, 121–126. https://doi.org/10.1111/j.1432-1033.1984.tb08258.x
70. Houten, S.M., Wanders, R.J.A., 2010. A general introduction to the biochemistry of mitochondrial fatty acid β-oxidation. Journal of Inherited Metabolic Disease 33, 469–477. https://doi.org/10.1007/s10545-010-9061-2
71. Thurber, A.R., 2007. Diets of Antarctic sponges: links between the pelagic microbial loop and benthic metazoan food web. Marine Ecology Progress Series 351, 77–89. https://doi.org/10.3354/meps07122
72. Hochmuth, T., Niederkrüger, H., Gernert, C., Siegl, A., Taudien, S., Platzer, M., Crews, P., Hentschel, U., Piel, J., 2010. Linking chemical and microbial diversity in marine sponges: possible role for Poribacteria as producers of methyl-branched fatty acids. ChemBioChem 11, 2572–2578. https://doi.org/10.1002/cbic.201000510
73. Aiello, A., Fattorusso, E., Menna, M., 1999. Steroids from sponges: recent reports. Steroids 64, 687–714. https://doi.org/10.1016/S0039-128X(99)00032-X
74. Elenkov, I., Dragova, B., Andreev, S., Popov, S., 1997. 4α-Methyl sterols from the sponges *Haliclona cinerea* and *Haliclona flavescens*. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 118, 155–157. https://doi.org/10.1016/S0305-0491(97)00029-1
75. Findlay, J.A., Patil, A.D., 1985. Novel sterols from the finger sponge *Haliclona oculata*. Canadian Journal of Chemistry 63, 2406–2410. https://doi.org/10.1139/v85-398
76. Viegelmann, C., Parker, J., Ooi, T., Clements, C., Abbott, G., Young, L., Kennedy, J., Dobson, A.D.W., Edrada-Ebel, R., 2014. Isolation and identification of antitrypanosomal and antimycobacterial active steroids from the sponge *Haliclona simulans*. Marine Drugs 12, 2937–2952. https://doi.org/10.3390/md12052937
77. Silva, C.J., Wünsche, L., Djerassi, C., 1991. Biosynthetic studies of marine lipid 35. The demonstration of *de* novo sterol biosynthesis in sponges using radiolabeled isoprenoid precursors. Comparative Biochemistry and Physiology Part B: Comparative Biochemistry 99, 763–773. https://doi.org/10.1016/0305-0491(91)90140-9
78. Volkman, J., 2003. Sterols in microorganisms. Applied Microbiology and Biotechnology 60, 495–506. https://doi.org/10.1007/s00253-002-1172-8
79. Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M.E.A., Mitros, T., Richards, G.S., Conaco, C., Dacre, M., Hellsten, U., Larroux, C., Putnam, N.H., Stanke, M., Adamska, M., Darling, A., Degnan, S.M., Oakley, T.H., Plachetzki, D.C., Zhai, Y., Adamski, M., Calcino, A., Cummins, S.F., Goodstein, D.M., Harris, C., Jackson, D.J., Leys, S.P., Shu, S., Woodcroft, B.J., Vervoort, M., Kosik, K.S., Manning, G., Degnan, B.M., Rokhsar, D.S., 2010. The *Amphimedon queenslandica* genome and the evolution of animal complexity. Nature 466, 720–726. https://doi.org/10.1038/nature09201
80. Takishita, K., Takaki, Y., Chikaraishi, Y., Ikuta, T., Ozawa, G., Yoshida, T., Ohkouchi, N., Fujikura, K., 2017. Genomic evidence that methanotrophic endosymbionts likely provide deep-sea *Bathymodiolus* mussels

with a sterol intermediate in cholesterol biosynthesis. Genome Biology and Evolution 9, 1148–1160. https://doi.org/10.1093/gbe/evx082

81. Gallo, C., Landi, S., d'Ippolito, G., Nuzzo, G., Manzo, E., Sardo, A., Fontana, A., 2020. Diatoms synthesize sterols by inclusion of animal and fungal genes in the plant pathway. Scientific Reports 10, 4204. https://doi.org/10.1038/s41598-020-60993-5
82. Kerr, R.G., Kerr, S.L., Malik, S., Djerassi, C., 1992. Biosynthetic studies of marine lipids. 38. Mechanism and scope of sterol side chain dealkylation in sponges: evidence for concurrent alkylation and dealkylation. The Journal of the American Chemical Society 114, 299–303. https://doi.org/10.1021/ja00027a038
83. Tomazic, M.L., Najle, S.R., Nusblat, A.D., Uttaro, A.D., Nudel, C.B., 2011. A novel sterol desaturase-like protein promoting dealkylation of phytosterols in *Tetrahymena thermophila*. Eukaryotic Cell 10, 423–434. https://doi.org/10.1128/EC.00259-10

Chapter 3

Genome Sequencing, Assembly, and Mining of *Haliclona indistincta*

Part of this chapter contributed to the as of yet unpublished manuscript:
Sandoval, K., McCormack, G.P., 2022. The *Haliclona indistincta* Genome reveals a New Family of Metazoan Hybrid Nonribosomal Peptide Synthetase-Polyketide Synthase Genes.

Introduction

The completion of the human genome project was a tremendous achievement which required thirteen years of research and an investment of around $2.7 billion [1]. Such a feat was made possible by capillary-array electrophoresis using the Sanger method—a so-called first-generation technique [2]. Since this accomplishment, the rise of second-generation sequencing, also known as next-generation sequencing (NGS), has dramatically increased the availability of whole genome sequencing to laboratories worldwide by improving on its predecessors. In particular, technologies such as the sequencing by synthesis method of Illumina can now sequence a whole human genome in less than a day, for below $1,000 and with higher throughput [3]. The reads produced by this method possess a high sequencing accuracy rate of >99.5% but are limited in their potential lengths of up to 300 bp depending on the system used [4]. While cheap and accessible, there are drawbacks to this method. The short read length particularly hampers the ability to assemble contiguous genomes as the reads cannot bridge repetitive regions longer than the sequenced length. Furthermore, this technology is reliant on PCR amplification and thus can introduce biases in the sequenced reads.

These aforementioned issues were addressed with third-generation sequencing technology such as those of Pacific Biosciences and Oxford Nanopore Technologies [5,6]. With these methods, longer DNA molecules in the kilobase pair range are directly sequenced by the system, potentially resulting in reads significantly longer than that of Illumina. As a result, the issue posed by repetitive regions of a genome are addressed allowing for a more contiguous assembly. However, this long-read technology possesses significantly higher error rates of 75-90%. For this reason a number of techniques are often employed such as deep-sequencing with long-reads so that the errors self-correct upon assembly when the reads overlap, incorporating short-read data to improve upon the accuracy *via* methods such as polishing a draft assembly from long-reads, or even new sequencing methods such as the high fidelity reads of Pacific Biosciences [6]. While long-read sequencing improves the contiguity of downstream assemblies, it is insufficient for achieving scaffolds on the chromosome level. For this purpose, additional methods such as chromosome conformation capture which utilize the interaction between chromosomes to improve the scaffolding process have been developed [7].

The increased accessibility of genome sequencing combined with the small genomes of microorganisms has also revolutionized the field of drug discovery. Because biosynthetic genes associated with specialized metabolism are often clustered together in the genomes of bacteria and fungi, it is also possible to detect the patterns exhibited by these biosynthetic gene clusters (BGCs) [8]. For this reason the method of genome mining, or computationally screening the genome of an organism for BGCs, has become a widely used technique to study specialized metabolism [9,10]. Particularly, genes encoding for megasynthases such as

polyketide synthase (PKS), nonribosomal peptide synthetase (NRPS) and hybrid systems of the two (PKS-NRPS; NRPS-PKS) are readily detectable due to the utilization of several conserved domains which are often organized in a predictable fashion. As a result several software packages have been developed to take advantage of these features and identify known or novel BGCs in an automated manner [10-12]. Furthermore, when considering the megasynthase genes there have been efforts to computationally predict what types of substrates will be utilized by means such as phylogenetic analysis [13,14]. This is due to biosynthetic domains of a class, such as the aforementioned megasynthases, often falling into different clades depending on what substrates they utilize or reactions they catalyse.

Sponges and their microbial consortia represent one of the most promising sources of novel specialized metabolites for applied purposes. For this reason, genome sequencing and mining represents a promising method for identifying the biosynthetic origin of both known and novel bioactive compounds. All three generations of sequencing technology have been utilized to sequence the genomes of sponges. The first was in 2010 *Amphimedon queenslandica* was sequenced using Sanger technology which revealed insights into the origin and early evolution of animals [15]. In turn second-generation sequencing has been used to sequence the genomes of *Oscarella pearsei* (then *Oscarella carmela*), *Sycon ciliatum*, *Stylissa carteri*, *Xestospongia testudinaria*, *Tethya wilhelma* and *Lubomirskia baikalensis* [16-21]. The most well assembled sponge genome to date is that of *Ephydatia muelleri* [22]. By combining long-read Pacific Biosciences sequencing with Chicago and Hi-C data for scaffolding, a chromosome-level assembly was achieved. Furthermore, the genomes of *A. queenslandica* and *E. muelleri* were respectively derived from larvae and gemmules. As a result, these genomes are the most accurate representations of sponge genomes with limited contamination from associated microorganisms. However, to date no high quality genomes of sponges associated with bioactive specialized metabolites have been produced; while the sequenced *S. carteri* and *X. testudinaria* are known for their specialized metabolism their assemblies are highly fragmented and contaminated [23,24]. Rather, focus has been directed towards sequencing the microbial symbionts of sponges as they are often the true producers of sponge-associated bioactive compounds [25]. Due to many sponge-associated symbionts being unculturable, methods ranging from shotgun metagenomics to single-cell genomics have been utilized to understand their biosynthetic potential [26,27]. In the case of the former, in which a mixture of host sponge and microbial symbiont NGS data is mixed together, bioinformatics tools to bin out the microbial contigs are often necessary to acquire metagenome-assembled genomes (MAGs) [28].

Sponges of the Order Haplosclerida are particularly known for being the source of a unique class of compounds known as 3-alkylpyridines alkaloids (3-APs) [29]. Interestingly, these alkaloids appear to be largely unique to this Order which has resulted in ambiguity whether they are produced by microbial symbionts, like some characterized sponge-derived compounds, or are actually derived from the sponges themselves. To date, the only published

evidence is the reported inconsistent production of manzamine A, a 3-AP derivative, from a *Micromonospora* sp. isolated from *Acanthostrongylophora ingens* [30]. In particular, the biosynthetic origin of the 3-APs could be the result of an AMP-binding enzyme, such as a CoA-ligase or NRPS adenylation domain, activating nicotinic acid (NTA) for polyketide biosynthesis (Figure 1.11). Such a mechanism is the likely way in which the similar, linear, monomeric haminol alkaloids are synthesized by the mollusc *Haminoea orbignyana* [31,32]. The incorporation of NTA into polyketide biosynthesis has also been proven on the genetic level, albeit only in bacteria and fungi. Specifically, the biosynthesis of pyripyroene A from *Aspergillus fumigatus* strain Af293, 5-deoxyoxalicine B from *Penicillium canescens* strain ATCC 10419 and pyridinopyrone A by the *Streptomyces albus* J1074 appear to involve a CoA-ligase (Figure 1.11, Figure 1.12) [33-35]. Alternatively, the kosinostatin biosynthetic pathway possesses a standalone adenylation domain capable of promiscuously utilizing NTA and represents an alternative way in which this building block could hypothetically be activated for incorporation into a polyketide [36]. Overall, there is a precedence for animals, fungi, and bacteria to perform the enzymatic steps hypothesized to be responsible for 3-AP biosynthesis.

The native Irish sponges *Haliclona indistincta* and *Haliclona viscosa* are sister species with both being a known source of 3-AP polymers. In turn they fall within the so-called Clade C in which most 3-AP-associated haplosclerids belong (Grace McCormack, personal communication). As they are both low-microbial abundance sponges, they represent excellent targets for genome sequencing and mining to identify the biosynthetic origin of 3-APs as the number of potential producing organisms is greatly reduced [37]. However, these species also pose several challenges as targets for genome sequencing. For one, upon being damaged they secrete a detergent-like substance which can potentially complicate the process of extracting clean, high-molecular-weight DNA (personal observation). Furthermore, specialized metabolites such as the very 3-APs which make these interesting targets for biodiscovery can be co-purified with DNA and interfere with the sequencing process (personal observation). Finally, despite its low-microbial-abundance status, *H. indistincta* possesses a diverse community of epibiotic organisms living in association with it such as algae, crustaceans and polychaetes which can complicate the acquisition of non-contaminated DNA (personal observation).

With the specific intention of identifying genes responsible for the biosynthesis of 3-APs, an assembled (meta)genome of a sponge associated with these compounds offers several advantages over the previously produced transcriptomes which utilized poly(A) selection and short-read sequencing (Chapter 2). First, it does not exclude associated microorganisms which could be the true producers of these compounds. Second, the influence of gene expression is not a potential issue. Third, technologies such as long-read sequencing can better capture the entire length of polyketide synthase genes. Fourth, the assembly of genomic contigs allows for the identification of BGCs which may encode for different

enzymes involved in the same biosynthetic process. With these qualities in mind the primary objective of using of both second- and third-generation sequencing methods to produce contiguous genomes of *H. indistincta*, *H. viscosa*, and their associated microorganisms was pursued. In turn, the complimentary aim of applying genome mining techniques to identify PKS genes possessing characteristics associated with the biosynthesis of 3-APs could then be attempted. Finally, functional prediction of identified PKS genes *via* computational tools and manual analysis of residues could then be pursued in order to estimate their role in the biosynthesis of 3-APs.

Materials and Methods

Sample Collection

Haliclona indistincta (MIIG1388) (Figure 2.2A) was collected at Corranroo on 17/05/2019. *Haliclona viscosa* (MIIG1389) (Figure 2.2B) was collected at Bridges of Ross on 01/08/2019. Both specimens were identified by Grace P. McCormack. Visible epibionts were removed. The sponges were rinsed in sterile artificial seawater, dissected into ~2cm^3 pieces and flash-frozen with liquid nitrogen. MIIG1388 and MIIG1389 samples were stored at -70 °C until further use. Voucher specimens of both sponges were stored in ethanol.

DNA Extraction

A ~1 cm^3 tissue sample of MIIG1388 was ground to a powder using an autoclaved mortar and pestle. The tissue was transferred to a sterile 50 mL centrifuge tube containing 10 mL of Carlson Lysis Buffer (100 mM Tris pH 9.5; 2% cetyltrimethylammonium bromide; 1.4 M sodium chloride; 1% polyethylene glycol 6000; 20 mM ethylenediaminetetraacetic acid (EDTA). 25 µL of β-mercaptoethanol was added to the solution. The sample was incubated in a 65 °C water bath for 2 hrs with gentle swirling every 30 min. After incubation the sample was allowed to cool to room temperature. The cell lysate was extracted with an equal volume of 25:24:1 phenol:chloroform:isoamyl alcohol (P:C:IAA) (pH 8.0). The extraction was gently inverted five times and then centrifuged at 5,000 × g for 10 min at 4 °C. The upper aqueous phase was transported to a new 50 mL centrifuge tube and the extraction process was repeated once more. The aqueous layer was transported to a new 50 mL centrifuge tube and gently mixed with 2 mL of 3 M sodium acetate. The sample was then gently mixed with 22 mL of isopropanol and allowed to precipitate at 4 °C overnight.

The sample was centrifuged at 4,500 × g for 90 min at 4 °C. The supernatant was discarded, and the pellet washed with 40 mL 70% ice cold ethanol. The sample was once again centrifuged at 4,500 × g for 90 min at 4 °C. The supernatant was discarded, and the DNA pellet was allowed to air dry until there was no residual ethanol in the tube. The DNA pellet was gently resuspended in 10 mL of QIAGEN buffer G2. 20 µL of 100 mg/mL RNase A was mixed into the suspension. The sample was allowed to incubate for 5 min at room temperature. 100 µL of 20 mg/mL proteinase K was then gently mixed into the sample. The

sample was incubated at 50 °C for 1 hr and then allowed to cool to room temperature. While the sample cooled a QIAGEN Genomic Tip 100/G column was calibrated with 10 mL of QIAGEN buffer QBT. The sample was then placed into the Genomic Tip and allowed to flow through the column *via* gravity. Following this the sample was washed three times with 15 mL of QIAGEN wash buffer QC. Finally, the DNA sample was eluted with 15 mL of QIAGEN elution buffer QF which was preheated to 50 °C. 10.5 mL of isopropanol was added to the eluted DNA which was gently swirled and allowed to precipitate at 4 °C overnight.

The eluted DNA was centrifuged at >4,500 × g for 90 min at 4 °C. The supernatant was discarded, and the DNA pellet was rinsed with 40 mL of ice cold 70% ethanol. The sample was centrifuged at >4,500 × g for 30 min at 4 °C. The supernatant was discarded, and the DNA pellet was allowed to air dry until no ethanol remained. 500 µL to 1 mL of tris-EDTA (Ph 8.0) + 100 mM sodium chloride was added to the pellet. The pellet was placed in a 37 °C incubator with gentle swirling overnight. The DNA was quantified with both Nanodrop and Qubit systems. DNA length was estimated by running on a gel electrophoresis alongside a 20 kbp ladder. The DNA was separated into microcentrifuge tubes with aliquots containing 50 µg DNA per tube and stored at -70 °C until further use. An identical protocol was used for MIIG1389, albeit the pH of 25:24:1 P:C:IAA was 6.0.

Next-Generation Sequencing

DNA samples were sent to the University of Maryland Institute for Genome Sciences for sequencing. MIIG1388 was sequenced with one Pacific Biosciences Sequel ii single-molecule real-time sequencing cell. Chimeric PacBio reads were split into proper subreads using the pbclip tool [38]. MIIG1388 was additionally sequenced on one-fourth of an Illumina NovaSeq 6000 S1 Sequencing Lane using a 150 bp PE library. MIIG1389 was sequenced using Illumina technology in the same manner. Raw Illumina reads were quality controlled using fastp version 0.2 on default settings [39].

Genome Assembly

Three different *de novo* genome assemblers capable of utilizing Pacific Biosciences data were used and compared for assembly of the *H. indistincta* genome: Raven version 1.1.5 [40], Flye version 2.8.1 [41] and Canu version 2.0 [42]. Genome size was designated as 250 Mb if requested by an assembly tool, which is a rough estimate between the size of *A. queenslandica* and *E. muelleri* [15,22]. The quality and completeness of the assembled genomes was assessed using BUSCO version 5.1.2; specifically, the assemblies were queried against the latest version of the eukaryota_odb10 dataset (downloaded 13/04/2021) [43]. After comparing assemblers, a single assembly was then chosen for further improvement and analysis. Polishing of the assembly with both short and long reads was achieved by using HyPo version 1.0.3 [44]. The BUSCO analysis was then repeated to check whether polishing improved the completeness of the assembly. Illumina data of MIIG1389 was assembled with SPAdes version 3.14.0 on default parameters [45]. Genome assembly and sequential

analyses were performed with accounts at the Leibniz Supercomputing Centre and the Irish Centre for High-End Computing.

Quality Control

Due to their filter feeding nature sponges are often in close association with microorganisms such as bacteria. Thus, it is practically impossible to get a sample of adult tissue that is free of microbial contamination. For this reason, the MIIG1388 assembly was processed with a variety of tools to remove possible contaminating contigs not of sponge origin. Coverage, GC content and taxonomic assignment of contigs from the genome assembly were determined using BlobTools version 2.2.0 [46]. Binning of the contigs was then performed using MetaBAT version 2.12.1 with default parameters [47]. Bins containing contigs taxonomically assigned as being Porifera were retained as being part of the *H. indistincta* genome. Bins containing contigs taxonomically assigned as being other eukaryotes were also included as being part of the *H. indistincta* genome. This was decided as contigs within these bins not assigned as Porifera, but as other eukaryotes, contained open reading frames whose translated proteins had significant matches to sequences from the *A. queenslandica* proteome (data not shown). Any other bins which were clearly of bacterial origin were kept separate from the final assembly. The completeness of the binned *H. indistincta* genome was again assessed using BUSCO as described prior. The completeness and contamination of the separated bacterial bins was assessed using CheckM [48]. The 5S, 16S and 28S ribosomal subunits of binned bacteria were extracted using RNAmmer version 1.2 [49].

Genome Masking and Annotation

Repeat libraries were produced from the *H. indistincta* assembly using RepeatModeler version 2.0.1 with the "LTRStruct" parameter [50]. These libraries were then utilized by RepeatMasker version 4.1.1 to soft-mask the genome (RepeatMasker at https://repeatmasker.org). Protein-coding regions were identified with the BRAKER pipeline version 2.1.5 using complimentary RNA-seq data [51]. Predicted proteins were then annotated with EggNOG-mapper version 2.1.0 using default settings and version 5.0.2 of the EggNOG database [52,53].

Identification of Novel Biosynthetic Genes

The final genome assembly of MIIG1388 was screened for BGCs using a local installation of antiSMASH version 5.1.2 under two different settings [54]. The first run was set for bacterial taxa utilizing the prodigal gene-finding tool. The second run was set for fungal taxa utilizing the cassis option and the glimmerhmm gene-finding tool. The metazoan origin of biosynthetic genes was assessed *via* mapping of raw reads in two ways. First, long DNA reads greater than 1 kbp in length were mapped back to the MIIG1388 contig containing a biosynthetic gene cluster using minimap2 and the bamutils module of NGSUtils to assess if an even level of coverage existed across biosynthetic genes of interest and surrounding

metazoan genes [55,56]. Second, short RNA-seq reads produced from poly(A) selected libraries were mapped back to the aforementioned contig using GSNAP (Chapter 2) [57]. Mapped reads and coverage were visualized using IGV version 2.8.0 [58]. In addition, the annotated final assembly was also analysed by plantiSMASH version 1.0.0 to possibly detect biosynthetic genes not identified by the antiSMASH algorithm [59]. Enzymatic domains of translated proteins from biosynthetic genes of interest were analysed using both the NCBI conserved domain database and the Pfam database [60,61].

The predicted proteins of the MIIG1388 assembly were also screened for novel multi-domain enzymes, such as megasynthases like PKS, which may be involved in specialized metabolism as such enzymes are often the result of gene fusion. The Pfam domains of predicted proteins were first identified with a local installation of Pfam-scan version 1.6 [62]. The output was then used with the CO-ED tool to generate a network of Pfam domains with nodes representing detected domains and edges representing proteins containing co-occurrences of these domains [63]. The network was then visualized in Cytoscape version 3.8.0 [64].

Identification of Novel Sponge Megasynthases in Publicly Available Data

Publicly available sponge genome [15-22] and transcriptome [17,21,65-85] assemblies were screened for similar megasynthase enzymes which were found in the MIIG1388 assembly using tblastn [86]. If an assembly was not available, transcriptomes were assembled with Trinity version 2.8.5 on default settings [87]. The raw Illumina data for *S. carteri* was also reassembled using SPAdes version 3.14.0 on default parameters [45]. A contiguous sequence encoding for a megasynthase of interest from *S. carteri* was then acquired by manipulating the graphical fragment assembly file with the software Bandage [88].

Predicted Functional Characterization of Novel Sponge Megasynthases

The functional characterization of identified condensation and ketosynthase domains was predicted using NaPDoS [13]. The substrate specificity of the adenylation domains was predicted using AdenylPred [14]The four AMP-binding enzymes associated with preparing NTA for polyketide biosynthesis were also analysed by antiSMASH features and AdenylPred for comparison with any biosynthetic genes of interest found in MIIG1388 [33-36]. Identified biosynthetic proteins of interest were also queried against the Pfam database using hmmscan version 3.2.1 [89]. The Stachelhaus codes, the conserved residues of AMP-binding enzymes associated with substrate selectivity, of all identified adenylation domains were also manually assessed to estimate what substrates they utilize [90,91].

Phylogenetic Analysis of the Ketosynthase Domain

Ketosynthase domains were extracted from the chosen megasynthases using NaPDoS [13]. The most similar bacterial, fungal, and non-fungal eukaryotic sequences from the NCBI non-redundant database as well as the most similar bacterial sequences from the MIBiG database were then extracted for phylogenetic analysis (Appendix 2) [92,93]. Before extraction of the

ketosynthase region from these additional sequences it was ensured with NaPDoS that all chosen ketosynthase sequences were of the same type as those from the sponges [13]. All alignments were performed using MAFFT 7.490, with the L-INS-i alignment method using default settings [94]. Maximum likelihood trees were constructed in IQ-TREE version 2.1.4 with 1000 bootstrap pseudoreplicates [95]. The best model was automatically detected by IQ-TREE to be Q.Pfam+R5. The resulting phylogenetic trees were modified using the Interactive Tree of Life (iTOL) v6 [96].

Results

Genome Sequencing

A total of 46.7 µg of DNA with around 50% of the fragments being greater than 15 kbp were isolated from MIIG1388 and sent for long-read sequencing. Approximately 60.22 Gbp comprising 6.24 M reads (both base pair and read number values rounded up here and for every instance hereafter) were sequenced with the Pacific Biosciences Sequel II system. Despite the reported size selection and construction of a library with an average length of around 18 kbp by the sequencing company, the mean length and N50 of the raw reads were 8.34 kbp and 14.19 kbp respectively (Appendix 3). After filtering with pbclip the read statistics lowered somewhat with a total of 6.79 M reads, a total of 52.88 Gbp, a mean length of 7.79 kbp, and an N50 of 13.31 kbp. In addition, a total of 212.65 M 151 bp reads were sequenced with Illumina NovaSeq 6000 resulting in a total of 31.27 Gbp after filtering with fastp. A total of 26.2 µg of DNA was isolated from MIIG1389. However, the DNA was degraded and around 10 kbp in length. For that reason, this sample was not sequenced on the Sequel II system. A total of 167.03 M 151 bp reads were sequenced with Illumina NovaSeq 6000 resulting in a total of 24.61 Gbp after filtering with fastp.

Genome Assembly, Binning and Annotation

Three different assemblies of differing quality were acquired using Pacific Biosciences long read data (Table 3.1). In comparison to Flye and Raven, Canu produced a larger assembly with a higher N50 value and complete BUSCO score. The Canu assembly also exhibited a higher BUSCO duplication rate in comparison to Raven and Flye. Because long-read sequencing was unfeasible for MIIG1389, its Illumina data was assembled alone using SPAdes; statistics of the assembly included a total size of 326.90 Mbp, 932,573 contigs, an N50 of 99.96 kbp, and a GC content of 42.4%.

Based on its higher N50 value and BUSCO score the Canu assembly was chosen for further processing and analysis. In addition, another factor which influenced this decision was an initial screening of the Canu draft assembly using AntiSMASH which indicated that a contig which contained a biosynthetic gene cluster of interest was better assembled in comparison to those of Raven and Flye. Polishing of the Canu assembly using HyPo had minimal effect on the genome completeness as indicated by the corresponding BUSCO score. The BlobTools2

pipeline calculated a GC content of 41.59% for the total assembly and assigned 66.19% of the total assembly as being of sponge origin (Figure 3.1). Additionally, 17.36% of the assembly was assigned as being of chordate origin. Analysing the genes of several contigs assigned as Chordata with blast indicated sponge origin; for this reason, these contigs were assumed to originate from *H. indistincta*. Overall, at least 83.54% of the polished Canu assembly is assumed to be of sponge origin.

Table 3.1 Statistics of three different de novo long-read assemblers for MIIG1388.

	Assembler	**Raven**	**Flye**	**Canu**	**Canu + HyPo**	**Canu + HyPo + Metabat2**
Assembly Stats	Sum (Mbp)	222.00	244.77	323.22	323.23	295.64
	average (kbp)	150.41	108.35	114.05	114.05	200.57
	N50 (kbp)	202.99	293.62	370.41	370.42	418.01
	N contigs	1476	2259	2834	2834	1474
	GC content (%)	41.21	41.64	41.71	41.71	41.69
BUSCO (eukaryota)	Complete (%)	85.50	85.40	88.30	88.70	87.80
	Single (%)	77.30	72.90	21.20	22.40	23.90
	Duplicate (%)	8.20	12.50	67.10	66.30	63.90
	Fragmented (%)	8.20	9.80	7.50	7.10	7.80
	Missing (%)	6.30	4.80	4.20	4.20	4.40
	N BUSCO	255	255	255	255	255

A total of 38 bins were produced using MetaBat2 on the polished Canu assembly. By comparing the contigs of each bin with their respective taxonomic identification by BlobTools2, it was determined that 25 bins possessed contigs of sponge origin. These contigs were assumed to be derived from *H. indistincta* rather than any possible contaminants. An additional four bins were identified as having contigs of eukaryotic origin not assigned as Porifera by BlobTools2 but possessing genes which most closely aligned to those from *A. queenslandica*. Therefore, a total of 29 bins were chosen to encompass the final assembly of *H. indistincta*. This binned assembly has an approximate size of 295.64 Mbp, an N50 value of 418.01 kbp and consists of 1,474 contigs (Table 3.1). The complete BUSCO score also slightly decreased from 88.7 to 87.8% after binning. Approximately 52.77% of the binned MIIG1388 assembly was masked using RepeatModeler and RepeatMasker. In turn, the Braker2 pipeline identified 42,428 protein-coding genes. Eggnog-Mapper was then able to assign 38,510 of these proteins to different Cluster of Orthologous Groups (COGs) categories (Figure 3.2). COGs associated with RNA processing and modification (14.71%), energy production and conversion (6.16%), cell cycle control, cell division, and chromosome partitioning (7.97%), cell motility (10.17%), inorganic ion transport and metabolism (8.45%), signal transduction

mechanisms (6.78%) and general function prediction (8.27%) represented the largest proportions of assigned COGS. In contrast, only 0.13% of the total COGs were assigned as being associated with specialized (secondary) metabolism.

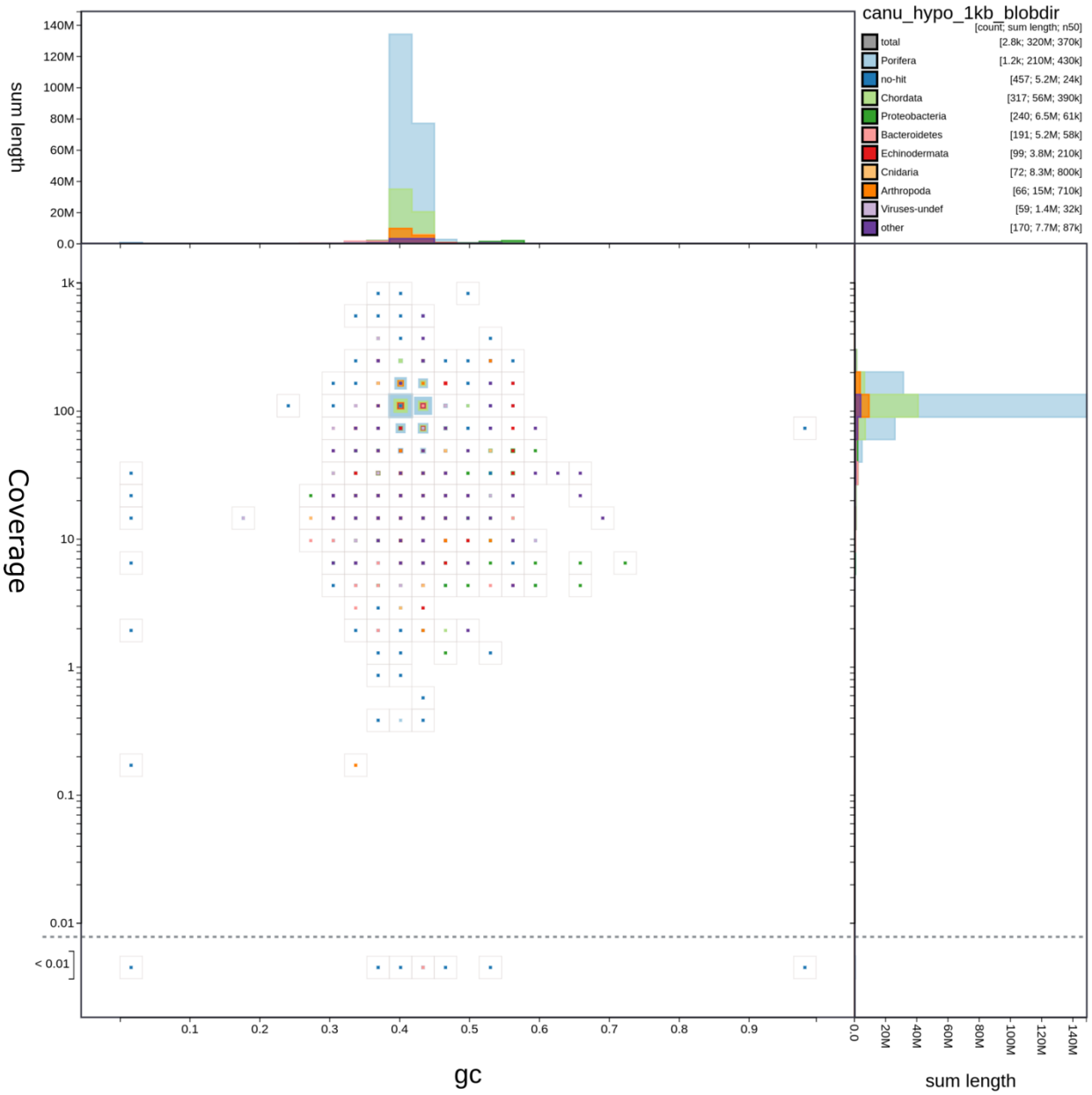


Figure 3.1 BlobTools2 scatter plot histogram of GC coverage and GC content of the contigs from the MIIG1388 Canu assembly with 1 kbp polishing using HyPo. Coverage is derived from mapping the raw reads of 1kb or greater length back to the polished assembly. GC represents the proportion of G and C bases. Sum length represents the cumulative length of contigs associated with a section of either axis.

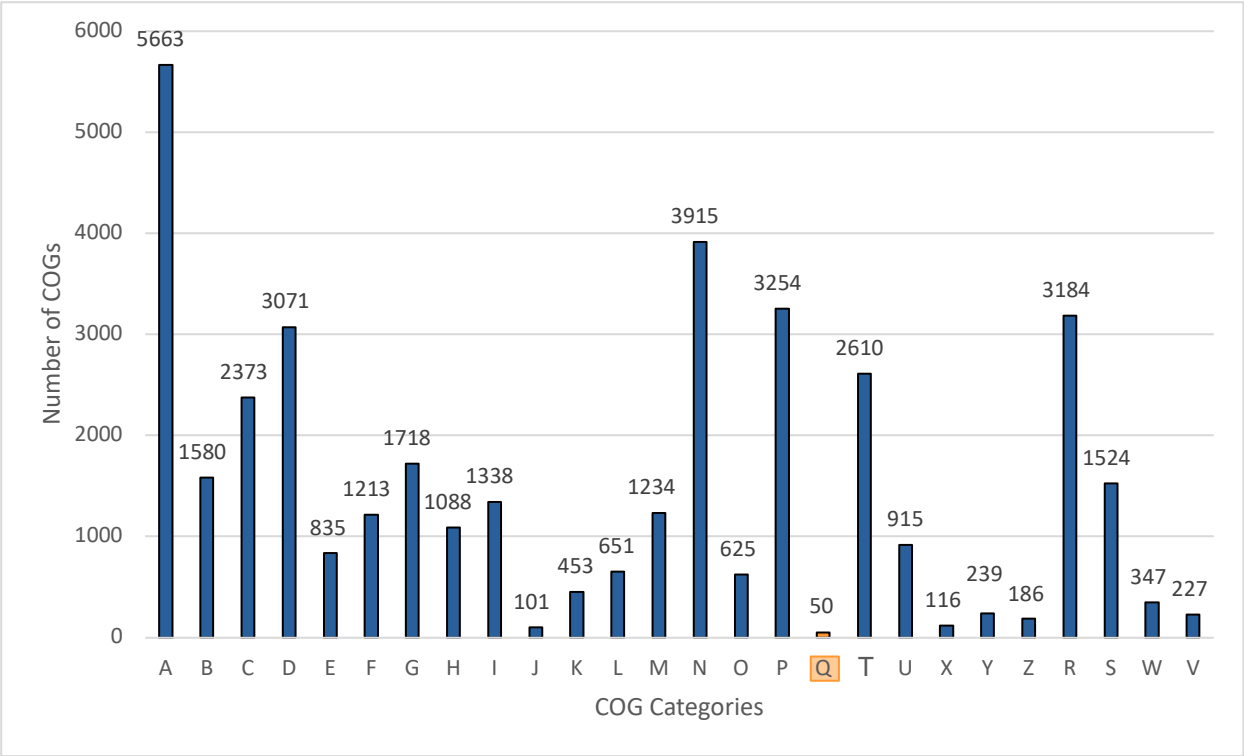


Figure 3.2 Number of genes from the *H. indistincta* genome which fall within various COG categories. COG categories are as follows: **A**, RNA Processing and Modification; **B**, Chromatin Structure and Dynamics; **C**, Energy Production and Conversion; **D**, Cell Cycle Control, Cell Division, Chromosome Partitioning; **E**, Amino Acid Transport and Metabolism; **F**, Nucleotide Transport and Metabolism; **G**, Carbohydrate Transport and Metabolism; **H**, Coenzyme Transport and Metabolism; **I**, Lipid Transport and Metabolism; **J**, Translation, Ribosomal Structure and Biogenesis; **K**, Transcription; **L**, Replication, Recombination and Repair; **M**, Cell Wall/Membrane/Envelope Biogenesis; **N**, Cell Motility; **O**, Posttranslational Modification, Protein Turnover, Chaperones; **P**, Inorganic Ion Transport and Metabolism; **Q**, Secondary Metabolites Biosynthesis, Transport and Catabolism; **R**, General Function Prediction Only; **S**, Function Unknown; **T**, Signal Transduction Mechanisms; **U**, Intracellular Trafficking, Secretion, and Vesicular Transport; **V**, Defence Mechanisms; **W**, Extracellular Structures; **X**, Mobilome, Prophages, Transposons; **Y**, Nuclear Structure; **Z** Cytoskeleton. COG category Q is indicated by the colour orange.

A total of 4 bacterial bins were separately acquired using MetaBat2 although bin 3 was too incomplete for further analysis (Table 3.2). CheckM indicated that bins 8 and 12 were of high completeness, while 12 and 28 had little contamination. The strain heterogeneity, which is the proportion of contamination derived from the same or similar strains, was 100% for Bins 8 and 28 whereas bin 12 was 50%. The 5S, 16S and 23S ribosomal subunit genes were extracted from the other three for blastp analysis against the NCBI nt database (Table 3.3). The sequences of bin 8 had relatively low percent identity in comparison to those of the other two bins and overall appeared to be a type of gammaproteobacteria. Bin 12 appears to be of the CFB group bacteria with high percent identity to a Flavobacteria. Bin 28 exhibited high similarity to members of the Order Rhodobacterales such as *Planktomarina temperata* RCA23.

Table 3.2 CheckM statistics of bacterial bins acquired from the MIIG1388 assembly using MetaBat2.

	Bin	**3**	**8**	**12**	**28**
MetaBat2	Sum (bp)	314810	2520641	1517996	2314170
	average (bp)	10493.67	38779.09	79894.53	100616.09
	N50 (bp)	14588	56002	125674	132681
	N contigs	30	65	19	23
CheckM	Completeness (%)	21.30	96.55	76.07	93.23
	Contamination (%)	0.00	28.29	0.66	0.30
	Strain heterogeneity (%)	0.00	100.00	50.00	100.00

Table 3.3 blastn results of ribosomal sequences from bacterial bins against the NCBI nonredundant database.

Bin	**Query**	**Top Hit**	**Max Score**	**% ID**	**E-value**	**Accession**
8	5S	*Pseudomonas syringae* pv. tomato	169	93.81	4e-38	CP047072.1
8	16S	gamma proteobacterium EHK-1	1788	88.38	0.0	AF228694.1
8	23S	*Microbulbife*r sp. GL-2	3247	87.12	0.0	AP019807.1
12	5S	*Cellulophaga* sp. L1A9	150	90.99	1e-32	CP047027.1
12	16S	Flavobacteria bacterium Yb001	2571	99.16	0.0	AB496658.1
12	23S	*Cellulophaga baltica* NN016038	4189	93.42	0.0	CP009887.1
28	5S	*Planktomarina temperata* RCA23	211	100	7e-51	CP003984.1
28	16S	Rhodobacteraceae bacterium FZCC0040	2684	100	0.0	MK335923.1
28	23S	*Planktomarina temperata* RCA23	5371	99.97	0.0	CP003984.1

Identification of Biosynthetic Genes

AntiSMASH analysis of the Canu-assembled draft genome before binning revealed the presence of six BGCs which appeared to be of bacterial origin based on the lack of introns as well homology to bacterial entries in the NCBI non-redundant database. Three encoded for terpene biosynthesis but were not included in any of the major bins produced by MetaBat2. Two additional BGCs were categorized as being responsible for terpene production. One terpene BGC belonged to bacterial bin 12 with 75% similarity to the MIBiG entry for carotenoid biosynthesis of *Arenibacter nanhaiticus* strain CGMCC 1.8863. A second terpene BGC belonged to bacterial bin 28 with 100% similarity to carotenoid biosynthesis of *P. temperata* RCA23. In addition, one BGC was categorized as being responsible for the production of a ribosomally synthesized and post-translationally modified peptide (RiPP) and belonged to the bacteria of bin 8 with highest similarity to a MIBiG entry for bacteriocin biosynthesis from *Sideroxydans lithotrophicus* ES-1. The genes of this BGC showed low similarity when queried against the NCBI nr database and no similar BGC was identified in the MIIG1389 assembly.

A seventh contig containing a gene coding for an NRPS-PKS was also detected. Unlike the aforementioned BGCs, this NRPS-PKS appeared to be encoded on a single ORF. Furthermore, it was located on a 215.17 kbp contig in which the surrounding genes were of putative sponge origin. This contig was classified as being of sponge origin by BlobTools2 and placed in a bin containing other contigs of sponge origin by MetaBat2. The gene for this hybrid NRPS-PKS was also detected in Flye and Raven assemblies, albeit on a shorter contig.

Plantismash analysis on the binned MIIG1388 assembly resulted in five additional BGCs being identified. Two identical clusters were of the putative, or unclassified, type and consisted of one methyltransferase gene and four AMP-binding genes. Two more identical clusters were classified as saccharide biosynthesis and contained genes encoding for glycosyltransferase and amino oxidase. A third saccharide BGC was also identified and included glycoslytransferase, epimerase and dioxygenase genes. All BGCs identified by plantiSMASH appeared to be of metazoan origin based on using their translated sequences as a blastp query against the nr database.

CO-ED analysis identified approximately 278 instances in which two enzymatic domains co-occurred within a single protein (Figure 3.3). Correspondingly a total of 251 multi-domain proteins were identified. With the UniProt, Brenda and MIBiG database respectively being represented by “U”, “B” and “M”, the co-occurrences can be classified as existing in the following databases: B = 6, M = 5, U = 2, U+B = 114, U+M = 6 and U+M+B = 22. A total of 123 co-occurrences were not determined to occur in any of these three databases by CO-ED. The aforementioned NRPS-PKS protein was also clearly identified by the CO-ED analysis. The majority of the enzymatic domain co-occurrences of this translated protein existed in the three aforementioned databases. However, one exception is that this NRPS-PKS contains both the formyl transferase and acyl transferase domains; according to CO-ED this is a novel co-occurrence not found in the UniProt, Brenda or MIBiG databases. No other multi-domain enzymes which obviously corresponded to the hypothetical polyketide origin of the 3-APs were detected with CO-ED.

The *H. indistincta* hybrid NRPS-PKS gene was chosen for further analysis due to two qualities. First, based on the hypothesis that 3-APs are produced *via* a polyketide synthase which accepts NTA as a starter unit by a feature such as an adenylation domain, this system appeared to be a possible candidate for how these compounds may be created in the *H. indistincta* holobiome. Second, this system represented the first identified instance of a megasynthase being derived from the host sponge itself rather than a microbial symbiont and thus was of high novelty.

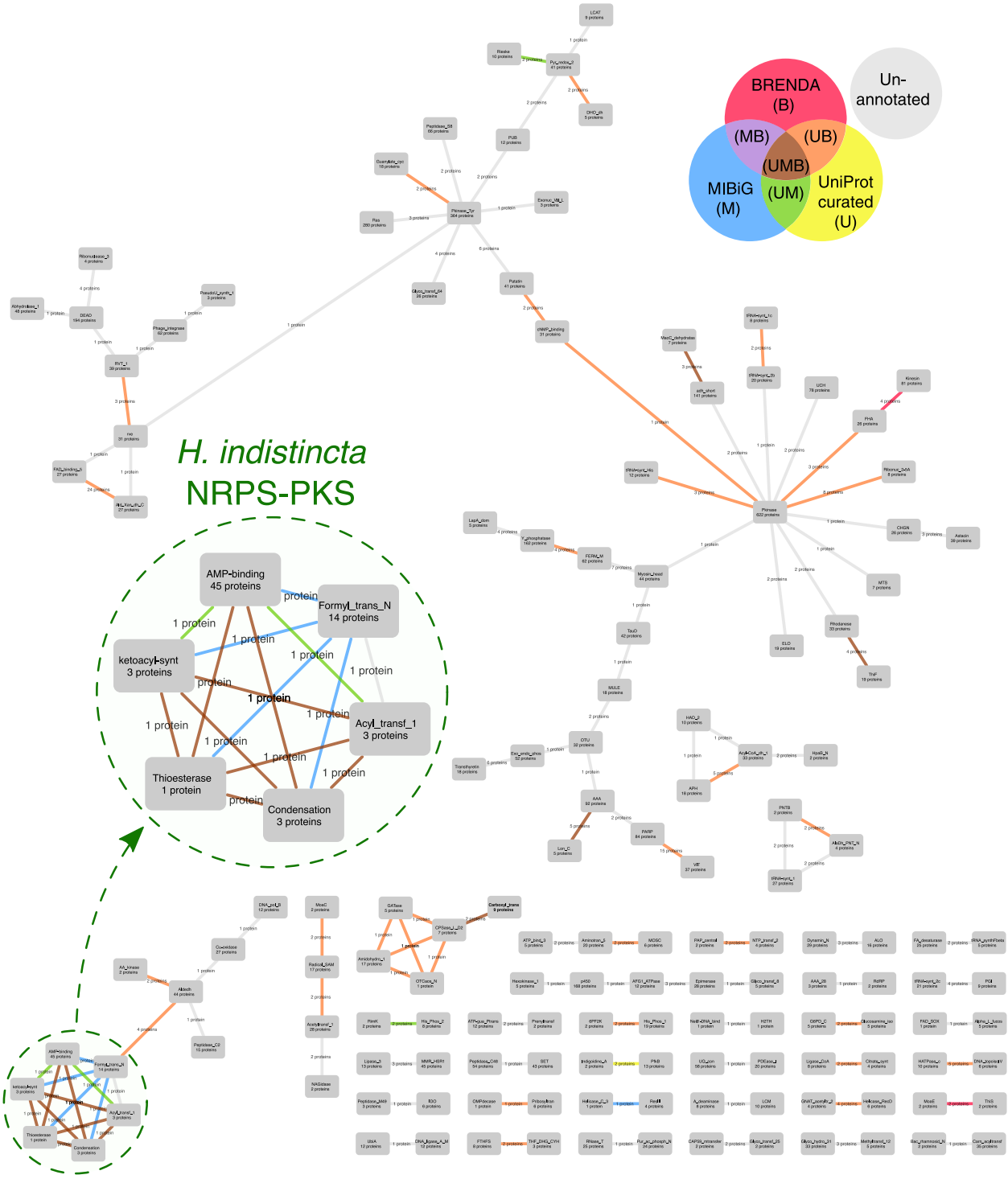


Figure 3.3 Enzymatic domain co-occurrences detected by CO-ED. The *H. indistincta* NRPS-PKS is highlighted in green. A zoomed in view of the *H. indistincta* NRPS-PKS is indicated by the arrow. Nodes represent different types of enzymatic domains detected in the *H. indistincta* proteome. Edges represent instances in which a protein contained two types of enzymatic domains. Whether the detected enzymatic domain co-occurrences are present in entries from the curated UniProt, BRENDA and MIBiG databases are coloured in the same manner as the Venn diagram.

Characterization of a Novel Metazoan NRPS-PKS

Submission of the translated amino acid sequence of the MIIG1388 NRPS-PKS for analysis against the Pfam database revealed the presence of two NRPS and one PKS modules (Figure 3.4). Furthermore, Pfam identified an N-terminal formylation domain not detected by antiSMASH. Pfam detected matches for each of the NRPS-PKS enzymatic domains with significantly low E-values save for the exception of the phosphopantetheine attachment site of the second module (Table 3.4). Based on typical PKS and NRPS proteins, it would be expected for this domain to be present on each module. However, Pfam did not detect such a domain on the PKS module.

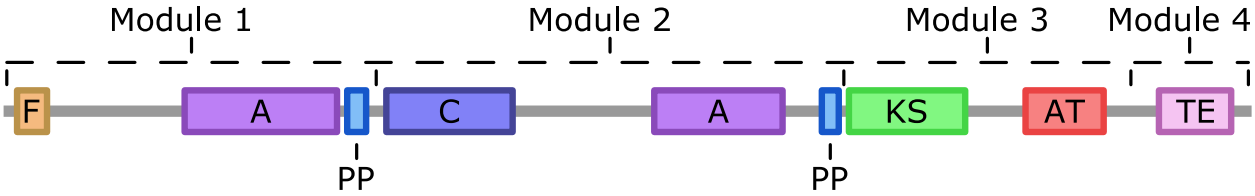


Figure 3.4 Domain organization of the hybrid NRPS-PKS identified in *H. indistincta*. Domain naming is as follows: F, formylation; A, adenylation; C, condensation; PP, phosphopantetheine attachment site; KS, ketosynthase; AT, acyltransferase; TE, thioesterase. Modules 1 and 2 correspond to the NRPS portion. Module 3 corresponds to the PKS portion. Module 4 corresponds to the termination domain.

Table 3.4 Pfam domains detected in the *H. indistincta* NRPS-PKS.

Domain	Pfam #	Clan	Start	End	Bit Score	E-value
Formylation	PF00551.19	n/a	66	167	53.7	2.10e-14
Adenylation	PF00501.28	CL0378	595	993	208.2	1.60e-61
Phosphopantetheine attachment site	PF00550.25	CL0314	1105	1172	23.1	6.80e-05
Condensation	PF00668.20	CL0149	1229	1632	99.7	1.50e-28
Adenylation	PF00501.28	CL0378	2073	2482	173.1	7.00e-51
Phosphopantetheine attachment site	PF00550.25	CL0314	2601	2659	3.7	80
Ketosynthase, N-term	PF00109.26	CL0046	2687	2934	152.5	1.50e-44
Ketosynthase, C-term	PF02801.22	CL0046	2942	3060	95.4	2.20e-27
Acyltransferase	PF00698.21	CL0323	3244	3495	82.4	4.00e-23
Thioesterase	PF00975.20	CL0028	3666	3894	60.1	3.30e-16

Multiple sequence alignments of the *H. indistincta* NRPS-PKS with conserved domains of the NCBI database revealed whether the different catalytic domains possessed all expected active site residues. Specifically, the formylation domain possessed 23/23 expected active sites, the first and second adenylation domains possessed 23/23, and the ketosynthase domain possessed 3/3. The condensation domain possessed the motif HHIVFDQ which differs from the conserved HHXXXDG motif of other condensation domains in the CD

database by a single glycine. The acyltransferase domain possessed a GHSLG motif which fulfilled the expected requirements for the conserved motif GHSXG involved in the catalytic activity of this domain [97]. Manual analysis of the region between the acyltransferase and the thioesterase which lacked a phosphopantetheine attachment site indicated that, while the conserved GXDS motif of these domains were not present, a conserved DS motif was.

No similar proteins were identified in the nr database when using the entire NRPS-PKS sequence as well as sequences of the individual catalytic domains (Table 3.5). Despite this NRPS-PKS gene being found on a sponge contig, the closest taxonomical hit for the entire translated gene was of bacteria. This was also observed for the formyltransferase, adenylation, condensation and acyltransferase domains. The ketosynthase domain displayed the highest significant alignment with a sequenced polyketide synthase derived from an "uncultured Porifera" (AAX62314.1) with a percent identity of 51.23%. This sequence was derived from PCR amplification of the ketosynthase domain from a species possibly of the family Raspailiidae which was collected in Curaçao (Joe Lopez, personal communication). Several other ketosynthase sequences were also amplified from this same specimen and uploaded to the NCBI database but were not in the top 100 hits against the nr database. The thioesterase domain also showed highest similarity to eukaryotic sequences, albeit from fungi such as *Melanogaster broomeanus* with a percent identify of 46.34%.

A complete homolog to the NRPS-PKS was not detected in the SPAdes assembly of MIIG1389 (*H. viscosa*) nor in the assembled transcriptomes previously detailed (Chapter 2). However, a single contig with a length of 234 bp and 2.823x read coverage from the MIIG1389 assembly did align with the nucleotide sequence of the MIIG1388 NRPS-PKS gene with 85% identity; such an alignment is far higher than any alignments this NRPS-PKS gene shows when queried against NCBI databases. Similarly, this MIIG1389 contig has no significant alignments against the NCBI nt database. An additional blast analysis using the merged Illumina reads of MIIG1389 allowed the identification of several raw reads which significantly aligned with the NRPS-PKS gene of MIIG1388. In particular, one merged read pair of 175 bp had a 100% identity to a region of the second adenylation domain in the MIIG1388 NRPS-PKS. In comparison, the most significant hit against the NCBI nr database was an adenylation domain of an NRPS from the CFB group bacteria *Pedobacter africanus* with a percent identity of 40.35% and query coverage of 97%.

Alignment of raw long reads greater than 1 kbp on the contig from MIIG1388 which contained the NRPS-PKS gene showed an even coverage across this target as well as two predicted downstream genes, g7534.t1 and g7535.t1, which most aligned to metazoan proteins (Figure 3.5). The highest alignment of g7534.t1 is with a NACHT, LRR and PYD domains-containing protein 14 isoform X2 (XP_028517167.1) from the cnidarian *Exaiptasia diaphana* with a percent identity of 22.60%. Also, this protein had a less significant alignment to an entry from *A. queenslandica* of the same functional annotation. The highest alignment of g7535.t1 is with a malignant fibrous histiocytoma-amplified sequence 1 homolog

(XP_035693581.1) from the lancelet *Branchiostoma floridae* with a percent identity of 23.06%. Furthermore, long reads which encompassed regions of both the NRPS-PKS gene as well as these two downstream genes were detected. Upstream of the NRPS-PKS gene was a 1292 bp region which exhibited significantly higher coverage. No significant alignments were detected when running blastn of this region against the NCBI nr/nt database. Augustus predicted fourteen genes which lay between the beginning of the contig and this region of higher coverage; the majority were found to significantly align with proteins of sponge or other metazoan origin. Mapping of the RNA-seq short-reads indicated that the NRPS-PKS gene was weakly transcribed at the time of sampling (data not shown). Fragmented transcripts which aligned to the gene were detected in the Trinity-assembled transcriptome.

Table 3.5 blastp results of the NRPS-PKS and its individual domains against the NCBI nonredundant database. Enzymatic domain naming follows the same scheme as previously used (Figure 3.4).

Query	Function	Taxonomy	Max Score	% ID	E-value	Accession
NRPS-PKS	amino acid adenylation domain-containing protein	*Pyxidicoccus* sp. SCPEA002	526	27.14	7e-147	WP_206727271.1
F	hybrid non-ribosomal peptide synthetase/type I polyketide synthase	*Legionella maceachernii*	96.3	52.00	7e-21	WP_065239994.1
A1	non-ribosomal peptide synthetase	*Caldithrix abyssi*	175	31.40	8e-44	WP_006929449.1
C	non-ribosomal peptide synthetase	*Streptococcus macacae*	104	26.65	3e-20	WP_003080184.1
A2	amino acid adenylation domain-containing protein	*Paenibacillus zanthoxyli*	145	35.07	2e-34	WP_025688024.1
KS	Polyketide synthase	uncultured Porifera	262	51.23	1e-82	AAX62314.1
AT	SDR family NAD(P)-dependent oxidoreductase	*Okeania* sp. SIO2G5	130	37.63	8e-31	NEP72547.1
TE	BcPKS20, polyketide synthase	*Rhexocercosporidium* sp. MPI-PUGE-AT-0058	79.7	32.65	5e-13	KAH7304361.1

Identification of other sponge megasynthase Genes

The genomes of *E. muelleri*, *S. carteri*, *O. pearsei* and *L. baikalensis* as well as the transcriptomes of *Ephydatia fluviatilis*, *Lubomirskia abietina*, *Baikalospongia bacillifera*, *Cinachyrella alloclada*, *Phakellia ventilabrum* and *Plakina jani* were determined to be the source of similar genes to that of the *H. indistincta* NRPS-PKS (Figure 3.6). Specifically, hybrid NRPS-PKS systems were limited to several sponges of the class Demospongiae whereas a NRPS system was limited to two from the class Homoscleromorpha. No similar system was detected in the Calcarea or Hexactinellida.

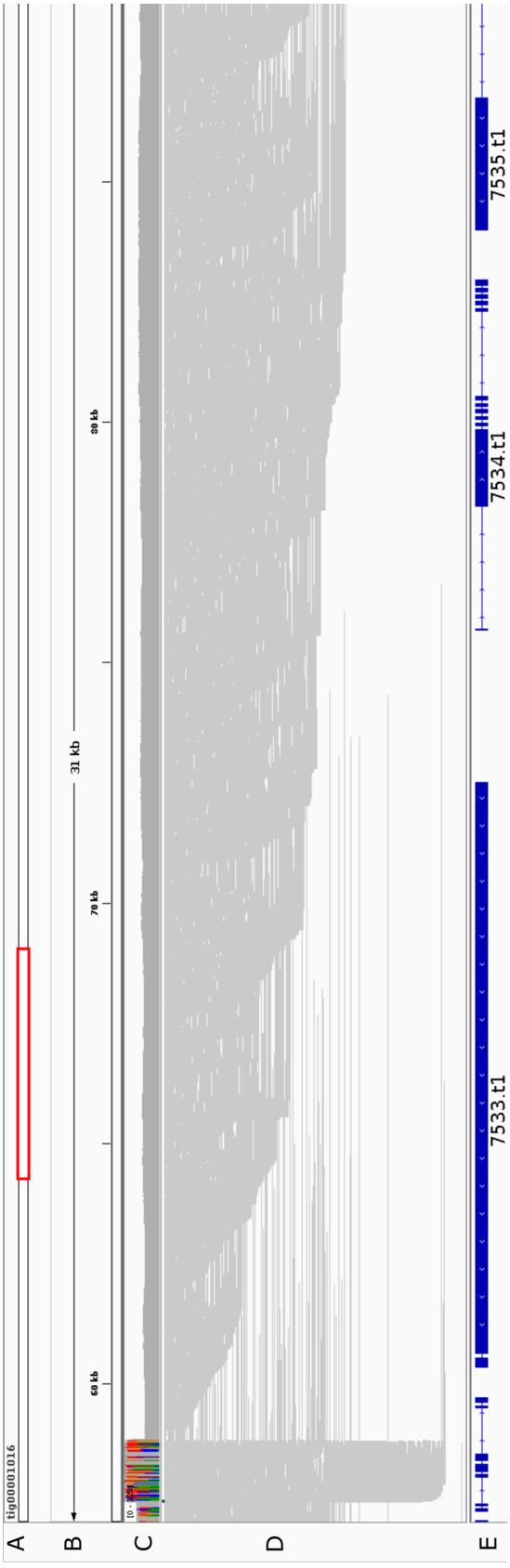


Figure 3.5 Mapping of raw Pacific Biosciences reads greater than 1 kbp on the MIIG1388 contig which contains the NRPS-PKS gene (tig00001016). Row A represents the contig with the region containing the NRPS-PKS gene highlighted in red. Row B represents the length of the region containing the NRPS-PKS gene. Row C represents the average coverage per nucleotide. Row D represents the individual, raw long reads of 1 kbp or greater mapped back to the contig. Row D represents the predicted genes: 7533.t1, NRPS-PKS; 7534.t1 and 7535.t1, genes with highest sequence similarity to those from metazoans.

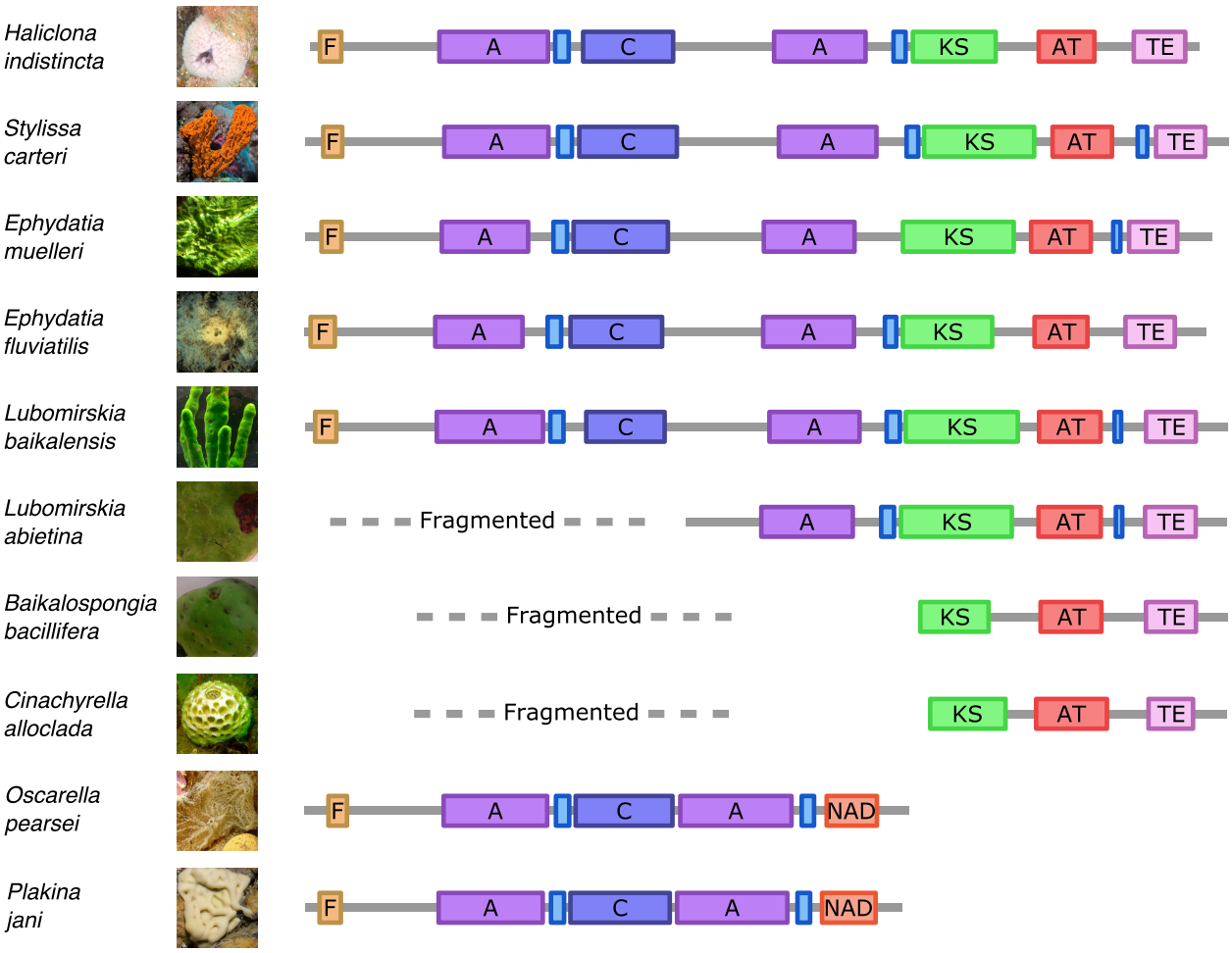


Figure 3.6 Complete or partial NRPS-PKS or NRPS enzymes identified in sponge genomes and transcriptomes. If a fragmented gene was identified, the largest fragment was incorporated into the figure. That of *P. ventilabrum* was too fragmented to include. Enzymatic domain naming follows the same scheme as previously used (Figure 3.4); NAD stands for NAD-binding. Image credits of sponges are as follows: *H. indistincta*, Daniel Rodrigues; *S. carteri*, Tane Sinclair-Taylor [19]; *E. muelleri*, Sally Leys; *E. fluviatilis*, Echte zoetwaterspons, Saxifraga-Eric Gibcus; *L. baikalensis*, Nathan Kenny and Valeria Itskovich [21]; *L abietina*, Nathan Kenny and Valeria Itskovich [21]; *B. bacillifera*, Nathan Kenny and Valeria Itskovich [21]; *C. alloclada*, Rob Ruzicka and Rob van Soest [98]; O. *pearsei*, Jennyfer Mitchell [17]; *P. jani*, Bernard Picton.

Predictive Functional Characterization of the *H. indistincta* NRPS-PKS

Online tools to classify catalytic domains and predict their substrate specificity were generally not able to do so with high confidence when run on the *H. indistincta* NRPS-PKS. The NRPSPredictor2 feature of antiSMASH was incapable of predicting the substrate which would be utilized by either adenylation domain of the NRPS-PKS; the ATSignature and minowa features both predicted malonyl-CoA as being the substrate that would be utilized by the PKS portion of the hybrid system with respective scores of 54.2% and 43.2%. In agreement with its placement directly next to a NRPS module, NaPDoS classified the ketosynthase domain of this system as belonging to the hybrid ketosynthase domain class

with the closest match being a ketosynthase involved in epothilone biosynthesis with an E-value of 3e-71. In turn NaPDoS classified the condensation domain as being of the DCL class which are involved in linking an L-amino acid to a D-amino acid with the closest hit being a fengycin condensation domain with an E-value of 5e-12. However, upon constructing a phylogenetic tree with NaPDoS, the *H. indistincta* condensation domain appears closest to a condensation domain of the dual class involved in HC-toxin biosynthesis by the fungus *Cochliobolus carbonum* (data not shown). The AdenylPred software predicted with high confidence that the first and second adenylation domains of the *H. indistincta* NRPS-PKS were of the NRPS class with respective percent confidences of 93% and 88% (Table 3.6). However, the substrate prediction was less clear. The first adenylation domain was predicted to utilize bulky mainly phenyl derivative substrates with a confidence of 34%. The second was predicted to utilize cyclic aliphatic substrates with a confidence of 30%.

Table 3.6 AdenylPred results for four AMP-binding enzymes which activate NTA as well as the two adenylation domains A1 and A2 found in the *H. indistincta* NRPS-PKS. FC stands for functional class. SS stands for substrate specificity.

Species	**Predicted functional class (FC)**	**FC prediction probability**	**Predicted substrate specificity (SS)**	**SS prediction probability**
H. indistincta A1	NRPS	0.9	bulky mainly phenyl derivatives	0.34
H. indistincta A2	NRPS	0.87	cyclic aliphatic	0.29
S. albus	Aryl-CoA ligase	0.73	aryl and biaryl derivatives	0.7
Micromonospora sp. TP-A0468	NRPS	0.78	polar and charged	0.32
A. fumigatus	Aryl-CoA ligase	0.64	aryl and biaryl derivatives	0.54
P. canescens	Aryl-CoA ligase	0.58	aryl and biaryl derivatives	0.36

Four AMP-binding enzyme domains are known to prepare NTA for polyketide biosynthesis. The characterization of the *H. indistincta* NRPS-PKS adenylation domains by antiSMASH and Adenylpred was not consistent with that made for these enzymatic domains. The features of antiSMASH specific for adenylation domains were only able to predict function for that of the *S. albus* enzyme with the physiochemical class of hydrophobic-aromatic being predicted. Unlike the adenylation domains of the *H. indistincta* NRPS-PKS, AdenylPred classified the enzymes of *S. albus*, *A. fumigatus*, and *P. canescens* as belonging to the Aryl-CoA ligase functional class with probability of 73%, 64% and 58% respectively (Table 3.6). The adenylation domain of the *Micromonospora* sp. TP-A0468 was classified as belonging to the NRPS functional class and having a substrate specificity for polar and charged amino acids

with respective probabilities of 78% and 32%. Querying the four AMP-binding enzymes which prepare NTA as a starter unit for polyketide biosynthesis against the MIIG1388 assembly resulted in three significant alignments. Two identical MIIG1388 proteins aligned with the enzymes of *S. albus*, *A. fumigatus* and *P. canescens.* When queried against the nr database these two MIIG1388 proteins significantly aligned with a predicted 4-coumarate—CoA ligase from *A. queenslandica* (XP_003386175.1) and the luciferase polypeptide from *Suberites domuncula* (CAR31336.1). The adenylation domain of *Micromonospora* sp. TP-A0468 which accepts NTA was found to significantly align with the second adenylation domain of the *H. indistincta* NRPS-PKS, albeit with a low bit score of 57.4.

Manual alignments of the Stachelhaus codes of both *H. indistincta* adenylation domains did not reveal much consistency with codes from biochemically characterized adenylation domains (Table 3.7). Specifically, the first had a match of six residues with those domains known to utilize the substrates D-glutamate, isoleucine, and N^{δ}-cis-anhydromevalonyl-N^{δ}-hydroxy-L-ornithine. The second had a lower match of five residues with those domains known to utilize 6-chloro-L-typtophan, N6-decanoyl-N6-hydroxy-L-lysine, (S)-β-tyrosine, and (2S,4R)-4-propyl-L-proline. Even less consistency was observed when comparing the *H. indistincta* Stachelhaus codes with those from the AMP-binding enzymes known to activate NTA for polyketide biosynthesis in that a match of 3 was the highest detected consistency (Table 3.8). Unexpectedly, the highest observed consistency of seven residues was between the second adenylation domain of *H. indistincta* with the corresponding domain from the NRPS-PKS of *E. muelleri* (Table 3.9).

Phylogenetic Analysis of Hybrid Ketosynthase Domains

As can be seen in the unrooted hybrid ketosynthase phylogenetic tree, those of the sponges form a well-resolved clade with high statistical support (Figure 3.7). This sponge clade is both distinct and distant from those of bacteria and other eukaryotes. Similarly, well-resolved, statistically supported clades of hybrid ketosynthases from algae, nematodes, and fungi were also produced. The topology of the tree indicates that the sponge sequences have a stronger phylogenetic signal towards those of bacteria than other eukaryotes. One sequence from a *Streptomyces* sp. (WP_228078005.1) branches off early from that of the sponges, but there is no statistical support for this sequence grouping either with those of the sponges or the other bacteria. To expand on statistical support, it should be noted that the deeper nodes throughout this tree do not possess bootstrap values above 75% and therefore deeper relationships cannot be determined with accuracy. No improvement in statistical support for the deeper nodes was observed upon trimming the edges of the alignment which had lower quality and more gaps (data not shown). Several additional hybrid ketosynthases purportedly derived from metazoans (e.g., several from arthropods and rotifers) were excluded from the tree as they consistently nestled within the bacterial sequences (data not shown). Such a quality was interpreted as possible bacterial contamination at the time of sequencing and thus omitted.

Table 3.7 Alignment of *H. indistincta* Stachelhaus codes with those from bacteria and fungus having the highest number of matching residues. Residues are coloured according to the Clustalx colour code. Abbreviations are as follows: Pos., position; Hi, *H. indistincta*. McyE, BacA, SidNA3, FmoA1, MbtE, SgcC1 and LmbC represent adenylation domains from genes associated with the biosynthesis corresponding compounds. The matching numbers refer to the consistency between the Stachelhaus code of *H. indistincta* adenylation domains and those which are most similar and biochemically characterized.

Enzyme	Substrate	Pos. 235	Pos. 236	Pos. 239	Pos. 278	Pos. 299	Pos. 301	Pos. 322	Pos. 330	Pos. 331	Pos. 517	Compound	Match
Hi A1	Unknown	D	G	H	F	S	G	V	I	G	K	Unknown	-
McyE	D-Glu	D	P	R	H	S	G	V	V	G	K	Microcystin	6
BacA	Ile	D	G	F	F	L	G	V	V	Y	K	Bacitracin	6
SidNA3	cis-AMHOa	D	V	G	G	G	G	V	I	G	K	Fungal siderophore	6
Enzyme	Substrate	Pos. 235	Pos. 236	Pos. 239	Pos. 278	Pos. 299	Pos. 301	Pos. 322	Pos. 330	Pos. 331	Pos. 517	Compound	Match
Hi A2	Unknown	D	V	A	F	F	A	W	I	A	K	Unknown	-
FmoA1	6-chloro-L-Trp	D	G	W	C	V	A	V	I	A	K	JBIR-34	5
MbtE	N6-decanoyl-N6-hydroxy-L-Lys	D	V	A	H	P	G	F	I	N	K	Mycobactin	5
SgcC1	(S)-β-Tyr	D	P	A	Q	L	M	L	I	A	K	C-1027	5
LmbC	(2S,4R)-4-propyl-L-Pro	D	V	A	L	V	A	I	G	C	K	Lincomycin	5

Table 3.8 Alignment of *H. indistincta* Stachelhaus codes with those from AMP-binding enzymes which activate NTA for incorporation into polyketides. Residues are coloured according to the Clustalx colour code. Abbreviations are as follows: Pos., position; Hi, *H. indistincta*. The matching numbers refer to the consistency between the Stachelhaus code of *H. indistincta* adenylation domains and those which are most similar and biochemically characterized.

Adenylation Sequence	Substrate	Pos. 235	Pos. 236	Pos. 239	Pos. 278	Pos. 299	Pos. 301	Pos. 322	Pos. 330	Pos. 331	Pos. 517	Compound	Match
Hi A1	Unknown	D	G	H	F	S	G	V	I	G	K	Unknown	-
Micromonospora sp.	Nicotinic acid	D	L	L	Q	V	T	L	N	D	K	Kosinostatin	2
S. albus	Nicotinic acid	S	F	P	A	Q	G	V	L	C	K	Pyridinopyrone A	3
A. fumigatus	Nicotinic acid	F	G	W	Y	G	S	L	V	V	K	Pyripyroene A	2
P. canescens	Nicotinic acid	F	G	W	Y	G	S	L	V	I	K	5-Deoxyoxalicine B	2
Sequence	Substrate	Pos. 235	Pos. 236	Pos. 239	Pos. 278	Pos. 299	Pos. 301	Pos. 322	Pos. 330	Pos. 331	Pos. 517		Match
Hi A2	Unknown	D	V	A	F	F	A	W	I	A	K	Unknown	-
Micromonospora sp.	Nicotinic acid	D	L	L	Q	V	T	L	N	D	K	Kosinostatin	2
S. albus	Nicotinic acid	S	F	P	A	Q	G	V	L	C	K	Pyridinopyrone A	1
A. fumigatus	Nicotinic acid	F	G	W	Y	G	S	L	V	V	K	Pyripyroene A	1
P. canescens	Nicotinic acid	F	G	W	Y	G	S	L	V	I	K	5-Deoxyoxalicine B	1

Table 3.9 Alignment of *H. indistincta* Stachelhaus codes with those from other sponge NRPS-PKS and NRPS enzymes. Residues are coloured according to the Clustalx colour code. Abbreviations are as follows: Hi, *H. indistincta*; Ef, *E. fluviatilis*; Em, *E. muelleri*; Lb, *L. baikalensis*; Sc, *S. carteri*; Op, *O. pearsei*; Pj, *P. jani*; A1, adenylation 1; A2, adenylation 2. The matching numbers refer to the consistency between the Stachelhaus code of *H. indistincta* adenylation domains and those which are most similar and biochemically characterized.

Sequence	Pos. 235	Pos. 236	Pos. 239	Pos. 278	Pos. 299	Pos. 301	Pos. 322	Pos. 330	Pos. 331	Pos. 517	Match
Hi A1	D	G	H	F	S	G	V	I	G	K	-
Ef A1	D	T	L	S	A	G	L	V	I	R	2
Em A1	D	S	L	L	A	G	V	V	I	K	4
Lb A1	D	T	L	S	A	G	L	V	I	K	3
Sc A1	D	A	H	H	S	G	I	V	A	K	5
Op A1	D	G	C	S	I	G	C	V	V	K	4
Pj A1	D	G	C	C	I	G	C	I	V	K	5
Sequence	Pos. 235	Pos. 236	Pos. 239	Pos. 278	Pos. 299	Pos. 301	Pos. 322	Pos. 330	Pos. 331	Pos. 517	Match
Hi A2	D	V	A	F	F	A	W	I	A	K	-
Ef A2	D	I	A	H	F	T	W	V	E	K	5
Em A2	D	L	A	H	F	G	W	I	A	K	7
Lb A2	D	I	A	H	F	T	W	V	E	K	5
Sc A2	D	G	L	F	I	I	W	I	A	K	6
Op A2	D	I	V	C	L	V	A	G	I	K	2
Pj A2	D	M	S	H	Y	L	L	V	A	K	3

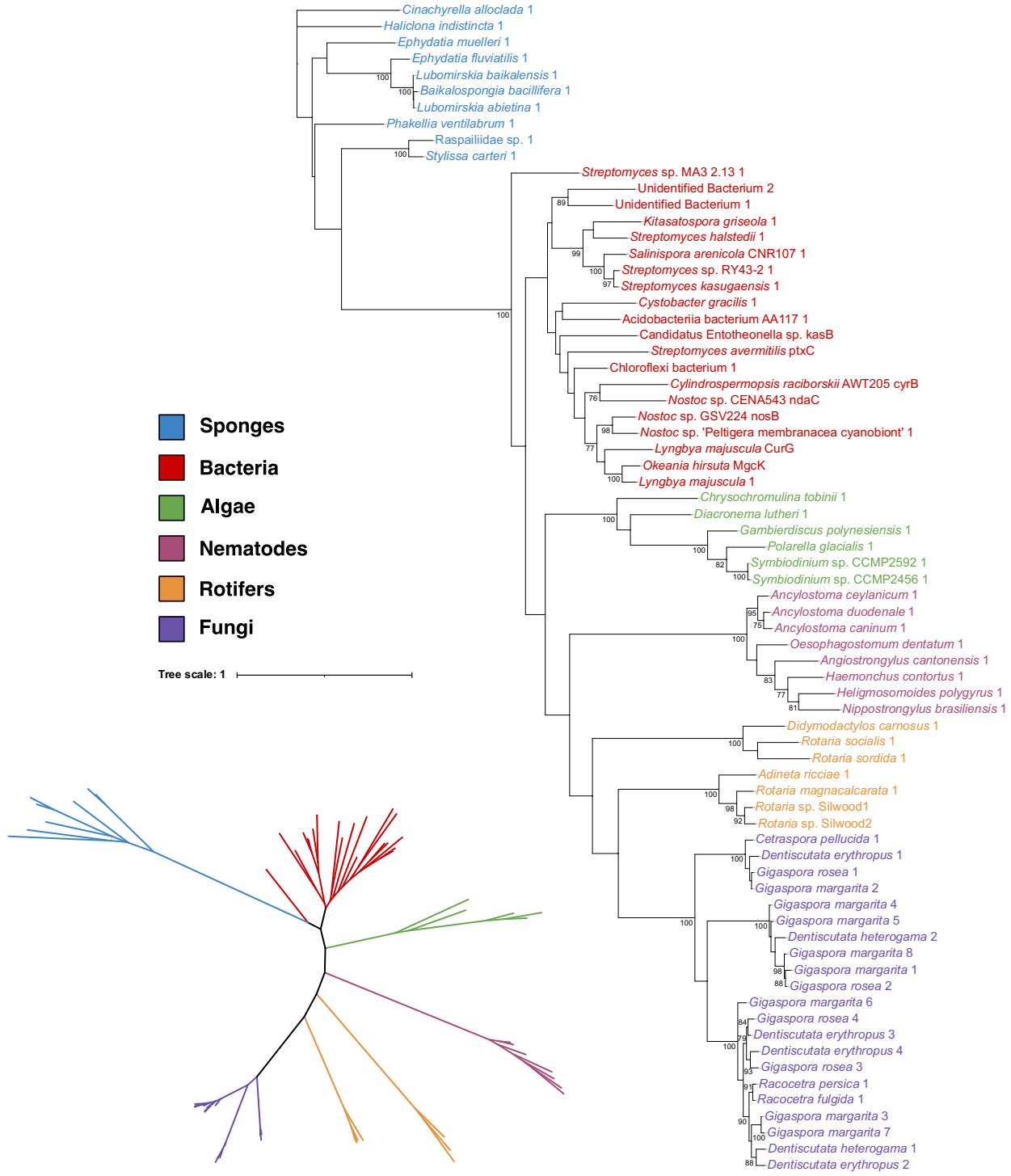


Figure 3.7 Unrooted phylogenetic trees of hybrid ketosynthase domains from sponges, bacteria, and other eukaryotes. Numbers on nodes represent maximum likelihood bootstrap values above 75%.

Discussion

In this chapter the *de novo* genome of the Irish sponge *H. indistincta* was sequenced using both short- and long-read technology. This has allowed for the assembly of a highly contiguous genome of the host organism separate from associated microbial genomes. In turn, genome mining identified the presence of a single megasynthase gene, although its role in the biosynthesis of 3-APs is questionable. Specifically, a novel hybrid NRPS-PKS gene has been identified in a contig which appears to originate from a chromosome of *H. indistincta*. This represents the first instance of a megasynthase-type enzyme found to be encoded in a sponge chromosome. As a result, it is likely that the specialized metabolism of sponges should not be overshadowed by their microbiome when trying to understand the biosynthetic origin of specialized metabolites derived from these animals.

The final binned *H. indistincta* genome possessed an 87.8% complete BUSCO score from the eukaryotic dataset which represents a high level of predicted completion for this assembly. In comparison the most well-assembled sponge genome publicly available, that of *E. muelleri* assembled to the chromosome level, scores lower at 83.83% [22]. However, the *H. indistincta* genome is lower than that of the first sponge to have its genome sequenced, *A. queenslandica*, which scores at 90.43% [15]. The final assembly from Canu also exhibited a high duplication rate in comparison to the Flye and Raven assemblies. This is most likely due to fewer collapsed haplotypes and an overall more diploid assembly [99]. The benefits of long-read sequencing to achieve a high degree of contiguity is evident in this genome assembly. Even though the length of sequenced reads did not reflect the size of the library prepared, a final N50 value of 418.01 kbp was still achieved which is second only to the assembly of *E. muelleri*. It is possible that a contaminating compound, such as the 3-APs, interfered with the polymerase of the sequencing technology; the acquisition of HMW gDNA suitable for long-read sequencing is notoriously difficult for marine invertebrates due to the production of polysaccharides, polyphenols and other specialized metabolites [100]. Some caution should be accompanied when interpreting the result of how complete the genome is based on the eukaryotic BUSCO score. Contamination was minimized in both the sequencing of *A. queenslandica* and *E. muelleri* by selecting for specific types of tissue which would be relatively sterile in comparison to adult tissue—embryos and gemmules respectively. In contrast adult tissue was used to sequence *H. indistincta* as a microbial origin of the 3-APs could not be ruled out at the time. *Haliclona indistincta* harbours a diverse array of eukaryotic epibionts such as crustaceans and polychaetes (personal observation). While these contaminating eukaryotes were removed prior to DNA extraction and the assembly binned to remove contigs not of sponge origin, the possibility of eukaryotic contamination remaining in the final assembly is something that cannot be completely ruled out.

Overall, the *H. indistincta* microbiome was a scarce resource for novel BGCs detectable by current genome mining tools despite sponge-associated microorganisms being

acknowledged as a promising resource for this purpose [101]. This is not unexpected as the species is considered a low microbial abundance sponge based on transmission electron microscopy as well as 16S rRNA sequencing [37]. The ecological function of specialized metabolites is often believed to be for defence from predators and pathogens. This is consistent with 3-AP polymers such as halitoxin and amphitoxin which have several reported activities with these ecological functions. This includes ichthyotoxicity as well as selective antibacterial activity to bacteria isolated from the surrounding seawater of sponges producing these compounds, but not to isolates from the sponge itself [102,103]. In the case of *H. indistincta* it may be that the halitoxin-like 3-AP it produces fulfils such roles. Thus, due to the halitoxin-like compound there may be no need for many microbial symbionts that produce other bioactive compounds to defend the sponge holobiome.

Terpene BGCs, many appearing to be for carotenoids, were the most numerously detected as evidenced by the presence of phytoene synthase and lycopene cyclase genes. These molecules have numerous functions including the absorption of light for photosynthesis as well as photoprotection [104]. Furthermore, sponges are known to be a source of diverse carotenoids with unique modifications and these compounds are often associated with the colours of these animals, but it is suspected that symbiotic microorganisms are the true source of sponge-derived carotenoids [104]. Given the intertidal nature of *H. indistincta* it likely experiences a larger amount of sunlight in comparison to sponges which live in deeper waters. This is evident by the large number of algae it is often in association with [105]. Thus, the presence of bacteria with high-sequencing coverage which contain genes for carotenoid biosynthesis is sensible and may contribute to the health of the *H. indistincta* holobiome. Carotenoids have garnered some attention due to their antioxidant activities [106]. For these reasons *H. indistincta* may represent a potential source of novel carotenoids with possible applied functions. However, it should also be considered that these bacteria may not necessarily be in association with *H. indistincta* but instead prominent inhabitants of the seawater in the area. For example, bin 28 is of the Order Rhodobacterales and members of this order have been found to be commonly observed in seawater samples from which *H. indistincta* can be collected from [37].

The RiPP-like BGC identified in one of the bacterial bins is primarily characterized by the presence of a gene encoding for a protein with two domains of unknown function (DUF), DUF692 and DUF2063. However, despite its categorization as being RiPP-like, no gene encoding for a precursor RiPP peptide was identified in this BGC. Surrounding this ORF are several genes related to fatty acid biosynthesis as well as one gene encoding an alanine racemase. Microbial symbionts of sponges have been shown to be a rewarding source of novel RiPP natural products [107]. Furthermore, some members of the DUF692 family are involved in the biosynthesis of RiPPs such as methanobactin [108]. Based on the low sequence similarity of the ribosomal sequences of bin 8 as well as the ORFs associated with its RiPP gene cluster, this BGC potentially represents a novel RiPP natural product. This class

of compounds display a wide variety of biological activities and are thus of interest for pharmaceutical applications [109]. For this reason, the novel RiPP-like BGC of bin 8 may necessitate future investigation for the sake of discovering new drugs. However, the genes of this BGC are inconsistent with the hypothesis on the biosynthesis of 3-APs involving an AMP-binding enzyme preparing NTA as a starter unit for polyketide biosynthesis. In addition, a similar BGC was not detected in the genome assembly of MIIG1389. For these reasons this BGC was assumed to not be responsible for the 3-APs of *H. indistincta*.

While plantiSMASH did detect BGCs in the binned *H. indistincta* genome, it is difficult to assess whether these are truly related to 3-AP biosynthesis. None of these identified BGCs included genes that logically fit in with the PKS hypothesis on 3-AP biosynthesis. The main exception are the BGCs containing several long chain fatty acyl-CoA ligases which, if related to 3-AP biosynthesis, may represent a scenario where a complete fatty acid is condensed with NTA (Figure 1.13). However, such an enzymatic reaction has not been reported in the literature compared to utilizing NTA as a starter unit for biosynthesis and thus these BGCs were not pursued for further analysis.

The high number of enzymatic domain co-occurrences not found in the UniProt, BRENDA or MIBiG databases as determined by CO-ED indicates that there may be novel enzymatic reactions capable by the *H. indistincta* proteome. However, whether the enzymes containing these novel co-occurrences are involved in specialized metabolism is difficult to predict without functional characterization. Because megasynthases, such as PKSs, comprise enzymes which possess multiple enzymatic domains they are readily detectable in the CO-ED output. As CO-ED only identified a single megasynthase, the NRPS-PKS, in the *H. indistincta* proteome this greatly narrows down the potential candidates of this enzymatic class which could be responsible for the biosynthesis of 3-APs via a PKS mechanism.

The presence of a gene encoding a hybrid NRPS-PKS within the genome of *H. indistincta* is unprecedented. While sponges have long been recognized as prolific sources of bioactive specialized metabolites, little evidence on the genetic level has been published showing that the host sponge itself produces such compounds rather than microbial symbionts. The only published pieces of evidence have been derived from the genome of the model sponge *A. queenslandica* which possesses predicted genes responsible for the biosynthesis of sterols and ovothiols as well as the deep-sea species *Geodia barretti* which is the source of the barrettide peptides [110-112]. However, genes for megasynthase enzymes such as PKS, NRPS and hybrid systems have not been reported from these species or other sponges. Indeed, all published literature on megasynthase genes from sponge holobiomes have been derived from symbiotic microorganisms such as bacteria [27,113]. Yet, it has recently become apparent that metazoans are capable of synthesizing specialized metabolites [114]. In particular functional PKS genes have been shown to exist in the genomes of the birds *Melopsittacus undulatus* and *Anas platyrhynchos*, the sea urchin *Hemicentrotus pulcherrimus* and the sea slug *Elysia chlorotica* [115-118]. Only a single hybrid system has been reported

in animals with the discovery of the biosynthetic gene cluster for nemamides in the nematode *Caenorhabditis elegans*, although similar systems exist in other nematodes [119].

The metazoan origin of the *H. indistincta* NRPS-PKS gene is supported by several pieces of evidence. First, mapping of the raw Pacific Biosciences read longer than 1 kbp upon the contig which contains the NRPS-PKS gene indicate an even level of coverage across this gene of interest as well as those of putative metazoan origin. If this region of the contig was the result of a misassembly an uneven level of coverage may instead be expected separating the NRPS-PKS from downstream metazoan genes [120]. In addition, several reads long enough to cover parts of both the NRPS-PKS gene as well as the aforementioned metazoan genes were also identified which indicates that both come from the same chromosome (data not shown). Furthermore, the mapping of raw RNAseq reads to the NRPS-PKS gene is also evidence of eukaryotic rather than bacterial origin. This is because the library from which this RNAseq data was derived underwent poly(A) selection which avoids mRNA from bacteria [121]. In addition, it shows that this gene was being expressed at the time of sampling and thus has a biological function rather than being a pseudogene. In contrast to a typical BGC, the NRPS-PKS of *H. indistincta* appears to be encoded on a single ORF as indicated by both antiSMASH analysis and AUGUSTUS gene-prediction. The two nearby downstream genes, g7534.t1 and g7535.t1, appear to be of clear metazoan origin. However their predicted function does not seem to indicate any role in the biosynthesis of a specialized metabolite, but instead inflammation and innate immunity [122,123]. Such functions may be consistent with the defensive purposes of specialized metabolites such as those which may be produced by the hybrid NRPS-PKS and for this reason these downstream genes may serve a regulatory purpose.

The PKS portion of the *H. indistincta* NRPS-PKS bears some anomalous features when trying to connect it to 3-AP biosynthesis. When considering the aliphatic chain of the 3-AP polymer associated with *H. indistincta*, one would expect enzymatic domains to fully reduce the ketone backbone as seen in systems such as highly reducing iterative PKSs [124]. However, the identified NRPS-PKS does not possess ketoreductase, dehydratase or enoylreductase domains. Running an hmmscan search using the NRPS-PKS sequence as a query against the Pfam database indicated that the acyltransferase domain does have low homology to the starter unit:ACP transacylase (SAT; Pfam PF16073.5) domain involved in aflatoxin biosynthesis by *Aspergillus parasiticus* [125]. This domain selects a hexanoyl starter unit, which possesses a fully reduced aliphatic chain, and incorporates it into polyketide biosynthesis [126]. The use of a similar fatty acid building block could explain the nature of the aliphatic chains that the *H. indistincta* 3-AP polymer possesses. Alternatively, it may be that the NRPS-PKS sequentially adds extender units derived from malonyl-CoA upon a NTA starter unit in a typical PKS fashion. Following this, the elongated ketone backbone could be fully reduced by separate *trans*-acting enzymes akin to the ketoreductase involved in the biosynthesis of the polyketide dimer SIA7248 [127]. Interestingly, the BGC found in the

MIBiG database with the closest similarity to the *H. indistincta* NRPS-PKS is a PKS responsible for the biosynthesis of 1-heptadecene from *Cyanothece* sp. PCC 7822, although only with a similarity score of 0.3 [128]. Such a connection bears some relevance to the notion of a fatty acid origin for the alkyl chain of the 3-APs. NaPDoS identifying the ketosynthase domain as being closest to the hybrid ketosynthase involved in epothilone biosynthesis has some consistency with the hypothesis on 3-AP biosynthesis [129]. Specifically, this ketosynthase incorporates a methylthiazolylcarboxy starter unit, which is heterocyclic like NTA, into polyketide biosynthesis. Furthermore, there would logically be a third phosphopantetheine attachment site within the PKS portion of this system to transfer the complete product to the thiosterase domain for release. However, the presence of conserved DS residues, while not entirely consistent with the GXDS motif of acyl-carrier proteins, is a good indication that this function exists in the PKS portion [130].

Considering the NRPS portion of the hybrid enzyme, it is unclear why two domains would be necessary to produce 3-APs considering the hypothesis that a single AMP-binding enzyme would prepare NTA as a starter unit in polyketide biosynthesis. The condensation domain of a NRPS is typically used to create a peptide bond between two L-amino acids which is a feature not seen in 3-APs. NaPDoS classified the *H. indistincta* condensation domain as belonging to the more unique DCL class, which creates a peptide bond between a D-amino acid and an L-amino acid, but at a low percent identity. Based on the low confidence, it is difficult to conclude whether a peptide bond is truly formed by this NRPS-PKS as condensation domains can also possess a number of alternative functions in natural products biosynthesis such as the incorporation of fatty acids [131]. Such alternative functions would more logically align with hypotheses on the biosynthesis of 3-APs. The formylation domain at the N-terminus of the NRPS-PKS represents another unique feature not seen before in megasynthases from other metazoans. Typically, this domain is located within the loading module of an NRPS and is responsible for adding a formyl group to the first amino acid incorporated into peptides such as gramicidin A [132]. As the common structure of 3-APs lacks any sort of amino acid, the presence of this domain also downplays the notion that the NRPS-PKS is responsible for the biosynthesis of these compounds. The only potential relevance of the reaction catalysed by a formylation domain is how pyridine rings can also be formed by non-enzymatic reactions involving an aldehyde and an amine. Specifically, quinolinic acid, a precursor of NTA, can be produced from 2-amino-3-carboxymuconate semialdehyde which is derived from tryptophan metabolism [133]. However, this differs from the function of a NRPS formylation domain which directly formylates an amino acid. Furthermore, as previously mentioned, fully formed NTA molecules are incorporated into 3-AP biosynthesis independent of quinolinic acid meaning that such a non-enzymatic reaction is unnecessary [134]. As such, it is likely that this formylation domain is involved in catalysing a reaction not involved in 3-AP biosynthesis.

It is difficult to conclude why only one small contig and several reads from the *H. viscosa* data significantly align with the MIIG1388 NRPS-PKS far better than anything in the NCBI nr database. This is especially so because a SPAdes assembly of the *H. indistincta* data on default settings resulted in the complete assembly of a contig containing all of the NRPS-PKS gene (data not shown). Because *H. viscosa* is the sister species to *H. indistincta* it would be expected that it also possesses a similar NRPS-PKS system. Furthermore, *H. viscosa* is also a known producer of a polymeric, halitoxin-like 3-AP and if this hybrid system is responsible for these compounds, it would be expected that *H. viscosa* possesses a similar modular enzyme. The DNA of *H. viscosa* was mistakenly extracted with 25:24:1 P:C:IAA with a pH of 6.0 which was not adjusted to an appropriate pH for DNA extraction. For this reason, most of the DNA would not be partitioned into the aqueous phase which was extracted during the procedure. Instead, at a pH below 7.0 the DNA would be denatured and primarily located in the organic phase (as directed by the manufacturer). Such an impure sample may explain the anomalous results of the *H. viscosa* Illumina data. Overall, the presence of reads with high sequence similarity to the *H. indistincta* NRPS-PKS seems to indicate that this system is present in *H. viscosa*. Future sequencing of this sister species should confirm this prediction.

The AMP-binding enzymes associated with preparing NTA for polyketide biosynthesis, specifically for the compounds pyridinopyrone A, pyripyroene A, 5-deoxyoxalicine B, as well as the standalone adenylation domain of kosinostatin biosynthesis which is capable of promiscuously activating NTA, had little similarity to the domains of the *H. indistincta* NRPS-PKS [33-36]. In particular, those for the first three compounds were classified as being independent aryl-CoA ligases rather than being of the NRPS class. Similar to when these AMP-binding enzymes were used as tblastn query searches against the *Haliclona* transcriptomes, the closest tblastn hits these aryl-CoA ligases have in the MIIG1388 genome were most similar to 4-coumarate—CoA ligase from *A. queenslandica* (XP_003386175.1) and the luciferase polypeptide from *S. domuncula* (CAR31336.1) in the NCBI non-redundant database [135]. For this reason, the similar sequences detected in the MIIG1388 genome are likely not responsible for 3-AP biosynthesis. It is interesting that the fourth AMP-binding enzyme, a standalone adenylation domain which prepares NTA for the biosynthesis of the pyrrolopyrrole moiety of kosinostatin, has a degree of sequence similarity to only one of the adenylation domains of the *H. indistincta* NRPS-PKS. However, the inconsistent and low confidence substrate prediction for both of these domains by AdenylPred makes it difficult to assume that that of the *H. indistincta* NRPS-PKS is associated with NTA.

Manual analysis of the Stachelhaus codes of both adenylation domains from the *H. indistincta* NRPS-PKS also led to inconclusive results regarding substrate specificity. It would be expected that if these domains were to utilize the same substrate as a biochemically characterized one then the code would be identical or similar; such a relationship was not observed with the *H. indistincta* adenylation domains [90]. This was particularly so when comparing the codes of the *H. indistincta* adenylation domains with those associated with

activating NTA as next to no consistency was observed [33-36]. Based on this observation, it is difficult to hypothesize that the *H. indistincta* NRPS-PKS utilizes NTA and is thus responsible for the biosynthesis of 3-APs. However, such a hypothesis should first be assessed biochemically before being discredited. Interestingly, the Stachelhaus codes also varied greatly between the different sponge NRPS-PKS and NRPS enzymes. This implies that they may utilize different substrates and thus may produce different compounds despite the similarity in enzymatic domain structure and organization they share.

The presence of similar genes in other sponge genomes and transcriptomes is an additional piece of evidence that this family of genes are derived from metazoan genomes rather than contamination. For one, many of the sponge NRPS-PKS and NRPS genes were identified in transcriptomes whose libraries were prepared with poly(A) selection which would select for eukaryotic mRNA. However, perhaps the strongest observation supporting this notion is the fact that one was identified in the genome of *E. muelleri* [22]. To date, this is the best assembled sponge genome as it is at the chromosome level thanks to techniques such as chromosome conformation capture. Considering this technique is based on interactions between chromosomes and that bacterial chromosome should not be interacting with those of *E. muelleri*, such data would be able to improve the separation of sponge and bacterial contigs during the process of scaffolding. Furthermore, the DNA acquired for this assembly was from gemmules which can be sterilized on the surface prior to DNA extractions. While the identification of these genes is compelling evidence that sponges themselves can create peptide-polyketide natural products, their presence in species not known to produce 3-APs downplays the notion that the one of *H. indistincta* is involved in the biosynthesis of its 3-AP polymers. Actually, nothing is known about the specialized metabolism of almost all species from which a NRPS-PKS or NRPS gene was identified. The only exception is *S. carteri* from which a diverse range of pyrrole 2-aminoimidazole alkaloids has been isolated [136].

The well-formed, statistically supported clade the sponge hybrid ketoynthases formed also serves as a strong indication that they are not the result of simple bacterial contamination, but instead share a common ancestor distinct from modern bacterial and eukaryotic sequences. Unfortunately, a lack of statistical support in the deeper nodes renders inference on the origin of this portion of the sponge NRPS-PKSs impossible. Several scenarios can be envisioned. Gene duplication from a different megasynthase, such as a fatty acid synthase, has been the explanation for the origin of the ketosynthase domains in the PKS-NRPS of *Caenorhabditis elegans* [137]. However, as *H. indistincta* does not possess a fatty acid synthase from which a gene duplication event could occur it is not possible to perform such an analysis (Figure 2.7). This absence is likely not due to the completeness of the genome, but rather a quality of sponges as seen with *A. queenslandica* [110]. Assuming gene duplication from a fatty acid synthase is the origin, a later loss of the originating gene in the *H. indistincta* genome would be required. Considering that there is a lack of similar NRPS-PKS enzymes in non-metazoan eukaryotes such as choanoflagellates, direct inheritance from

a eukaryotic ancestor also appears unlikely unless a series of gene losses also occurred. This leaves the possibility that this gene, or various portions of it, are the result of horizontal gene transfer from a microorganism such as a bacterium. Such an event has been recorded previously between a sponge and a bacterium, although with a much smaller gene associated with biomineralization [138]. Given that the sponge clade is not nestled more closely within the bacterial sequences, this may be evidence of a more ancient horizontal transfer from bacteria. In turn, the long branch may be an indication of divergence across a long period of time. Overall, this origin *via* horizontal transfer would be consistent with the fact that not only the entire NRPS-PKS gene, but also most of its enzymatic domains, show the highest sequence similarity to bacterial sequences.

A popular hypothesis on the biosynthesis of 3-APs is that it is *via* a PKS which accepts NTA as a starter unit (Figure 1.11B) [139]. If this hypothesis were true, it would be expected that such a PKS would possess ketoreductase, dehydratase and enoylreductase domains to fully reduce the alkyl chain as the 3-AP polymer from *H. indistincta* exhibits this quality. No such PKS was identified in neither the MAGs nor the genome of *H. indistincta*. This suggests that 3-APs are likely not synthesized in such a manner. Furthermore, the notion that they are synthesized by a megasynthase at all is also questionable based on the observations of this work. The only identified gene encoding for PKS enzymatic domains is the *H. indistincta* NRPS-PKS which notably lacks the aforementioned domains for chain reduction. Furthermore, its possession of a condensation domain between two NRPS modules, which is orthodoxically utilized to create peptide bonds, questions its potential role in 3-AP biosynthesis. Furthermore, there is a lack of evidence that the adenylation domains of this NRPS-PKS activate NTA which would be expected of AMP-binding enzymes involved in 3-AP biosynthesis. However, it should be acknowledged that the attempt to functionally characterize the NRPS-PKS in this work is strictly computational prediction upon a protein dissimilar to anything in known databases.

Overall, the work in the *de novo* sequencing and assembly of the *H. indistincta* genome has allowed for the identification of a novel biosynthetic gene encoding for a NRPS-PKS. What is significant is that this gene appears to be present on a chromosome of the sponge rather than being derived from an associated microorganism. Such a quality has never been reported before in the literature. Furthermore, metazoan NRPS-PKSs are exceptionally rare and to date only reported in nematodes. For these reasons, this novel NRPS-PKS gene represents a possible new aspect in both sponge and metazoan specialized metabolism. Yet, the novelty of this hybrid NRPS-PKS has also made it difficult to characterize and understand through bioinformatics methods. Low sequence similarity to anything in databases such as MIBiG, NaPDoS and those of NCBI make predicting the specific function of each enzymatic domain a challenge. As such, functional characterization in the laboratory is needed to truly understand what role this fascinating but enigmatic NRPS-PKS serves to the sponge.

References

1. The Cost of Sequencing a Human Genome [WWW Document], n.d. Genome.gov. URL https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost (accessed 4.21.21).
2. Dovichi, N.J., Zhang, J., 2000. How capillary electrophoresis Ssequenced the human genome. Angewandte Chemie International Edition 39, 4463–4468. https://doi.org/10.1002/1521-3773(20001215)39:24<4463::AID-ANIE4463>3.0.CO;2-8
3. Giani, A.M., Gallo, G.R., Gianfranceschi, L., Formenti, G., 2020. Long walk to genomics: history and current approaches to genome sequencing and assembly. Computational and Structural Biotechnology Journal 18, 9–19. https://doi.org/10.1016/j.csbj.2019.11.002
4. Goodwin, S., McPherson, J.D., McCombie, W.R., 2016. Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics 17, 333–351. https://doi.org/10.1038/nrg.2016.49
5. Jain, M., Olsen, H.E., Paten, B., Akeson, M., 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biology 17, 239. https://doi.org/10.1186/s13059-016-1103-0
6. Wenger, A.M., Peluso, P., Rowell, W.J., Chang, P.-C., Hall, R.J., Concepcion, G.T., Ebler, J., Fungtammasan, A., Kolesnikov, A., Olson, N.D., Töpfer, A., Alonge, M., Mahmoud, M., Qian, Y., Chin, C.-S., Phillippy, A.M., Schatz, M.C., Myers, G., DePristo, M.A., Ruan, J., Marschall, T., Sedlazeck, F.J., Zook, J.M., Li, H., Koren, S., Carroll, A., Rank, D.R., Hunkapiller, M.W., 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nature Biotechnology 37, 1155–1162. https://doi.org/10.1038/s41587-019-0217-9
7. Belton, J.-M., McCord, R.P., Gibcus, J.H., Naumova, N., Zhan, Y., Dekker, J., 2012. Hi–C: A comprehensive technique to capture the conformation of genomes. Methods, 3D chromatin architecture 58, 268–276. https://doi.org/10.1016/j.ymeth.2012.05.001
8. Medema, M.H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J.B., Blin, K., de Bruijn, I., Chooi, Y.H., Claesen, J., Coates, R.C., Cruz-Morales, P., Duddela, S., Düsterhus, S., Edwards, D.J., Fewer, D.P., Garg, N., Geiger, C., Gomez-Escribano, J.P., Greule, A., Hadjithomas, M., Haines, A.S., Helfrich, E.J.N., Hillwig, M.L., Ishida, K., Jones, A.C., Jones, C.S., Jungmann, K., Kegler, C., Kim, H.U., Kötter, P., Krug, D., Masschelein, J., Melnik, A.V., Mantovani, S.M., Monroe, E.A., Moore, M., Moss, N., Nützmann, H.-W., Pan, G., Pati, A., Petras, D., Reen, F.J., Rosconi, F., Rui, Z., Tian, Z., Tobias, N.J., Tsunematsu, Y., Wiemann, P., Wyckoff, E., Yan, X., Yim, G., Yu, F., Xie, Y., Aigle, B., Apel, A.K., Balibar, C.J., Balskus, E.P., Barona-Gómez, F., Bechthold, A., Bode, H.B., Borriss, R., Brady, S.F., Brakhage, A.A., Caffrey, P., Cheng, Y.-Q., Clardy, J., Cox, R.J., De Mot, R., Donadio, S., Donia, M.S., van der Donk, W.A., Dorrestein, P.C., Doyle, S., Driessen, A.J.M., Ehling-Schulz, M., Entian, K.-D., Fischbach, M.A., Gerwick, L., Gerwick, W.H., Gross, H., Gust, B., Hertweck, C., Höfte, M., Jensen, S.E., Ju, J., Katz, L., Kaysser, L., Klassen, J.L., Keller, N.P., Kormanec, J., Kuipers, O.P., Kuzuyama, T., Kyrpides, N.C., Kwon, H.-J., Lautru, S., Lavigne, R., Lee, C.Y., Linquan, B., Liu, X., Liu, W., Luzhetskyy, A., Mahmud, T., Mast, Y., Méndez, C., Metsä-Ketelä, M., Micklefield, J., Mitchell, D.A., Moore, B.S., Moreira, L.M., Müller, R., Neilan, B.A., Nett, M., Nielsen, J., O'Gara, F., Oikawa, H., Osbourn, A., Osburne, M.S., Ostash, B., Payne, S.M., Pernodet, J.-L., Petricek, M., Piel, J., Ploux, O., Raaijmakers, J.M., Salas, J.A., Schmitt, E.K., Scott, B., Seipke, R.F., Shen, B., Sherman, D.H., Sivonen, K., Smanski, M.J., Sosio, M., Stegmann, E., Süssmuth, R.D., Tahlan, K., Thomas, C.M., Tang, Y., Truman, A.W., Viaud, M., Walton, J.D., Walsh, C.T., Weber, T., van Wezel, G.P., Wilkinson, B., Willey, J.M., Wohlleben, W., Wright, G.D., Ziemert, N., Zhang, C., Zotchev, S.B., Breitling, R., Takano, E., Glöckner, F.O., 2015. Minimum Information about a Biosynthetic Gene cluster. Nature Chemical Biology 11, 625–631. https://doi.org/10.1038/nchembio.1890
9. Ziemert, N., Alanjary, M., Weber, T., 2016. The evolution of genome mining in microbes – a review. Natural Product Reports 33, 988–1005. https://doi.org/10.1039/C6NP00025H

10. Medema, M.H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M.A., Weber, T., Takano, E., Breitling, R., 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Research 39, W339–W346. https://doi.org/10.1093/nar/gkr466
11. Skinnider, M.A., Dejong, C.A., Rees, P.N., Johnston, C.W., Li, H., Webster, A.L.H., Wyatt, M.A., Magarvey, N.A., 2015. Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM). Nucleic Acids Research 43, 9645–9662. https://doi.org/10.1093/nar/gkv1012
12. Agrawal, P., Khater, S., Gupta, M., Sain, N., Mohanty, D., 2017. RiPPMiner: a bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links. Nucleic Acids Research 45, W80–W88. https://doi.org/10.1093/nar/gkx408
13. Ziemert, N., Podell, S., Penn, K., Badger, J.H., Allen, E., Jensen, P.R., 2012. The Natural Product Domain Seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. PLOS ONE 7, e34064. https://doi.org/10.1371/journal.pone.0034064
14. Robinson, S.L., Terlouw, B.R., Smith, M.D., Pidot, S.J., Stinear, T.P., Medema, M.H., Wackett, L.P., 2020. Global analysis of adenylate-forming enzymes reveals β-lactone biosynthesis pathway in pathogenic *Nocardia*. Journal of Biological Chemistry 295, 14826–14839. https://doi.org/10.1074/jbc.RA120.013528
15. Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M.E.A., Mitros, T., Richards, G.S., Conaco, C., Dacre, M., Hellsten, U., Larroux, C., Putnam, N.H., Stanke, M., Adamska, M., Darling, A., Degnan, S.M., Oakley, T.H., Plachetzki, D.C., Zhai, Y., Adamski, M., Calcino, A., Cummins, S.F., Goodstein, D.M., Harris, C., Jackson, D.J., Leys, S.P., Shu, S., Woodcroft, B.J., Vervoort, M., Kosik, K.S., Manning, G., Degnan, B.M., Rokhsar, D.S., 2010. The *Amphimedon queenslandica* genome and the evolution of animal complexity. Nature 466, 720–726. https://doi.org/10.1038/nature09201
16. Nichols, S.A., Roberts, B.W., Richter, D.J., Fairclough, S.R., King, N., 2012. Origin of metazoan cadherin diversity and the antiquity of the classical cadherin/β-catenin complex. The Proceedings of the National Academy of Sciences 109, 13046–13051. https://doi.org/10.1073/pnas.1120685109
17. Ereskovsky, A.V., Richter, D.J., Lavrov, D.V., Schippers, K.J., Nichols, S.A., 2017. Transcriptome sequencing and delimitation of sympatric *Oscarella* species (*O. carmela* and *O. pearsei* sp. nov) from California, USA. PLOS ONE 12, e0183002. https://doi.org/10.1371/journal.pone.0183002
18. Fortunato, S.A.V., Adamski, M., Ramos, O.M., Leininger, S., Liu, J., Ferrier, D.E.K., Adamska, M., 2014. Calcisponges have a ParaHox gene and dynamic expression of dispersed NK homeobox genes. Nature 514, 620–623. https://doi.org/10.1038/nature13881
19. Ryu, T., Seridi, L., Moitinho-Silva, L., Oates, M., Liew, Y.J., Mavromatis, C., Wang, X., Haywood, A., Lafi, F.F., Kupresanin, M., Sougrat, R., Alzahrani, M.A., Giles, E., Ghosheh, Y., Schunter, C., Baumgarten, S., Berumen, M.L., Gao, X., Aranda, M., Foret, S., Gough, J., Voolstra, C.R., Hentschel, U., Ravasi, T., 2016. Hologenome analysis of two marine sponges with different microbiomes. BioMed Central Genomics 17, 158. https://doi.org/10.1186/s12864-016-2501-0
20. Francis, W.R., Eitel, M., Vargas, S., Adamski, M., Haddock, S.H.D., Krebs, S., Blum, H., Erpenbeck, D., Wörheide, G., 2017. The genome of the contractile demosponge *Tethya wilhelma* and the evolution of metazoan neural signalling pathways. bioRxiv 120998. https://doi.org/10.1101/120998
21. Kenny, N.J., Plese, B., Riesgo, A., Itskovich, V.B., 2019. Symbiosis, selection, and novelty: freshwater adaptation in the unique sponges of Lake Baikal. Molecular Biology and Evolution 36, 2462–2480. https://doi.org/10.1093/molbev/msz151
22. Kenny, N.J., Francis, W.R., Rivera-Vicéns, R.E., Juravel, K., de Mendoza, A., Díez-Vives, C., Lister, R., Bezares-Calderón, L.A., Grombacher, L., Roller, M., Barlow, L.D., Camilli, S., Ryan, J.F., Wörheide, G., Hill, A.L., Riesgo, A., Leys, S.P., 2020. Tracing animal genomic evolution with the chromosomal-level assembly of the freshwater sponge *Ephydatia muelleri*. Nature Communications 11, 3676. https://doi.org/10.1038/s41467-020-17397-w

23. Eder, C., Proksch, P., Wray, V., Steube, K., Bringmann, G., van Soest, R.W.M., Sudarsono, Ferdinandus, E., Pattisina, L.A., Wiryowidagdo, S., Moka, W., 1999. New Alkaloids from the Indopacific sponge *Stylissa carteri*. Journal of Natural Products 62, 184–187. https://doi.org/10.1021/np980315g
24. Bourguet-Kondracki, M.L., Rakotoarisoa, M.T., Martin, M.T., Guyot, M., 1992. Bioactive bromopolyacetylenes from the marine sponge *Xestospongia testudinaria*. Tetrahedron Letters 33, 225–226. https://doi.org/10.1016/0040-4039(92)88056-B
25. Brinkmann, C.M., Marker, A., Kurtböke, D.İ., 2017. An overview on marine sponge-symbiotic bacteria as unexhausted sources for natural Product discovery. Diversity 9, 40. https://doi.org/10.3390/d9040040
26. Nakashima, Y., Egami, Y., Kimura, M., Wakimoto, T., Abe, I., 2016. Metagenomic analysis of the sponge *Discodermia* reveals the production of the cyanobacterial natural product Kasumigamide by '*Entotheonella*.' PLOS ONE 11, e0164468. https://doi.org/10.1371/journal.pone.0164468
27. Mori, T., Cahn, J.K.B., Wilson, M.C., Meoded, R.A., Wiebach, V., Martinez, A.F.C., Helfrich, E.J.N., Albersmeier, A., Wibberg, D., Dätwyler, S., Keren, R., Lavy, A., Rückert, C., Ilan, M., Kalinowski, J., Matsunaga, S., Takeyama, H., Piel, J., 2018. Single-bacterial genomics validates rich and varied specialized metabolism of uncultivated *Entotheonella* sponge symbionts. The Proceedings of the National Academy of Sciences 115, 1718–1723. https://doi.org/10.1073/pnas.1715496115
28. Hugerth, L.W., Larsson, J., Alneberg, J., Lindh, M.V., Legrand, C., Pinhassi, J., Andersson, A.F., 2015. Metagenome-assembled genomes uncover a global brackish microbiome. Genome Biology 16, 279. https://doi.org/10.1186/s13059-015-0834-7
29. Tribalat, M.-A., Marra, M.V., McCormack, G., Thomas, O., 2016. Does the chemical diversity of the order Haplosclerida (phylum Porifera: class Demospongia) fit with current taxonomic classification? Planta Medica. https://doi.org/10.1055/s-0042-105879
30. Waters, A.L., Peraud, O., Kasanah, N., Sims, J.W., Kothalawala, N., Anderson, M.A., Abbas, S.H., Rao, K.V., Jupally, V.R., Kelly, M., Dass, A., Hill, R.T., Hamann, M.T., 2014. An analysis of the sponge *Acanthostrongylophora igens*' microbiome yields an actinomycete that produces the natural product manzamine A. Frontiers in Marine Science 1. https://doi.org/10.3389/fmars.2014.00054
31. Cutignano, A., Tramice, A., Caro, S.D., Villani, G., Cimino, G., Fontana, A., 2003. Biogenesis of 3-alkylpyridine alkaloids in the marine mollusc *Haminoea Orbignyana*. Angewandte Chemie 115, 2737–2740. https://doi.org/10.1002/ange.200250642
32. Cutignano, A., Cimino, G., Giordano, A., d'Ippolito, G., Fontana, A., 2004. Polyketide origin of 3-alkylpyridines in the marine mollusc *Haminoea orbignyana*. Tetrahedron Letters 45, 2627–2629. https://doi.org/10.1016/j.tetlet.2004.01.138
33. Itoh, T., Tokunaga, K., Matsuda, Y., Fujii, I., Abe, I., Ebizuka, Y., Kushiro, T., 2010. Reconstitution of a fungal meroterpenoid biosynthesis reveals the involvement of a novel family of terpene cyclases. Nature Chemistry 2, 858–864. https://doi.org/10.1038/nchem.764
34. Yaegashi, J., Romsdahl, J., Chiang, Y.-M., Wang, C.C.C., 2015. Genome mining and molecular characterization of the biosynthetic gene cluster of a diterpenic meroterpenoid, 15-deoxyoxalicine B, in *Penicillium canescens*. Chemical Science 6, 6537–6544. https://doi.org/10.1039/C5SC01965F
35. Myronovskyi, M., Rosenkränzer, B., Nadmid, S., Pujic, P., Normand, P., Luzhetskyy, A., 2018. Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters. Metabolic Engineering 49, 316–324. https://doi.org/10.1016/j.ymben.2018.09.004
36. Ma, H.-M., Zhou, Q., Tang, Y.-M., Zhang, Z., Chen, Y.-S., He, H.-Y., Pan, H.-X., Tang, M.-C., Gao, J.-F., Zhao, S.-Y., Igarashi, Y., Tang, G.-L., 2013. Unconventional origin and hybrid system for construction of pyrrolopyrrole moiety in Kosinostatin Biosynthesis. Chemistry & Biology 20, 796–805. https://doi.org/10.1016/j.chembiol.2013.04.013
37. Marra, M.V., 2019. Investigation of biological factors that may contribute to bioactivity in *Haliclona* (Porifera, Haplosclerida).

38. Kolmogorov, M., Bickhart, D.M., Behsaz, B., Gurevich, A., Rayko, M., Shin, S.B., Kuhn, K., Yuan, J., Polevikov, E., Smith, T.P.L., Pevzner, P.A., 2020. metaFlye: scalable long-read metagenome assembly using repeat graphs. Nature Methods 17, 1103–1110. https://doi.org/10.1038/s41592-020-00971-x
39. Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890. https://doi.org/10.1093/bioinformatics/bty560
40. Vaser, R., Šikić, M., 2021. Time- and memory-efficient genome assembly with Raven. Nature Computational Science 1, 332–336. https://doi.org/10.1038/s43588-021-00073-4
41. Kolmogorov, M., Yuan, J., Lin, Y., Pevzner, P.A., 2019. Assembly of long, error-prone reads using repeat graphs. Nature Biotechnology 37, 540–546. https://doi.org/10.1038/s41587-019-0072-8
42. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Research 27, 722–736. https://doi.org/10.1101/gr.215087.116
43. 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, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351a
44. Kundu, R., Casey, J., Sung, W.-K., 2019. HyPo: super fast & accurate polisher for long read genome assemblies. https://doi.org/10.1101/2019.12.19.882506
45. Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., Korobeynikov, A., 2020. Using SPAdes *de novo* assembler. Current Protocols in Bioinformatics 70, e102. https://doi.org/10.1002/cpbi.102
46. Challis, R., Richards, E., Rajan, J., Cochrane, G., Blaxter, M., 2020. BlobToolKit – interactive quality assessment of genome assemblies. G3 Genes|Genomes|Genetics 10, 1361–1374. https://doi.org/10.1534/g3.119.400908
47. Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., Wang, Z., 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 7, e7359. https://doi.org/10.7717/peerj.7359
48. Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research 25, 1043–1055. https://doi.org/10.1101/gr.186072.114
49. Lagesen, K., Hallin, P., Rødland, E.A., Stærfeldt, H.-H., Rognes, T., Ussery, D.W., 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Research 35, 3100–3108. https://doi.org/10.1093/nar/gkm160
50. Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., Smit, A.F., 2020. RepeatModeler2 for automated genomic discovery of transposable element families. The Proceedings of the National Academy of Sciences 117, 9451–9457. https://doi.org/10.1073/pnas.1921046117
51. Brůna, T., Hoff, K.J., Lomsadze, A., Stanke, M., Borodovsky, M., 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genomics and Bioinformatics 3. https://doi.org/10.1093/nargab/lqaa108
52. Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C., Bork, P., 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. Molecular Biology and Evolution 34, 2115–2122. https://doi.org/10.1093/molbev/msx148
53. Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., von Mering, C., Bork, P., 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Research 47, D309–D314. https://doi.org/10.1093/nar/gky1085
54. Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M.H., Weber, T., 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Research 47, W81–W87. https://doi.org/10.1093/nar/gkz310

55. Breese, M.R., Liu, Y., 2013. NGSUtils: a software suite for analyzing and manipulating next-generation sequencing datasets. Bioinformatics 29, 494–496. https://doi.org/10.1093/bioinformatics/bts731
56. Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191
57. Wu, T.D., Reeder, J., Lawrence, M., Becker, G., Brauer, M.J., 2016. GMAP and GSNAP for genomic sequence alignment: enhancements to speed, accuracy, and functionality, in: Mathé, E., Davis, S. (Eds.), Statistical Genomics: Methods and Protocols, Methods in Molecular Biology. Springer, New York, NY, pp. 283–334. https://doi.org/10.1007/978-1-4939-3578-9_15
58. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nature Biotechnology 29, 24–26. https://doi.org/10.1038/nbt.1754
59. Kautsar, S.A., Suarez Duran, H.G., Blin, K., Osbourn, A., Medema, M.H., 2017. plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. Nucleic Acids Research 45, W55–W63. https://doi.org/10.1093/nar/gkx305
60. Lu, S., Wang, J., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Marchler, G.H., Song, J.S., Thanki, N., Yamashita, R.A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C.J., Marchler-Bauer, A., 2020. CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Research 48, D265–D268. https://doi.org/10.1093/nar/gkz991
61. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., Finn, R.D., Bateman, A., 2021. Pfam: The protein families database in 2021. Nucleic Acids Research 49, D412–D419. https://doi.org/10.1093/nar/gkaa913
62. Madeira, F., Park, Y. mi, Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., Lopez, R., 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Research 47, W636–W641. https://doi.org/10.1093/nar/gkz268
63. de Rond, T., Asay, J.E., Moore, B.S., 2021. Co-occurrence of enzyme domains guides the discovery of an oxazolone synthetase. Nature Chemical Biology 17, 794–799. https://doi.org/10.1038/s41589-021-00808-4
64. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research 13, 2498–2504. https://doi.org/10.1101/gr.1239303
65. Desplat, Y., Warner, J.F., Lopez, J.V., 2022. Holo-transcriptome sequences from the tropical marine sponge *Cinachyrella alloclada*. Journal of Heredity 113, 184–187. https://doi.org/10.1093/jhered/esab075
66. Plese, B., Kenny, N.J., Rossi, M.E., Cárdenas, P., Schuster, A., Taboada, S., Koutsouveli, V., Riesgo, A., 2021. Mitochondrial evolution in the Demospongiae (Porifera): phylogeny, divergence time, and genome biology. Molecular Phylogenetics and Evolution 155, 107011. https://doi.org/10.1016/j.ympev.2020.107011
67. Koutsouveli, V., Cárdenas, P., Santodomingo, N., Marina, A., Morato, E., Rapp, H.T., Riesgo, A., 2020. The molecular machinery of gametogenesis in *Geodia* Demosponges (Porifera): evolutionary origins of a conserved toolkit across animals. Molecular Biology and Evolution 37, 3485–3506. https://doi.org/10.1093/molbev/msaa183
68. Finoshin, A.D., Adameyko, K.I., Mikhailov, K.V., Kravchuk, O.I., Georgiev, A.A., Gornostaev, N.G., Kosevich, I.A., Mikhailov, V.S., Gazizova, G.R., Shagimardanova, E.I., Gusev, O.A., Lyupina, Y.V., 2020. Iron metabolic pathways in the processes of sponge plasticity. PLOS ONE 15, e0228722. https://doi.org/10.1371/journal.pone.0228722
69. Manousaki, T., Koutsouveli, V., Lagnel, J., Kollias, S., Tsigenopoulos, C.S., Arvanitidis, C., Magoulas, A., Dounas, C., Dailianis, T., 2019. A *de novo* transcriptome assembly for the bath sponge *Spongia officinalis*, adjusting for microsymbionts. BioMed Central Research Notes 12, 813. https://doi.org/10.1186/s13104-019-4843-6

70. González-Aravena, M., Kenny, N.J., Osorio, M., Font, A., Riesgo, A., Cárdenas, C.A., 2019. Warm temperatures, cool sponges: the effect of increased temperatures on the Antarctic sponge *Isodictya* sp. PeerJ 7, e8088. https://doi.org/10.7717/peerj.8088
71. Leiva, C., Taboada, S., Kenny, N.J., Combosch, D., Giribet, G., Jombart, T., Riesgo, A., 2019. Population substructure and signals of divergent adaptive selection despite admixture in the sponge *Dendrilla antarctica* from shallow waters surrounding the Antarctic Peninsula. Molecular Ecology 28, 3151–3170. https://doi.org/10.1111/mec.15135
72. Pita, L., Hoeppner, M.P., Ribes, M., Hentschel, U., 2018. Differential expression of immune receptors in two marine sponges upon exposure to microbial-associated molecular patterns. Scientific Reports 8, 16081. https://doi.org/10.1038/s41598-018-34330-w
73. Revilla-i-Domingo, R., Schmidt, C., Zifko, C., Raible, F., 2018. Establishment of transgenesis in the Demosponge *Suberites domuncula*. Genetics 210, 435–443. https://doi.org/10.1534/genetics.118.301121
74. Kenny, N.J., de Goeij, J.M., de Bakker, D.M., Whalen, C.G., Berezikov, E., Riesgo, A., 2018. Towards the identification of ancestrally shared regenerative mechanisms across the metazoa: t transcriptomic case study in the Demosponge *Halisarca caerulea*. Marine Genomics 37, 135–147. https://doi.org/10.1016/j.margen.2017.11.001
75. Leys, S., 2017. *Ephydatia muelleri* Trinity transcriptome [WWW Document]. Education & Research Archive. https://doi.org/10.7939/R3WH2DV20
76. Leys, S., 2017. *Eunapius fragilis* Trinity transcriptome [WWW Document]. Education & Research Archive. https://doi.org/10.7939/R3794177K
77. Simion, P., Philippe, H., Baurain, D., Jager, M., Richter, D.J., Di Franco, A., Roure, B., Satoh, N., Quéinnec, É., Ereskovsky, A., Lapébie, P., Corre, E., Delsuc, F., King, N., Wörheide, G., Manuel, M., 2017. A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. Current Biology 27, 958–967. https://doi.org/10.1016/j.cub.2017.02.031
78. Díez-Vives, C., Moitinho-Silva, L., Nielsen, S., Reynolds, D., Thomas, T., 2017. Expression of eukaryotic-like protein in the microbiome of sponges. Molecular Ecology 26, 1432–1451. https://doi.org/10.1111/mec.14003
79. Borisenko, I., Adamski, M., Ereskovsky, A., Adamska, M., 2016. Surprisingly rich repertoire of Wnt genes in the demosponge *Halisarca dujardini*. BioMed Central Evolutionary Biology 16, 123. https://doi.org/10.1186/s12862-016-0700-6
80. Guzman, C., Conaco, C., 2016. Comparative transcriptome analysis reveals insights into the streamlined genomes of haplosclerid demosponges. Scientific Reports 6, 18774. https://doi.org/10.1038/srep18774
81. Alié, A., Hayashi, T., Sugimura, I., Manuel, M., Sugano, W., Mano, A., Satoh, N., Agata, K., Funayama, N., 2015. The ancestral gene repertoire of animal stem cells. The Proceedings of the National Academy of Sciences USA 112, E7093–E7100. https://doi.org/10.1073/pnas.1514789112
82. Qiu, F., Ding, S., Ou, H., Wang, D., Chen, J., Miyamoto, M.M., 2015. Transcriptome changes during the life cycle of the red sponge, *Mycale phyllophila* (Porifera, Demospongiae, Poecilosclerida). Genes 6, 1023–1052. https://doi.org/10.3390/genes6041023
83. Whelan, N.V., Kocot, K.M., Moroz, L.L., Halanych, K.M., 2015. Error, signal, and the placement of Ctenophora sister to all other animals. The Proceedings of the National Academy of Sciences 112, 5773–5778. https://doi.org/10.1073/pnas.1503453112
84. Riesgo, A., Peterson, K., Richardson, C., Heist, T., Strehlow, B., McCauley, M., Cotman, C., Hill, M., Hill, A., 2014. Transcriptomic analysis of differential host gene expression upon uptake of symbionts: a case study with *Symbiodinium* and the major bioeroding sponge *Cliona varians*. BioMed Central Genomics 15, 376. https://doi.org/10.1186/1471-2164-15-376
85. Riesgo, A., Farrar, N., Windsor, P.J., Giribet, G., Leys, S.P., 2014. The analysis of eight transcriptomes from all poriferan classes reveals surprising genetic complexity in sponges. Molecular Biology and Evolution 31, 1102–1120. https://doi.org/10.1093/molbev/msu057

86. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of Molecular Biology 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
87. Grabherr, M.G., Haas, 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., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nature Biotechnology 29, 644–652. https://doi.org/10.1038/nbt.1883
88. Wick, R.R., Schultz, M.B., Zobel, J., Holt, K.E., 2015. Bandage: interactive visualization of *de novo* genome assemblies. Bioinformatics 31, 3350–3352. https://doi.org/10.1093/bioinformatics/btv383
89. Mistry, J., Finn, R.D., Eddy, S.R., Bateman, A., Punta, M., 2013. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Research 41, e121–e121. https://doi.org/10.1093/nar/gkt263
90. Stachelhaus, T., Mootz, H.D., Marahiel, M.A., 1999. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chemistry & Biology 6, 493–505. https://doi.org/10.1016/S1074-5521(99)80082-9
91. Kudo, F., Miyanaga, A., Eguchi, T., 2019. Structural basis of the nonribosomal codes for nonproteinogenic amino acid selective adenylation enzymes in the biosynthesis of natural products. Journal of Industrial Microbiology and Biotechnology 46, 515–536. https://doi.org/10.1007/s10295-018-2084-7
92. Pruitt, K.D., Tatusova, T., Maglott, D.R., 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Research 35, D61–D65. https://doi.org/10.1093/nar/gkl842
93. Kautsar, S.A., Blin, K., Shaw, S., Navarro-Muñoz, J.C., Terlouw, B.R., van der Hooft, J.J.J., van Santen, J.A., Tracanna, V., Suarez Duran, H.G., Pascal Andreu, V., Selem-Mojica, N., Alanjary, M., Robinson, S.L., Lund, G., Epstein, S.C., Sisto, A.C., Charkoudian, L.K., Collemare, J., Linington, R.G., Weber, T., Medema, M.H., 2020. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. Nucleic Acids Research 48, D454–D458. https://doi.org/10.1093/nar/gkz882
94. Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 30, 772–780. https://doi.org/10.1093/molbev/mst010
95. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015
96. Letunic, I., Bork, P., 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Research 49, W293–W296. https://doi.org/10.1093/nar/gkab301
97. Poust, S., Yoon, I., Adams, P.D., Katz, L., Petzold, C.J., Keasling, J.D., 2014. Understanding the role of histidine in the GHSxG acyltransferase active site motif: evidence for histidine stabilization of the malonyl-enzyme intermediate. PLOS ONE 9, e109421. https://doi.org/10.1371/journal.pone.0109421
98. Freeman, C.J., Gleason, D.F., Ruzicka, R., McFall, G., n.d. A biogeographic comparison of sponge fauna from Gray's Reef National Marine Sanctuary and other hard-bottom reefs of coastal Georgia, U.S.A. 7.
99. Roach, M.J., Schmidt, S.A., Borneman, A.R., 2018. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BioMed Central Bioinformatics 19, 460. https://doi.org/10.1186/s12859-018-2485-7
100. Panova, M., Aronsson, H., Cameron, R.A., Dahl, P., Godhe, A., Lind, U., Ortega-Martinez, O., Pereyra, R., Tesson, S.V.M., Wrange, A.-L., Blomberg, A., Johannesson, K., 2016. DNA extraction protocols for whole-genome sequencing in marine organisms, in: Bourlat, S.J. (Ed.), Marine Genomics: Methods and Protocols, Methods in Molecular Biology. Springer, New York, NY, pp. 13–44. https://doi.org/10.1007/978-1-4939-3774-5_2

101. Robinson, S.L., Piel, J., Sunagawa, S., 2021. A roadmap for metagenomic enzyme discovery. Natural Products Reports https://doi.org/10.1039/D1NP00006C
102. Albrizio, S., Ciminiello, P., Fattorusso, E., Magno, S., Pawlik, J.R., 1995. Amphitoxin, a new high molecular weight antifeedant pyridinium salt from the Caribbean sponge *Amphimedon compressa*. Journal of Natural Products 58, 647–652. https://doi.org/10.1021/np50119a002
103. Kelman, D., Kashman, Y., Rosenberg, E., Ilan, M., Ifrach, I., Loya, Y., 2001. Antimicrobial activity of the reef sponge *Amphimedon viridis* from the Red Sea: evidence for selective toxicity. Aquatic Microbial Ecology 24, 9–16. https://doi.org/10.3354/ame024009
104. Matsuno, T., 2001. Aquatic animal carotenoids. Fisheries science 67, 771–783. https://doi.org/10.1046/j.1444-2906.2001.00323.x
105. Moniz, M.B.J., Rindi, F., Stephens, K., Maggi, E., Collins, P., McCormack, G.P., 2013. Composition and temporal variation of the algal assemblage associated with the haplosclerid sponge *Haliclona indistincta* (Bowerbank). Aquatic Botany 105, 50–53. https://doi.org/10.1016/j.aquabot.2012.12.002
106. Young, A.J., Lowe, G.L., 2018. Carotenoids—antioxidant properties. Antioxidants 7, 28. https://doi.org/10.3390/antiox7020028
107. Freeman, M.F., Vagstad, A.L., Piel, J., 2016. Polytheonamide biosynthesis showcasing the metabolic potential of sponge-associated uncultivated '*Entotheonella*' bacteria. Current Opinion in Chemical Biology, Biocatalysis and biotransformation bioinorganic chemistry 31, 8–14. https://doi.org/10.1016/j.cbpa.2015.11.002
108. Kenney, G.E., Dassama, L.M.K., Pandelia, M.-E., Gizzi, A.S., Martinie, R.J., Gao, P., DeHart, C.J., Schachner, L.F., Skinner, O.S., Ro, S.Y., Zhu, X., Sadek, M., Thomas, P.M., Almo, S.C., Bollinger, J.M., Krebs, C., Kelleher, N.L., Rosenzweig, A.C., 2018. The biosynthesis of methanobactin. Science 359, 1411–1416. https://doi.org/10.1126/science.aap9437
109. G. Arnison, P., J. Bibb, M., Bierbaum, G., A. Bowers, A., S. Bugni, T., Bulaj, G., A. Camarero, J., J. Campopiano, D., L. Challis, G., Clardy, J., D. Cotter, P., J. Craik, D., Dawson, M., Dittmann, E., Donadio, S., C. Dorrestein, P., Entian, K.-D., A. Fischbach, M., S. Garavelli, J., Göransson, U., W. Gruber, C., H. Haft, D., K. Hemscheidt, T., Hertweck, C., Hill, C., R. Horswill, A., Jaspars, M., L. Kelly, W., P. Klinman, J., P. Kuipers, O., James Link, A., Liu, W., A. Marahiel, M., A. Mitchell, D., N. Moll, G., S. Moore, B., Müller, R., K. Nair, S., F. Nes, I., E. Norris, G., M. Olivera, B., Onaka, H., L. Patchett, M., Piel, J., T. Reaney, M.J., Rebuffat, S., Paul Ross, R., Sahl, H.-G., W. Schmidt, E., E. Selsted, M., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., D. Süssmuth, R., R. Tagg, J., Tang, G.-L., W. Truman, A., C. Vederas, J., T. Walsh, C., D. Walton, J., C. Wenzel, S., M. Willey, J., Donk, W.A. van der, 2013. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Natural Product Reports 30, 108–160. https://doi.org/10.1039/C2NP20085F
110. Gold, D.A., O'Reilly, S.S., Watson, J., Degnan, B.M., Degnan, S.M., Krömer, J.O., Summons, R.E., 2017. Lipidomics of the sponge *Amphimedon queenslandica* and implication for biomarker geochemistry. Geobiology 15, 836–843. https://doi.org/10.1111/gbi.12253
111. Gerdol, M., Sollitto, M., Pallavicini, A., Castellano, I., 2019. The complex evolutionary history of sulfoxide synthase in ovothiol biosynthesis. Proceedings of the Royal Society B: Biological Sciences 286, 20191812. https://doi.org/10.1098/rspb.2019.1812
112. Steffen, K., Laborde, Q., Gunasekera, S., Payne, C.D., Rosengren, K.J., Riesgo, A., Göransson, U., Cárdenas, P., 2021. Barrettides: a peptide family specifically produced by the deep-sea sponge *Geodia barretti*. Journal of Natural Products 84, 3138–3146. https://doi.org/10.1021/acs.jnatprod.1c00938
113. Tianero, M.D., Balaich, J.N., Donia, M.S., 2019. Localized production of defence chemicals by intracellular symbionts of *Haliclona* sponges. Nature Microbiology 4, 1149–1159. https://doi.org/10.1038/s41564-019-0415-8
114. Torres, J.P., Schmidt, E.W., 2019. The biosynthetic diversity of the animal world. Journal of Biological Chemistry 294, 17684–17692. https://doi.org/10.1074/jbc.REV119.006130

115. Cooke, T.F., Fischer, C.R., Wu, P., Jiang, T.-X., Xie, K.T., Kuo, J., Doctorov, E., Zehnder, A., Khosla, C., Chuong, C.-M., Bustamante, C.D., 2017. Genetic mapping and biochemical basis of yellow feather pigmentation in budgerigars. Cell 171, 427-439.e21. https://doi.org/10.1016/j.cell.2017.08.016
116. Sabatini, M., Comba, S., Altabe, S., Recio-Balsells, A.I., Labadie, G.R., Takano, E., Gramajo, H., Arabolaza, A., 2018. Biochemical characterization of the minimal domains of an iterative eukaryotic polyketide synthase. The Federation of European Biochemical Societies Journal 285, 4494–4511. https://doi.org/10.1111/febs.14675
117. Torres, J.P., Lin, Z., Winter, J.M., Krug, P.J., Schmidt, E.W., 2020. Animal biosynthesis of complex polyketides in a photosynthetic partnership. Nature Communications 11, 2882. https://doi.org/10.1038/s41467-020-16376-5
118. Li, F., Lin, Z., Torres, J.P., Hill, E.A., Li, D., Townsend, C.A., Schmidt, E.W., 2022. Sea urchin polyketide synthase SpPks1 produces the naphthalene precursor to echinoderm pigments. The Journal of the American Chemical Society 144, 9363–9371. https://doi.org/10.1021/jacs.2c01416
119. Shou, Q., Feng, L., Long, Y., Han, J., Nunnery, J.K., Powell, D.H., Butcher, R.A., 2016. A hybrid polyketide–nonribosomal peptide in nematodes that promotes larval survival. Nature Chemical Biology 12, 770–772. https://doi.org/10.1038/nchembio.2144
120. Wang, P., Meng, F., Moore, B.M., Shiu, S.-H., 2021. Impact of short-read sequencing on the misassembly of a plant genome. BioMed Central Genomics 22, 99. https://doi.org/10.1186/s12864-021-07397-5
121. Kumar, N., Lin, M., Zhao, X., Ott, S., Santana-Cruz, I., Daugherty, S., Rikihisa, Y., Sadzewicz, L., Tallon, L.J., Fraser, C.M., Dunning Hotopp, J.C., 2016. Efficient enrichment of bacterial mRNA from host-bacteria total RNA Samples. Scientific Reports 6, 34850. https://doi.org/10.1038/srep34850
122. Tschopp, J., Martinon, F., Burns, K., 2003. NALPs: a novel protein family involved in inflammation. Nature Reviews Molecular Cell Biology 4, 95–104. https://doi.org/10.1038/nrm1019
123. Ng, A.C.Y., Eisenberg, J.M., Heath, R.J.W., Huett, A., Robinson, C.M., Nau, G.J., Xavier, R.J., 2011. Human leucine-rich repeat proteins: a genome-wide bioinformatic categorization and functional analysis in innate immunity. The Proceedings of the National Academy of Sciences 108, 4631–4638. https://doi.org/10.1073/pnas.1000093107
124. A. Herbst, D., A. Townsend, C., Maier, T., 2018. The architectures of iterative type I PKS and FAS. Natural Product Reports 35, 1046–1069. https://doi.org/10.1039/C8NP00039E
125. Crawford, J.M., Dancy, B.C.R., Hill, E.A., Udwary, D.W., Townsend, C.A., 2006. Identification of a starter unit acyl-carrier protein transacylase domain in an iterative type I polyketide synthase. The Proceedings of the National Academy of Sciences 103, 16728–16733. https://doi.org/10.1073/pnas.0604112103
126. Crawford, J.M., Korman, T.P., Labonte, J.W., Vagstad, A.L., Hill, E.A., Kamari-Bidkorpeh, O., Tsai, S.-C., Townsend, C.A., 2009. Structural basis for biosynthetic programming of fungal aromatic polyketide cyclization. Nature 461, 1139–1143. https://doi.org/10.1038/nature08475
127. Zou, Y., Yin, H., Kong, D., Deng, Z., Lin, S., 2013. A *trans*-acting ketoreductase in biosynthesis of a symmetric polyketide dimer SIA7248. ChemBioChem 14, 679–683. https://doi.org/10.1002/cbic.201300068
128. Coates, R.C., Podell, S., Korobeynikov, A., Lapidus, A., Pevzner, P., Sherman, D.H., Allen, E.E., Gerwick, L., Gerwick, W.H., 2014. Characterization of cyanobacterial hydrocarbon composition and distribution of biosynthetic pathways. PLoS One 9, e85140. https://doi.org/10.1371/journal.pone.0085140
129. Chen, H., O'Connor, S., Cane, D.E., Walsh, C.T., 2001. Epothilone biosynthesis: assembly of the methylthiazolylcarboxy starter unit on the EpoB subunit. Chemistry & Biology 8, 899–912. https://doi.org/10.1016/S1074-5521(01)00064-3
130. Farmer, R., Thomas, C.M., Winn, P.J., 2019. Structure, function and dynamics in acyl carrier proteins. PLOS ONE 14, e0219435. https://doi.org/10.1371/journal.pone.0219435

131. Dekimpe, S., Masschelein, J., 2021. Beyond peptide bond formation: the versatile role of condensation domains in natural product biosynthesis. Natural Products Reports https://doi.org/10.1039/D0NP00098A
132. Reimer, J.M., Eivaskhani, M., Harb, I., Guarné, A., Weigt, M., Schmeing, T.M., 2019. Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. Science 366. https://doi.org/10.1126/science.aaw4388
133. Colabroy, K.L., Begley, T.P., 2005. The Pyridine ring of NAD is formed by a nonenzymatic pericyclic reaction. The Journal of the American Chemical Society 127, 840–841. https://doi.org/10.1021/ja0446395
134. Tribalat, M.-A., 2016. Métabolismes spécialisés d'éponges méditerranéennes du genre *Haliclona* Grant, 1836 205.
135. Müller, W.E.G., Kasueske, M., Wang, X., Schröder, H.C., Wang, Y., Pisignano, D., Wiens, M., 2008. Luciferase a light source for the silica-based optical waveguides (spicules) in the demosponge *Suberites domuncula*. Cellular and Molecular Life Sciences 66, 537. https://doi.org/10.1007/s00018-008-8492-5
136. Hamed, A.N.E., Schmitz, R., Bergermann, A., Totzke, F., Kubbutat, M., Müller, W.E.G., Youssef, D.T.A., Bishr, M.M., Kamel, M.S., Edrada-Ebel, R., Wätjen, W., Proksch, P., 2018. Bioactive pyrrole alkaloids isolated from the Red Sea: marine sponge *Stylissa carteri*. Zeitschrift für Naturforschung C 73, 199–210. https://doi.org/10.1515/znc-2017-0161
137. O'Brien, R.V., Davis, R.W., Khosla, C., Hillenmeyer, M.E., 2014. Computational identification and analysis of orphan assembly-line polyketide synthases. The Journal of Antibiotics 67, 89–97. https://doi.org/10.1038/ja.2013.125
138. Jackson, D.J., Macis, L., Reitner, J., Wörheide, G., 2011. A horizontal gene transfer supported the evolution of an early metazoan biomineralization strategy. BioMed Central Evolutionary Biology 11, 238. https://doi.org/10.1186/1471-2148-11-238
139. Fontana, A., 2006. Biogenetic proposals and biosynthetic studies on secondary metabolites of opisthobranch molluscs, in: Cimino, G., Gavagnin, M. (Eds.), Molluscs: from chemo-ecological study to biotechnological application, Progress in molecular and subcellular biology. Springer, Berlin, Heidelberg, pp. 303–332. https://doi.org/10.1007/978-3-540-30880-5_14

Chapter 4

Heterologous Expression of a Novel Metazoan Megasynthase Gene

Part of this chapter contributed to the as of yet unpublished manuscript:
Sandoval, K., McCormack, G.P., 2022. The *Haliclona indistincta* Genome reveals a New Family of Metazoan Hybrid Nonribosomal Peptide Synthetase-Polyketide Synthase Genes.

Introduction

Usually, but not always, the steps responsible for producing natural products (NPs) are catalysed by biosynthetic enzymes. In turn, these enzymes are encoded for by genes in the genomes of the organisms from which the NPs originate from. Often, these genes are clustered together in regions of the genome known as biosynthetic gene clusters (BGCs) [1]. As the price of next-generation sequencing has drastically decreased, the genomic information, and thus biosynthetic genes, of diverse organisms have become increasingly accessible to laboratories worldwide [2]. Predictive software based on qualities such as sequence homology have paved the way for the rapid identification of biosynthetic genes when provided with the sequences of an organism of interest [3]. This has allowed for the efficient screening of genes responsible for several well-studied families of NPs such as polyketides, nonribosomal peptides, ribosomally synthesized and post-translationally modified peptides, and terpenes. While advances have been made to computationally predict what compounds are produced by these genes, often these methods do not work for unorthodox biosynthetic genes with low homology to those originally used to develop said computational tools [4]. This is especially true regarding biosynthetic genes from lesser studied organisms such as animals. To truly connect a NP to its responsible gene(s), physical evidence derived from the laboratory must be provided. Heterologous expression is a technique in which a non-native gene is expressed in a host organism different from that which the gene was derived from. Often, this technique is employed for the industrial production of pharmaceutically and biotechnologically relevant proteins [5]. However, of greater relevance to the field of NPs is the employment of this method to establish the connection between a NP and its corresponding gene [6]. In tandem with the affordability of DNA sequencing, the synthesis of DNA has made the acquisition of biosynthetic genes of interest a more accessible feat [7].

Before heterologous expression can be attempted, the gene(s) of interest must first be cloned into a suitable expression vector. Regarding NPs research, samples such as environmental DNA often harbour diverse microbial communities which in turn possess BGCs of potential interest. Early attempts to access the BGCs of both cultivable and uncultivable microorganisms were performed using clone library-based techniques. This involved random, fragmented genomic DNA from a sample being cloned into *Escherichia coli* and then BGCs of interest being screened for in the produced library. Specific examples of this technique include the use of cosmid, fosmid, bacterial artificial chromosome, P1 artificial chromosome, and fungal artificial chromosome libraries [8-12]. While these methods have allowed for the characterization of numerous BGCs, the increasingly available genomic information of diverse organisms has enabled more precise methodology. For one, the *in vitro* assembly of known BGC fragments produced by techniques such as PCR represent a more efficient alternative that directly targets said BGC of interest. One of the most popular of these methods is the Gibson assembly which utilizes a 5’ exonuclease, a DNA polymerase

and a DNA ligase to assemble larger DNA molecules from smaller fragments in a single isothermal reaction [13]. The Gibson assembly has particularly been utilized for the tandem reconstruction of smaller BGCs and their cloning into a desired vector [14]. Alternatively, *in vivo* methods such as homologous recombination in yeast have been utilized to assemble larger BGCs than can be accomplished with the Gibson assembly [15]. Finally, assembly of a BGC has been avoided using direct cloning methods such as transformation-associated recombination in *Saccharomyces cerevisiae*. In this way, fully intact BGCs have been successfully captured from the native organism and cloned directly into vectors specific to this process [16]. In the past all these methods involved acquiring the gene(s) of interest directly from DNA samples, but advances in DNA synthesis have allowed this step to be circumvented.

Following cloning of a gene or BGC of interest into an expression vector, transformation of a suitable heterologous host is necessary. A great number of both prokaryotic and eukaryotic systems have been developed and utilized for bacterial and fungal BGCs. Because they are a prolific source of specialized metabolites, Actinobacteria have been a popular target for the development of heterologous hosts for such a purpose. In particular, the genus *Streptomyces* is a popular one for this role with well-studied species, such as *S. coelicolor*, being repurposed *via* methods such as deletion of the native BGCs and introduction of mutations to enhance specialized metabolite production [17]. *Escherichia coli*, a common workhorse for the expression of heterologous proteins, has also been engineered to express biosynthetic genes as numerous strains, such as BAP1, have been designed to work with polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) enzymes [18]. This has allowed the bypassing of several limitations of Actinobacteria such as slow growth rates. Logically, fungal and other eukaryotic BGCs are instead often expressed in fungal systems as they are more suited to the task. Like *E. coli*, the baker's yeast *S. cerevisiae* is routinely used for the expression of heterologous proteins and has also been repurposed to express BGCs. For example, *S. cerevisiae* strain BJ5464-NpgA is engineered to possess a chromosomal copy of the *Aspergillus nidulans* phosphopantetheinyl transferase gene npgA which is more efficient in the pantetheinylation of PKSs than that of yeast [19]. However, several qualities such as differences in intron processing between yeast and filamentous fungi have led to the latter also being developed for the expression of fungal BGCs [12]. Once more, *A. nidulans* serves as an example of an organism with the genetic capabilities suitable for the expression of BGCs as it is one of the many aforementioned filamentous fungi which has been engineered for this purpose [20].

Like their microbial counterparts, metazoans also possess genes responsible for the biosynthesis of a wide variety of NPs [21]. However, in contrast to microbial BGCs the clustering of these metazoan biosynthetic genes is not as prevalent as seen with the seven nematode genes responsible for the biosynthesis of the nemamides and how they are located across four separate chromosomes [22]. Initially these metazoan biosynthetic genes were

only predictively determined *via in silico* analysis using genome mining techniques [23]. More recently, several types of metazoan specialized metabolites have been connected to their respective genes *via* heterologous expression. One of the first successful examples was the determination that sacoglossan polypropionates are the partial product of an iterative polyketide synthase encoded in the genome of the mollusc *Elysia chlorotica* by utilizing a strain of *S. cerevisiae* as a heterologous host [24]. Similarly, the precursors to echinochromes from the sea urchin *Strongylocentrotus purpuratus* were ascertained to be *via* a PKS derived from the animal using the same strain of *S. cerevisiae* [25]. The successful expression of such large metazoan megasynthase genes necessitated the use of eukaryotic expression systems such as yeast. However, the much smaller terpene cyclases of octocorals have been successfully expressed in strains of *E. coli* which elucidated an animal origin of precursors to several terpenes from this taxonomic group [26,27]. Pairing the prevalence of biosynthetic genes in the kingdom Animalia with the increasing affordability of both next generation sequencing and gene synthesis it is likely that the number of biosynthetic connections between animal NPs and their corresponding genes will continue to grow.

The discovery of NRPS-PKS genes in the genomes and transcriptomes of diverse demosponges is unprecedented in the topics of both NPs biosynthetic and sponge specialized metabolism (Chapter 3). However, characteristics of the NRPS-PKS gene in the *Haliclona indistincta* genome challenges the theory that the 3-alkylpyridine alkaloids are biosynthesized *via* a PKS which accepts NTA as a starter unit. To determine whether this gene is responsible for these alkaloids as well as to explore a novel aspect of sponge specialized metabolism heterologous expression of this gene in yeast was pursued. Furthermore, the presence of introns in the gene, which could theoretically interfere with expression in yeast systems, was also assessed *via* the construction of a cDNA library and attempted amplification of the NRPS-PKS gene.

Materials and Methods

Sample Collection

Haliclona indistincta (MIIG1404) was collected at Newquay on 19/10/2020. Visible epibionts were removed. The sponge was rinsed in sterile artificial seawater, dissected into ~1 cm^3 pieces and flash-frozen with liquid nitrogen. Samples were stored at -70 °C until further use. Voucher specimens were stored in ethanol.

Plasmid Construction and Strain Transformation

The entire 11,922 bp NRPS-PKS gene from the *H. indistincta* genome was synthesized and cloned into the XW55 yeast expression vector by GenScript without codon optimization (Chapter 3) (Figure 4.1) [28]. This plasmid was then used to transform *S. cerevisiae* BJ5464-NpgA using the S.C. EasyComp Transformation Kit (Sigma) [19]. Transformed colonies were selected for on uracil-deficient agar (1.39 g/L Yeast Synthetic Drop-out Media Supplements

without uracil (Sigma-Aldrich), 6.7 g/L Yeast Nitrogen Base (Sigma-Aldrich), 40 mL/L 50% glucose solution, agar 20 g/L) after incubation at 30 °C for 48 hrs.

Bioinformatic Analyses

The weight of the NRPS-PKS enzyme was predicted using the Expasy Compute pI/Mw tool [29]. The presence of transmembrane regions was assessed using DeepTMHMM version 1.0.11 [30].

Heterologous Expression

Attempted expression of the *H. indistincta* NRPS-PKS gene followed a prior protocol used for the expression of an iterative PKS from the mollusc *E. chlorotica* with several modifications [24]. Six separate colonies were used to inoculate separate seed cultures of uracil-deficient broth (5 mL; 1.39 g/L Yeast Synthetic Drop-out Media Supplements without uracil (Sigma-Aldrich), 6.7 g/L Yeast Nitrogen Base (Sigma-Aldrich), 40 mL/L 50% glucose solution) and incubated at 30 °C with shaking at 150 rpm. After 24 hrs, 1 mL of each seed culture was used to inoculate a respective 1 L volume of yeast peptone dextrose broth (10 g/L yeast extract, 20 g/L peptone, 20 g/L glucose). The culture was grown at 30 °C under constant shaking at 180 rpm. After 72 hrs the cells were pelleted at 3,739 g for 20 min at 4 °C. The cell pellets were resuspended in lysis buffer (50 mM NaH_2PO_4 ($2H_2O$ 7.8 g/L; anhydrous 6 g/L), 150 mM NaCl (8.766 g/L), 10 mM imidazole, pH 8.0) and lysed four times by a LM20 Microfluidizer™ Processor set at 20,000 psi. The lysed cells were centrifugated at 28,928 × g for 40 min at 4 °C to separate the supernatant from the cell debris. The supernatant was filtered using a 0.45 µm PVDF syringe filter (Millex-HV, Sigma) before adding Ni-NTA and incubating the mixture at 4 °C for 12 hrs. The solution was then loaded onto a Ni-NTA resin in a gravity column. After the flowthrough passed, the resin was washed with 25 mL each of 20 mM and 50 mM imidazole in 50 mM Tris-HCl buffer (500 mM NaCl, pH 8.0). A final volume of 3 x 5 mL 250 mM imidazole in 50 mM Tris-HCl buffered pH 8.0 was used to elute potential recombinant protein. The presence of the NRPS-PKS protein in the elution was assessed using an SDS-PAGE gel. The entire protocol was also attempted with *S. cerevisiae* strain BJ5464 in case activation of the NRPS-PKS by the NpgA gene resulted in the production of a toxic compound during fermentation. All work involving the heterologous expression of the NRPS-PKS gene was performed in the Marine Natural Products Laboratory of the University of Utah.

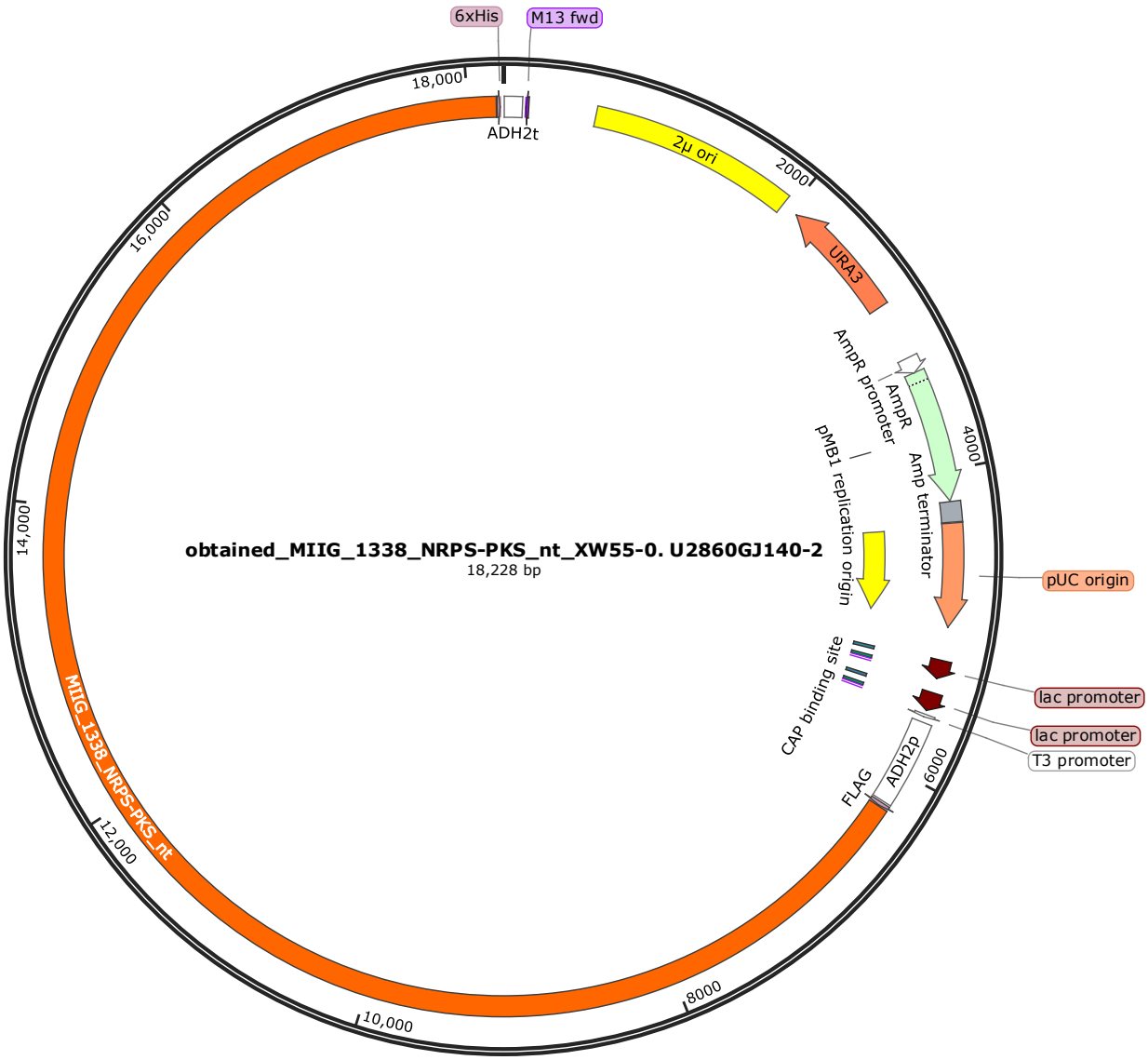


Figure 4.1 MIIG1388 NRPS-PKS gene within the XW55 expression vector.

RNAseq Mapping

The presence of introns in the *H. indistincta* NRPS-PKS gene was assessed by mapping raw cDNA reads upon the genomic contig containing the gene using gsnap version 2020-10-14 and then visualized with IGV version 2.12.3 (Chapter 2, Chapter 3) [31,32].

RNA Extraction and cDNA Synthesis

RNA was extracted from *H. indistincta* specimen MIIG1404 with the intention of generating a cDNA library. From this cDNA library, the presence of introns in the synthesized NRPS-PKS could be assessed *via* PCR amplification and sequencing. Flash-frozen tissue of around 1 cm^3 was first ground in a mortar and pestle with liquid nitrogen. The pulverized tissue was then transferred to 1 mL of Trizol and vortexed until even distribution of the tissue. The sample

was incubated for 5 min at room temperature, vortexed, and centrifuged at 10,000 × g for 5 min at room temperature. The supernatant was transferred to a new tube. 200 µL $CHCl_3$ was added to each supernatant aliquot per 1 mL of Trizol used, vortexed vigorously for 15 sec, incubated at room temperature for 3 min and centrifuged at 12,000 × g for 15 min at 4 °C. The upper aqueous phase was then transferred to a new tube to which 500 µL isopropanol was added to the aqueous phase for every 1 mL of Trizol used. The sample was mixed by inverting the tube several times, incubated at room temperature for 10 min and then centrifuged at 12,000 × g for 10 min at 4 °C to pellet RNA. The RNA pellet was washed with 1 mL of 75% ethanol by inverting the tube 4-5 times and then centrifuged at 12,000 × g for 10 min at 4 °C to remove the supernatant. The RNA pellet was allowed to air dry for 5 min, suspended in 500 µL of 2M LiCl until dissolved and then incubated at room temperature for 5 min. The sample was centrifuged at 12,000 × g at 4 °C for 15 min. The LiCl wash process was repeated two more times. The pellet was then dissolved in 500 µL 1X TE to which an equal volume of phenol/$CHCl_3$/isoamyl alcohol was added, vortexed vigorously, and then centrifuged at 12,000 × g at 4 °C for 15 min. The supernatant was transferred to a new tube. 500 µL of $CHCl_3$ was added to the sample, which was vortexed vigorously, and centrifuged at 12,000 × g at 4 °C for 15 min. The supernatant was removed, and the RNA pellet precipitated in 0.1 volumes of 3M sodium acetate, 2.5 volumes of 100% ethanol, and incubated at -80 °C overnight. The sample was centrifuged at 12,000 × g for 4 °C for 15 min to pellet the RNA. The supernatant was removed and 500 µL of 75% ethanol was added. The sample was mixed by inverting the tube several times, centrifuged at 12,000 × g at 4 °C for 15 min, and the supernatant was removed. The RNA pellet was allowed to air dry for 10 min, dissolved in 30 µL RNase free dH_2O, and the quality and quantity were checked using a Nanodrop spectrophotometer. Purification of the RNA sample followed the manufacturer's protocol of the Zymogen RNA clean and concentrator kit. Finally, cDNA libraries were prepared using the SuperScript™ Double-Stranded cDNA Synthesis Kit using the provided oligo(dT)20 primers and following the manufacturer's protocol.

Gene Amplification

Six sets of PCR primers designed for overlapping regions of the *H. indistincta* NRPS-PKS gene which corresponded to assembled transcripts from the transcriptome, which would lack introns, were used to attempt amplification with the Phusion® High-fidelity DNA polymerase upon a cDNA library produced from isolated mRNA from MIIG1404 (Chapter 2) (Figure 4.2, Appendix 4). The PCR amplification protocol was divided into two parts. The first part consisted of ten cycles in which the denaturation temperature was 94 °C for 10 sec, the annealing temperature (Tm) started at 45 °C for 30 sec and was raised in increments of 1 °C per cycle, and the extension temperature was 68 °C for 1 min. The second part consisted of twenty cycles in which the Tm was instead consistently 53 °C for 1 min for each cycle. Amplification was checked using gel electrophoresis.

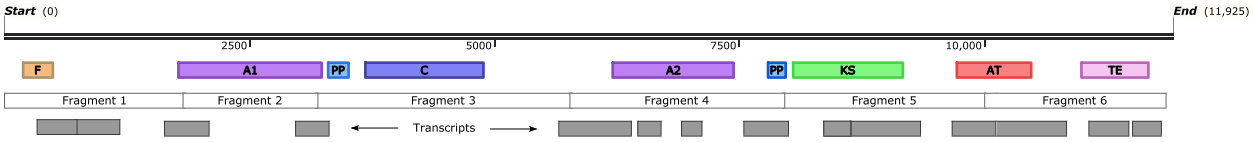


Figure 4.2 The *H. indistincta* NRPS-PKS gene, encoded enzymatic domains, corresponding fragments to be amplified by primers (Appendix 4) and regions covered by transcripts from the assembled transcriptome. Domain naming is as follows: F, formylation; A, adenylation; PP, phosphopantetheine attachment site; KS, ketosynthase; AT, acyltransferase; TE, thioesterase.

Results

The predicted weight of the *H. indistincta* NRPS-PKS is 442.680 kDa. No transmembrane regions were detected. The use of the EasyComp Transformation Kit resulted in a strain of yeast successfully transformed with XW55 containing the *H. indistincta* NRPS-PKS gene as indicated by colony growth on media deficient in uracil. After three days of fermentation and attempted protein extraction, no visible protein band above the 300 kDa marker was visible in the SDS-PAGE gel run after attempted protein purification indicating no isolation of the NRPS-PKS enzyme from the *S. cerevisiae* BJ5464-NpgA cultures (Figure 4.3). Furthermore, no expression was observed with *S. cerevisiae* BJ5464 even when the protein extraction procedure was performed in the timespan of one day to minimize potential protein degradation (Figure 4.4).

Mapping the raw cDNA sequences from the previously assembled transcriptome to the *H. indistincta* NRPS-PKS gene revealed that there were regions of low coverage, as low as a single mapped read, that were not assembled into transcripts (Chapter 2) (Figure 4.2, Figure 4.5). Furthermore, there were regions in which no mapping was detected.

After purification of the RNA extraction with the Zymogen RNA clean and concentrator kit, a sample concentration of 1,655.0 ng/µL, an A260/A280 of 2.20, and an A260/A230 of 2.28 was acquired. Visualizing the sample using gel electrophoresis indicated that RNA molecules greater than 10 kb were isolated (Figure 4.6). No amplification of the NRPS-PKS gene was observed using the designed primers upon the corresponding cDNA library (Figure 4.7).

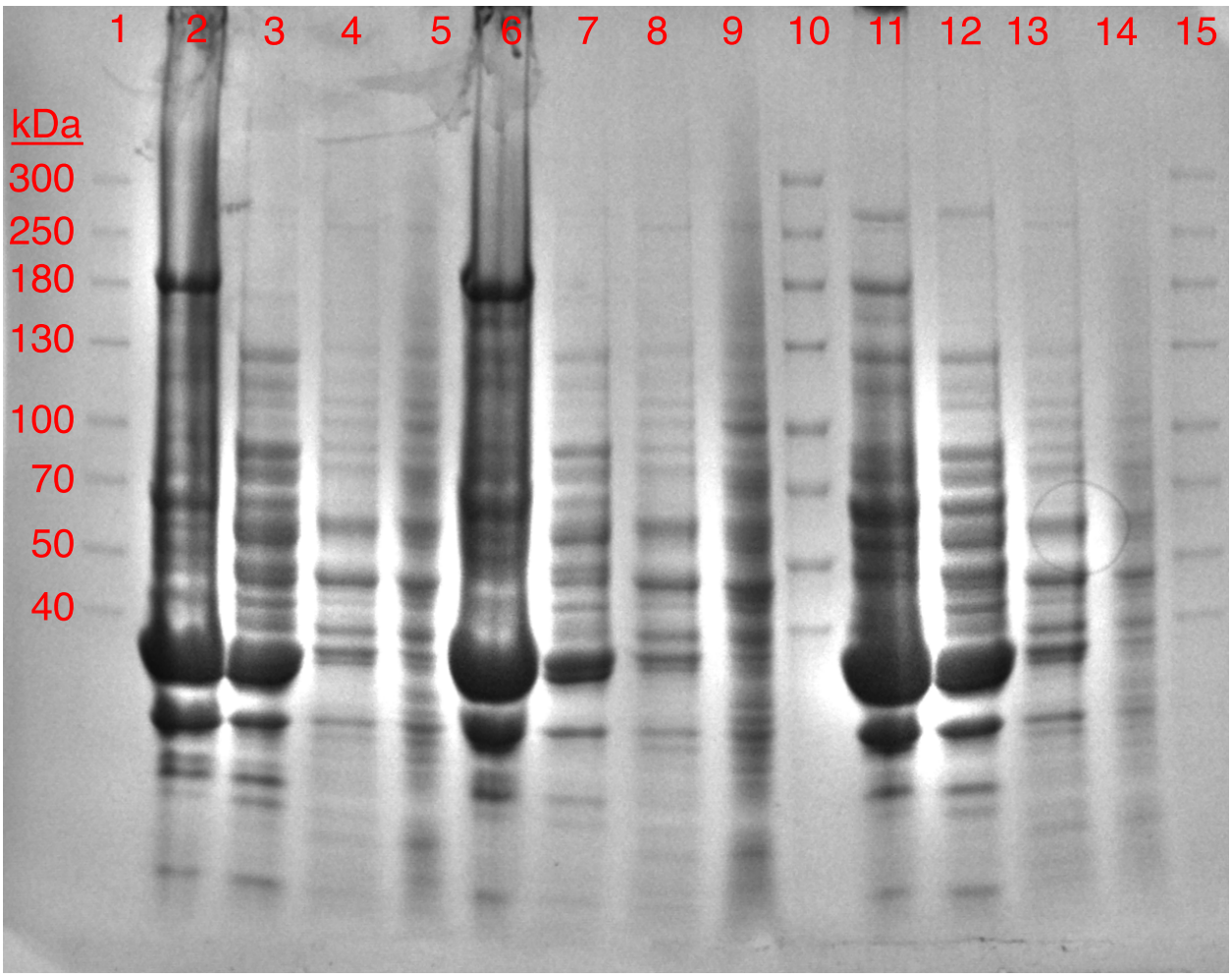


Figure 4.3 SDS-PAGE gel of a protein extraction from *S. cerevisiae* BJ5464-NpgA transformed with XW55 containing *the H. indistincta* NRPS-PKS gene. Columns 1, 10 and 15 represent a 300 kDa ladder. Columns 2, 3, 4 and 5 represent the flowthrough, 20 mM imidazole wash, 50 mM imidazole wash, and 250 mM imidazole elution respectively for one cultured colony. Columns 6-9 and 11-14 represent these same types of samples as 2-5, but for two additional colonies respectively. The circle on columns 13 and 14 is an artifact of the gel imaging process. If successfully expressed, the NRPS-PKS enzyme would be expected to be above the 300 kDa marker.

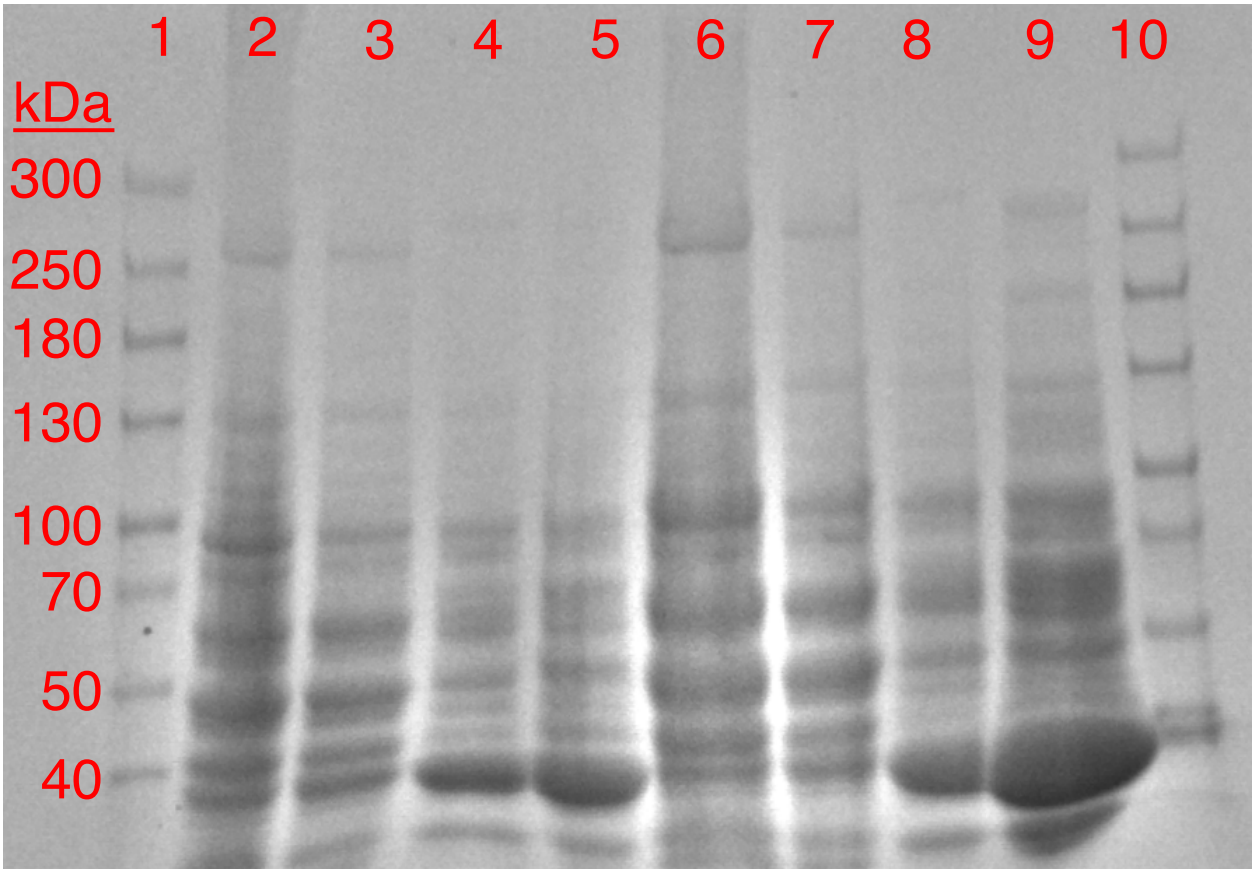


Figure 4.4 SDS-PAGE gel of a protein extraction from *S. cerevisiae* BJ5464 transformed with XW55 containing the *H. indistincta* NRPS-PKS gene. Columns 1 and 10 represent a 300 kDa ladder. Columns 2, 3, 4 and 5 represent the flowthrough, 20 mM imidazole wash, 50 mM imidazole wash, and 250 mM imidazole elution respectively for one cultured colony. Columns 6-9 represent these same types of samples as 2-5, but for a second colony. If successfully expressed, the NRPS-PKS enzyme would be expected to be above the 300 kDa marker.

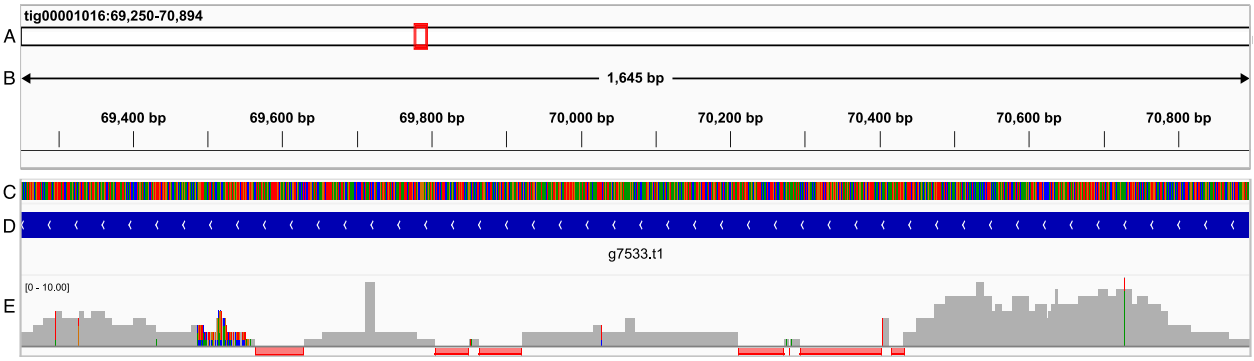


Figure 4.5 Mapping of raw cDNA reads produced and sequenced from extracted mRNA from MIIG1388 to the MIIG1388 contig which contains the NRPS-PKS gene (Chapter 2; Chapter 3) (tig00001016). Row A represents the contig with a section of the NRPS-PKS which is magnified in sequential rows highlighted in red. Row B represents the length of the magnified region of the NRPS-PKS. Row C represents the nucleotide composition of the magnified region. Row D represents the predicted NRPS-PKS ORF. Row E represents the nucleotide coverage from the raw cDNA reads in grey blocks and the absence of coverage with red highlights.

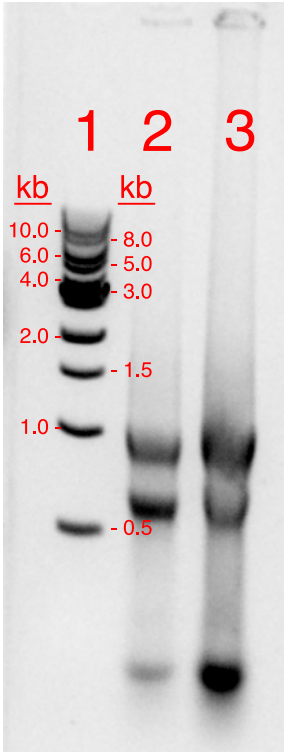


Figure 4.6 Agarose gel electrophoresis of an RNA extraction from *H. indistincta* specimen MIIG1404. Column 1 represents a 1kb DNA ladder. Columns 2 and 3 represent RNA extractions from around 1 cm^3 and 2 cm^3 pieces of tissue respectively.

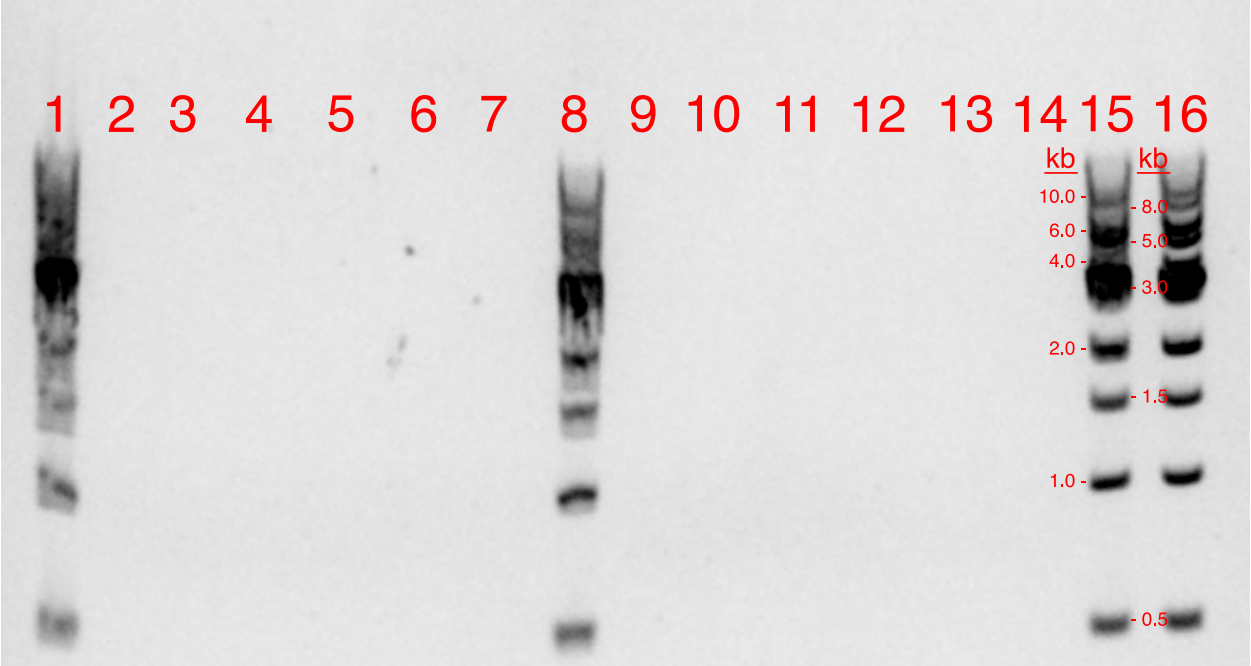


Figure 4.7 Agarose gel electrophoresis of attempted PCR amplification of the *H. indistincta* NRPS-PKS gene from a cDNA library. Columns 1, 8, 15 and 16 represent a 1 kb ladder. Columns 2-7 correspond to attempted amplification of six regions of the *H. indistincta* NRPS-PKS gene from the cDNA library. Columns 9-14 correspond to negative controls represented by attempted amplification of the *H. indistincta* NRPS-PKS gene from extracted RNA. The primers were designed to amplify fragments between 1377 to 2615 bp.

Discussion

Although the *H. indistinta* NRPS-PKS gene was successfully synthesized, cloned into the XW55 plasmid, and then transformed into *S. cerevisiae* strain BJ5464-NpgA, no production of the predicted protein was observed. The phosphopantetheinyl transferase NpgA is necessary for the activation of PKS and NRPS enzymes in the species *A. nidulans* [33]. Based on this, it is often included in *S. cerevisiae* strains, such as BJ5464-NpgA, with the purpose of activating an expressed PKS [19]. If the NRPS-PKS of *H. indistincta* was being expressed in small amounts, it could have been that its activated form was producing a compound toxic to the yeast. In turn, this may have resulted in effects such as the death of cells producing the compound and thus a lack of significant NRPS-PKS protein. With that in mind, expression with BJ5464, which would not produce an activated NRPS-PKS, was also attempted but still resulted in no isolated protein. In addition, eukaryotic transmembrane proteins present several challenges as targets for heterologous expression, but the lack of predicted transmembrane regions in this NRPS-PKS can in theory rule this out as a potential issue [34].

While these observations may address some possible problems with the attempted heterologous expression of the *H. indistincta* NRPS-PKS, there are a myriad of other factors which could have influenced the lack of protein production. For one, the mass is an important factor in the successful expression of a heterologous protein in *S. cerevisiae* and has been shown to influence qualities such as growth rate [35]. Typically, eukaryotic expression systems are necessary to produce proteins greater than 100 kDa [36]. Large megasynthase proteins, such as the FAS-like PKS of 247.973 kDa from *E. chlorotica*, have been successfully expressed in strains of yeast such as BJ5464-NpgA [24]. However, at 442.680 kDa, the NRPS-PKS of *H. indistincta* is nearly twice the size. The native protein size range of *S. cerevisiae* ranges from 25 to more than 4,100 amino acids with one of the largest entries in UniProt, midasin, being an estimated 559.308 kDa [37]. While the *H. Indistincta* NRPS-PKS falls within this range, it is closer to the upper end and thus may have resulted in complications with the successful transcription and translation of the final protein. Alternatively, the lack of codon optimization for *S. cerevisiae* may have caused issues during heterologous expression.

Another possible complication to the successful expression of the *H. indistincta* NRPS-PKS could be the presence of undetected introns. *Saccharomyces cerivisiae* possesses a particularly small number of native introns and their splicing mechanisms differ from that of more closely related organisms such as filamentous fungi [38,39]. In turn, this can introduce problems in the heterologous expression of filamentous fungi BGCs in *S. cerevisiae* [6]. As such, it is not unreasonable to consider that *S. cerevisiae* is incapable of handling the splicing of introns derived from a more distant organism such as a sponge. While the protein-coding gene prediction of BRAKER version 2.1.5 indicated that this gene was contained in a single, 11,925 bp open reading frame it is an *in silico* prediction on a non-model organism and thus its accuracy may be questionable (Chapter 3). In fact, metazoan megasynthases

such as the PKS from *E. chlorotica* comprise a multitude of exons [24]. This quality necessitated the synthesis of the mature gene without introns for its successful heterologous expression in BJ5464-NpgA (Feng Li, personal communication). When analysing the assembled transcripts of the *H. indistincta* NRPS-PKS gene and aligning them to the corresponding genomic sequence it was found that there are regions which would be expected to have function, such as the condensation domain, which are not accounted for by the assembled transcriptome (Figure 4.2). Furthermore, analysis of this transcriptome *via* mapping raw cDNA reads to the genomic NRPS-PKS gene could not clarify whether introns are present in this sequence. Rather, based on the low to non-existent coverage on various portions of the gene all that can instead be concluded is that a greater depth of sequencing is likely necessary to determine the presence of introns.

Inspecting the other sponge NRPS-PKS genes also does not provide a satisfying answer (data not shown). The genome of *Ephydatia muelleri* indicates that two copies of the NRPS-PKS gene exist on a single chromosome [40]. Analysis with AntiSMASH shows that one of these copies is contained in a single ORF, whereas another comprises three exons. Furthermore, using this single ORF sequence of *E. muelleri* to screen the transcriptome of this species does not result in the identification of a single, fully assembled transcript [41]. Similar to that of the *H. indistincta* transcriptome, there are also catalytic domains in the *E. muelleri* NRPS-PKS which are not accounted for in the corresponding transcriptome. In contrast, the transcriptome of the closely related *E. fluviatilis* contains a single transcript which encompass the entirety of the poriferan NRPS-PKS features [42]. All other sponge genomes and transcriptomes which contain similar NRPS-PKS genes are too fragmented to derive a proper conclusion regarding introns [43-46]. Alternatives to short-read sequencing, methods such as the sequencing of full-length cDNA molecules, as is done with Pacific Biosciences Iso-Seq, may offer one potential solution to accurately address this question [47].

While not definitive proof, the lack of amplification of all six primer pairs corresponding to different portions of the *H. indistincta* NRPS-PKS may indicate that an issue common to the entire set of primers occurred. It should be pointed out that several fundamental, overlooked flaws in the design of this experiment exist. The first and foremost is the lack of a positive control to confirm that the designed primers properly function with the *H. indistincta* NRPS-PKS gene. This could have been pursued by attempting to amplify the six respective regions from this gene which were cloned into XW55. Without this information, it is impossible to determine whether the issue is due to the designed primers. Assuming the issue is not attributed to the primer design, one possible setback could be the fact that a different specimen of *H. indistincta*, MIIG1404, was the only one available at the time for RNA extraction. This is in contrast to the specimen MIIG1388 from which the genome and transcriptome was acquired (Chapter 2, Chapter 3). Considering that the NRPS-PKS gene was already lowly expressed in MIIG1388, it may be that a similar or even lower expression was

occurring at the time of flash-freezing MIIG1404. In turn, this potential level of expression could be a reason as to the lack of amplification of any region of the gene. An issue with the PCR protocol should also be mentioned. The New England BioLabs Tm calculator indicates that each individual primer used in this study has a Tm between 52 and 55 °C. As such, the chosen Tm for the amplification protocol consisted of ten cycles from 45-55 °C in which each cycle increased by 1 °C, followed by twenty cycles at a Tm of 53 °C. However, it was overlooked that with the Phusion High-Fidelity DNA polymerase it is recommended to increase the Tm by several degrees. Based on the aforementioned Tm calculator, the recommended Tm should instead have been between 56-58 °C. Such an error could have resulted in nonspecific or no binding of the primers to the DNA template.

Overall, the first attempts at heterologous expression of the *H. indistincta* NRPS-PKS were not successful. It is difficult to ascertain the exact issue with the experimental design, but the presence of introns within the predicted ORF remains an unsolved question. As such, it is likely that repeated attempts at cDNA synthesis and PCR amplification, but this time with adjusted protocols, will be necessary to confirm whether introns are an issue. Alternatively, the sequencing of a deeper transcriptome could also answer this question if sufficient coverage of the NRPS-PKS gene is acquired. In reality, the heterologous expression of functioning metazoan megasynthases is a particularly young topic in the field of biosynthesis. Often, what goes unreported in published works is that a majority of metazoan megasynthase genes, particularly PKSs, do not successfully express in hosts such as *S. cerevisiae* even when potential problems such as introns are addressed (Feng Li, personal communication). It is likely that for the purpose of functional characterization many of these metazoan megasynthases will instead require the heterologous expression of individual enzymatic domains as has been done for a PKS from the duck *Anas platyrhynchos* [48]. The small size of the expressed enzymatic domains allowed for the use of *E. coli* as a heterologous host and the biochemical characterization of each individual enzymatic domain. Based on the complications encountered in this work, such an alternative approach should also be considered. Alternatively, animal heterologous expression systems such as the insect cell baculovirus system may be more suited for this metazoan gene and deserve consideration [49].

References

1. Medema, M.H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J.B., Blin, K., de Bruijn, I., Chooi, Y.H., Claesen, J., Coates, R.C., Cruz-Morales, P., Duddela, S., Düsterhus, S., Edwards, D.J., Fewer, D.P., Garg, N., Geiger, C., Gomez-Escribano, J.P., Greule, A., Hadjithomas, M., Haines, A.S., Helfrich, E.J.N., Hillwig, M.L., Ishida, K., Jones, A.C., Jones, C.S., Jungmann, K., Kegler, C., Kim, H.U., Kötter, P., Krug, D., Masschelein, J., Melnik, A.V., Mantovani, S.M., Monroe, E.A., Moore, M., Moss, N., Nützmann, H.-W., Pan, G., Pati, A., Petras, D., Reen, F.J., Rosconi, F., Rui, Z., Tian, Z., Tobias, N.J., Tsunematsu, Y., Wiemann, P., Wyckoff, E., Yan, X., Yim, G., Yu, F., Xie, Y., Aigle, B., Apel, A.K., Balibar, C.J., Balskus, E.P., Barona-Gómez, F., Bechthold, A., Bode, H.B., Borriss, R., Brady, S.F., Brakhage, A.A., Caffrey, P., Cheng, Y.-Q., Clardy, J., Cox, R.J., De Mot, R., Donadio, S., Donia, M.S., van der Donk, W.A., Dorrestein, P.C., Doyle, S., Driessen, A.J.M., Ehling-Schulz, M., Entian, K.-D., Fischbach, M.A., Gerwick, L., Gerwick, W.H., Gross, H., Gust, B., Hertweck, C., Höfte, M., Jensen, S.E., Ju, J., Katz, L., Kaysser, L., Klassen, J.L., Keller, N.P., Kormanec, J., Kuipers, O.P., Kuzuyama, T., Kyrpides, N.C., Kwon, H.-J., Lautru, S., Lavigne, R., Lee, C.Y., Linquan, B., Liu, X., Liu, W., Luzhetskyy, A., Mahmud, T., Mast, Y., Méndez, C., Metsä-Ketelä, M., Micklefield, J., Mitchell, D.A., Moore, B.S., Moreira, L.M., Müller, R., Neilan, B.A., Nett, M., Nielsen, J., O'Gara, F., Oikawa, H., Osbourn, A., Osburne, M.S., Ostash, B., Payne, S.M., Pernodet, J.-L., Petricek, M., Piel, J., Ploux, O., Raaijmakers, J.M., Salas, J.A., Schmitt, E.K., Scott, B., Seipke, R.F., Shen, B., Sherman, D.H., Sivonen, K., Smanski, M.J., Sosio, M., Stegmann, E., Süssmuth, R.D., Tahlan, K., Thomas, C.M., Tang, Y., Truman, A.W., Viaud, M., Walton, J.D., Walsh, C.T., Weber, T., van Wezel, G.P., Wilkinson, B., Willey, J.M., Wohlleben, W., Wright, G.D., Ziemert, N., Zhang, C., Zotchev, S.B., Breitling, R., Takano, E., Glöckner, F.O., 2015. Minimum Information about a Biosynthetic Gene cluster. Nature Chemical Biology 11, 625–631. https://doi.org/10.1038/nchembio.1890
2. Goodwin, S., McPherson, J.D., McCombie, W.R., 2016. Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics 17, 333–351. https://doi.org/10.1038/nrg.2016.49
3. Blin, K., Shaw, S., Kloosterman, A.M., Charlop-Powers, Z., van Wezel, G.P., Medema, M.H., Weber, T., 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Research 49, W29–W35. https://doi.org/10.1093/nar/gkab335
4. Robinson, S.L., Terlouw, B.R., Smith, M.D., Pidot, S.J., Stinear, T.P., Medema, M.H., Wackett, L.P., 2020. Global analysis of adenylate-forming enzymes reveals β-lactone biosynthesis pathway in pathogenic *Nocardia*. Journal of Biological Chemistry 295, 14826–14839. https://doi.org/10.1074/jbc.RA120.013528
5. Walsh, G., 2018. Biopharmaceutical benchmarks 2018. Nature Biotechnology 36, 1136–1145. https://doi.org/10.1038/nbt.4305
6. Jia Zhang, J., Tang, X., S. Moore, B., 2019. Genetic platforms for heterologous expression of microbial natural products. Natural Product Reports 36, 1313–1332. https://doi.org/10.1039/C9NP00025A
7. Kosuri, S., Church, G.M., 2014. Large-scale *de novo* DNA synthesis: technologies and applications. Nature Methods 11, 499–507. https://doi.org/10.1038/nmeth.2918
8. Flinspach, K., Westrich, L., Kaysser, L., Siebenberg, S., Gomez-Escribano, J.P., Bibb, M., Gust, B., Heide, L., 2010. Heterologous expression of the biosynthetic gene clusters of coumermycin A1, clorobiocin and caprazamycins in genetically modified *Streptomyces coelicolor* strains. Biopolymers 93, 823–832. https://doi.org/10.1002/bip.21493
9. Mantovani, S.M., Moore, B.S., 2013. Flavin-linked oxidase catalyzes pyrrolizine formation of dichloropyrrole-containing polyketide extender unit in chlorizidine A. The Journal of the American Chemical Society 135, 18032–18035. https://doi.org/10.1021/ja409520v
10. Tu, Q., Herrmann, J., Hu, S., Raju, R., Bian, X., Zhang, Y., Müller, R., 2016. Genetic engineering and heterologous expression of the disorazol biosynthetic gene cluster via Red/ET recombineering. Scientific Reports 6, 21066. https://doi.org/10.1038/srep21066

11. Jones, A.C., Gust, B., Kulik, A., Heide, L., Buttner, M.J., Bibb, M.J., 2013. Phage P1-derived artificial chromosomes facilitate heterologous expression of the FK506 gene cluster. PLOS ONE 8, e69319. https://doi.org/10.1371/journal.pone.0069319
12. Bok, J.W., Ye, R., Clevenger, K.D., Mead, D., Wagner, M., Krerowicz, A., Albright, J.C., Goering, A.W., Thomas, P.M., Kelleher, N.L., Keller, N.P., Wu, C.C., 2015. Fungal artificial chromosomes for mining of the fungal secondary metabolome. BioMed Central Genomics 16, 343. https://doi.org/10.1186/s12864-015-1561-x
13. Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., Smith, H.O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods 6, 343–345. https://doi.org/10.1038/nmeth.1318
14. Vior, N.M., Lacret, R., Chandra, G., Dorai-Raj, S., Trick, M., Truman, A.W., 2018. Discovery and biosynthesis of the antibiotic bicyclomycin in distantly related bacterial classes. Applied and Environmental Microbiology 84, e02828-17. https://doi.org/10.1128/AEM.02828-17
15. Zhang, J.J., Tang, X., Zhang, M., Nguyen, D., Moore, B.S., 2017. Broad-host-range expression reveals native and host regulatory elements that influence heterologous antibiotic production in Gram-negative bacteria. mBio 8, e01291-17. https://doi.org/10.1128/mBio.01291-17
16. Alberti, F., J. Leng, D., Wilkening, I., Song, L., Tosin, M., Corre, C., 2019. Triggering the expression of a silent gene cluster from genetically intractable bacteria results in scleric acid discovery. Chemical Science 10, 453–463. https://doi.org/10.1039/C8SC03814G
17. Gomez-Escribano, J.P., Bibb, M.J., 2011. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. Microbial Biotechnology 4, 207–215. https://doi.org/10.1111/j.1751-7915.2010.00219.x
18. Pfeifer, B.A., Admiraal, S.J., Gramajo, H., Cane, D.E., Khosla, C., 2001. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. Science 291, 1790–1792. https://doi.org/10.1126/science.1058092
19. Ma, S.M., Li, J.W.-H., Choi, J.W., Zhou, H., Lee, K.K.M., Moorthie, V.A., Xie, X., Kealey, J.T., Da Silva, N.A., Vederas, J.C., Tang, Y., 2009. Complete reconstitution of a highly-reducing iterative polyketide synthase. Science 326, 589–592. https://doi.org/10.1126/science.1175602
20. Chiang, Y.-M., Oakley, C.E., Ahuja, M., Entwistle, R., Schultz, A., Chang, S.-L., Sung, C.T., Wang, C.C.C., Oakley, B.R., 2013. An efficient system for heterologous expression of secondary metabolite genes in *Aspergillus nidulans*. The Journal of the American Chemical Society 135, 7720–7731. https://doi.org/10.1021/ja401945a
21. Torres, J.P., Schmidt, E.W., 2019. The biosynthetic diversity of the animal world. Journal of Biological Chemistry 294, 17684–17692. https://doi.org/10.1074/jbc.REV119.006130
22. Feng, L., Gordon, M.T., Liu, Y., Basso, K.B., Butcher, R.A., 2021. Mapping the biosynthetic pathway of a hybrid polyketide-nonribosomal peptide in a metazoan. Nature Communications 12, 4912. https://doi.org/10.1038/s41467-021-24682-9
23. O'Brien, R.V., Davis, R.W., Khosla, C., Hillenmeyer, M.E., 2014. Computational identification and analysis of orphan assembly-line polyketide synthases. The Journal of Antibiotics 67, 89–97. https://doi.org/10.1038/ja.2013.125
24. Torres, J.P., Lin, Z., Winter, J.M., Krug, P.J., Schmidt, E.W., 2020. Animal biosynthesis of complex polyketides in a photosynthetic partnership. Nature Communications 11, 2882. https://doi.org/10.1038/s41467-020-16376-5
25. Li, F., Lin, Z., Torres, J.P., Hill, E.A., Li, D., Townsend, C.A., Schmidt, E.W., 2022. Sea urchin polyketide synthase SpPks1 produces the naphthalene precursor to echinoderm pigments. The Journal of the American Chemical Society 144, 9363–9371. https://doi.org/10.1021/jacs.2c01416
26. Burkhardt, I., de Rond, T., Chen, P.Y.-T., Moore, B.S., 2022. Ancient plant-like terpene biosynthesis in corals. Nature Chemical Biology 18, 664–669. https://doi.org/10.1038/s41589-022-01026-2

27. Scesa, P.D., Lin, Z., Schmidt, E.W., 2022. Ancient defensive terpene biosynthetic gene clusters in the soft corals. Nature Chemical Biology 18, 659–663. https://doi.org/10.1038/s41589-022-01027-1
28. Cacho, R.A., Tang, Y., 2016. Reconstitution of fungal nonribosomal peptide synthetases in yeast and *in vitro*, in: Evans, B.S. (Ed.), Nonribosomal peptide and polyketide biosynthesis: methods and protocols, Methods in Molecular Biology. Springer, New York, NY, pp. 103–119. https://doi.org/10.1007/978-1-4939-3375-4_7
29. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy Server, in: Walker, J.M. (Ed.), The Proteomics Protocols Handbook, Springer Protocols Handbooks. Humana Press, Totowa, NJ, pp. 571–607. https://doi.org/10.1385/1-59259-890-0:571
30. Hallgren, J., Tsirigos, K.D., Pedersen, M.D., Armenteros, J.J.A., Marcatili, P., Nielsen, H., Krogh, A., Winther, O., 2022. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. https://doi.org/10.1101/2022.04.08.487609
31. Wu, T.D., Reeder, J., Lawrence, M., Becker, G., Brauer, M.J., 2016. GMAP and GSNAP for genomic sequence alignment: enhancements to speed, accuracy, and functionality, in: Mathé, E., Davis, S. (Eds.), Statistical genomics: methods and protocols, Methods in Molecular Biology. Springer, New York, NY, pp. 283–334. https://doi.org/10.1007/978-1-4939-3578-9_15
32. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nature Biotechnology 29, 24–26. https://doi.org/10.1038/nbt.1754
33. Márquez-Fernández, O., Trigos, Á., Ramos-Balderas, J.L., Viniegra-González, G., Deising, H.B., Aguirre, J., 2007. Phosphopantetheinyl transferase CfwA/NpgA is required for *Aspergillus nidulans* secondary metabolism and asexual development. Eukaryotic Cell 6, 710–720. https://doi.org/10.1128/EC.00362-06
34. Clark, K.M., Fedoriw, N., Robinson, K., Connelly, S.M., Randles, J., Malkowski, M.G., DeTitta, G.T., Dumont, M.E., 2010. Purification of transmembrane proteins from *Saccharomyces cerevisiae* for X-ray crystallography. Protein Expression and Purification 71, 207–223. https://doi.org/10.1016/j.pep.2009.12.012
35. de Ruijter, J.C., Koskela, E.V., Nandania, J., Frey, A.D., Velagapudi, V., 2018. Understanding the metabolic burden of recombinant antibody production in *Saccharomyces cerevisiae* using a quantitative metabolomics approach. Yeast 35, 331–341. https://doi.org/10.1002/yea.3298
36. Demain, A.L., Vaishnav, P., 2009. Production of recombinant proteins by microbes and higher organisms. Biotechnology Advances 27, 297–306. https://doi.org/10.1016/j.biotechadv.2009.01.008
37. Warringer, J., Blomberg, A., 2006. Evolutionary constraints on yeast protein size. BioMed Central Evolutionary Biology 6, 61. https://doi.org/10.1186/1471-2148-6-61
38. Spingola, M., Grate, L., Haussler, D., Ares, M., 1999. Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. RNA 5, 221–234. https://doi.org/10.1017/S1355838299981682
39. Kupfer, D.M., Drabenstot, S.D., Buchanan, K.L., Lai, H., Zhu, H., Dyer, D.W., Roe, B.A., Murphy, J.W., 2004. Introns and splicing elements of five diverse fungi. Eukaryotic Cell 3, 1088–1100. https://doi.org/10.1128/EC.3.5.1088-1100.2004
40. Kenny, N.J., Francis, W.R., Rivera-Vicéns, R.E., Juravel, K., de Mendoza, A., Díez-Vives, C., Lister, R., Bezares-Calderón, L.A., Grombacher, L., Roller, M., Barlow, L.D., Camilli, S., Ryan, J.F., Wörheide, G., Hill, A.L., Riesgo, A., Leys, S.P., 2020. Tracing animal genomic evolution with the chromosomal-level assembly of the freshwater sponge *Ephydatia muelleri*. Nature Communications 11, 3676. https://doi.org/10.1038/s41467-020-17397-w
41. Leys, S., 2017. *Ephydatia muelleri* Trinity transcriptome [WWW Document]. Education & Research Archive. https://doi.org/10.7939/R3WH2DV20
42. Alié, A., Hayashi, T., Sugimura, I., Manuel, M., Sugano, W., Mano, A., Satoh, N., Agata, K., Funayama, N., 2015. The ancestral gene repertoire of animal stem cells. The Proceedings of the National Academy of Sciences U S A 112, E7093–E7100. https://doi.org/10.1073/pnas.1514789112

43. Desplat, Y., Warner, J.F., Lopez, J.V., 2022. Holo-transcriptome sequences from the tropical marine sponge *Cinachyrella alloclada*. Journal of Heredity 113, 184–187. https://doi.org/10.1093/jhered/esab075
44. Plese, B., Kenny, N.J., Rossi, M.E., Cárdenas, P., Schuster, A., Taboada, S., Koutsouveli, V., Riesgo, A., 2021. Mitochondrial evolution in the Demospongiae (Porifera): phylogeny, divergence time, and genome biology. Molecular Phylogenetics and Evolution 155, 107011. https://doi.org/10.1016/j.ympev.2020.107011
45. Kenny, N.J., Plese, B., Riesgo, A., Itskovich, V.B., 2019. Symbiosis, selection, and novelty: freshwater adaptation in the unique sponges of Lake Baikal. Molecular Biology and Evolution 36, 2462–2480. https://doi.org/10.1093/molbev/msz151
46. Ryu, T., Seridi, L., Moitinho-Silva, L., Oates, M., Liew, Y.J., Mavromatis, C., Wang, X., Haywood, A., Lafi, F.F., Kupresanin, M., Sougrat, R., Alzahrani, M.A., Giles, E., Ghosheh, Y., Schunter, C., Baumgarten, S., Berumen, M.L., Gao, X., Aranda, M., Foret, S., Gough, J., Voolstra, C.R., Hentschel, U., Ravasi, T., 2016. Hologenome analysis of two marine sponges with different microbiomes. BioMed Central Genomics 17, 158. https://doi.org/10.1186/s12864-016-2501-0
47. Gonzalez-Garay, M.L., 2016. Introduction to isoform sequencing using Pacific Biosciences technology (Iso-Seq), in: Wu, J. (Ed.), Transcriptomics and gene regulation, translational bioinformatics. Springer Netherlands, Dordrecht, pp. 141–160. https://doi.org/10.1007/978-94-017-7450-5_6
48. Sabatini, M., Comba, S., Altabe, S., Recio-Balsells, A.I., Labadie, G.R., Takano, E., Gramajo, H., Arabolaza, A., 2018. Biochemical characterization of the minimal domains of an iterative eukaryotic polyketide synthase. The Federation of European Biochemical Societies Journal 285, 4494–4511. https://doi.org/10.1111/febs.14675
49. Jarvis, D.L., 2009. Chapter 14 Baculovirus–insect cell expression systems, in: Burgess, R.R., Deutscher, M.P. (Eds.), Methods in enzymology, guide to protein purification, 2nd Edition. Academic Press, pp. 191–222. https://doi.org/10.1016/S0076-6879(09)63014-7

Chapter 5

Actinoporin-like Proteins from the Phylum Porifera

Part of this chapter contributed to the peer-reviewed publication:
Sandoval, K., McCormack, G.P., 2022. Actinoporin-like Proteins Are Widely Distributed in the Phylum Porifera. Marine Drugs 20, 74.

Introduction

Actinoporins (APs) are proteinaceous α-pore-forming toxins originally isolated from and named after sea anemones [1]. This group of toxins typically exhibits several common characteristics, such as a common absence of cysteine residues, a high isoelectric point (>8.8), and a small size (~20 kDa) [2]. Furthermore, they comprise a compact β-sandwich flanked on each side by an α-helix, as indicated by the crystal structures of the well-studied equinatoxin II (EqT-II), stichyolysin II (Stn-II), and fragaceatoxin C (Fra-C) [3-5]. The molecular mechanism of cytolytic pore formation by APs has been extensively researched and appears to involve several steps, which are briefly summarized. First, lipid recognition and membrane binding are accomplished *via* the interfacial binding site (IBS), which features a cluster of prominent aromatic residues that bind to phosphocholine (POC) [4,6]. In particular, APs have an affinity for the POC group of sphingomyelin (SM) and are capable of discriminating between this target and other membrane lipids, such as phosphatidyl-choline [6,7]. After binding to a target membrane, APs then undergo a conformational change in which the N-terminal region, containing one of the α-helices, is translocated to lie flat upon the membrane surface [8,9]. This N-terminal region is then inserted into the target membrane and undergoes further conformational change to increase the overall length of the amphipathic α-helix relative to its unbound state [10,11]. The pore is finally formed when oligomerization occurs *via* the recruitment of additional AP monomers, which undergo the same process in the same region of the membrane to bring about the death of targeted cells by osmotic shock [12,13]. For a more in-depth explanation on the molecular mechanisms of pore formation by APs, the reader is referred to reviews which focus on this topic [13-15]. The qualities of APs which allow for their membrane-binding and pore-forming activity have attracted attention regarding potential biotechnological and therapeutic applications, such as the design of immunotoxins, nanopores, adjuvants, and SM-specific probes [16-19].

Historically, sea anemones have been the primary source of APs, although similar cytolytic proteins can be found in other anthozoans [20,21]. Indeed, an exhaustive bioinformatic analysis indicated that actinoporin-like proteins (ALPs) are distributed across multiple phyla with high structural similarity despite low sequence similarity [22]. In particular, APs and ALPs have been detected in chordates (primarily teleost fish), cnidarians, molluscs, mosses, and ferns. Furthermore, a structural similarity of APs and ALPs to fungal-fruit body lectins has also been determined. A phylogenetic analysis of these identified proteins revealed four distinct groups comprising ALPs primarily found in vertebrates, hydrozoan ALPs, APs from cnidarians and plants, and fungal fruit-body lectins, all of which were proposed to comprise the actinoporin-like proteins and fungal fruit-body lectins superfamily (AF). The presence of ALP genes in non-vertebrate bilaterians has been further illuminated in studies focused on polychaetes of the genus *Glycera*, the crustacean *Xibalbanus tulumensis*, the brachiopod *Lingula anatina*, and many molluscs of the classes Gastropoda and Bivalvia

[23-25]. Several ALPs have been functionally characterized, indicating that they can possess similar qualities to APs regarding membrane binding and cytolytic activity.

The first published example of an ALP was echotoxin II, isolated from the salivary glands of the predatory mollusc *Monoplex echo,* which also has an amphipathic N-terminal α-helix, a patch of aromatic residues, and haemolytic activity, but a specificity to gangliosides rather than SM [26]. In further contrast to APs, an ALP from the zebrafish, *D. rerio*, possessed no cytolytic activity and its membrane-binding activity was not specific to SM [22]. Yet bryoporin, from the moss *Physcomitrella patens*, showed consistency with its close phylogenetic grouping to APs in that it also exhibited specificity for SM as well as haemolytic activity, although its biological role appears to be related to dehydration stress [27]. Similarly, clamlysin B from the bivalve *Corbicula japonica* also exhibits SM-binding and cytolytic activity [28]. Finally, the ALP HALT-1 from the cnidarian *Hydra magnipapillata*, which is phylogenetically distinct from anthozoan APs, exhibits lower haemolytic activity, the creation of larger pores, and a lower affinity to SM in comparison to EqT-II [29]. Altogether, the observation that ALPs from non-anthozoans can possess similar biochemical properties to APs supports the notion that they may also be potential targets for the aforementioned biotechnological and therapeutic applications which have been investigated for EqT-II, Stn-II, and Fra-C.

Sponges of the phylum Porifera are benthic, filter-feeding animals which can be found in marine and freshwater environments throughout the world. Given the niche they fill, challenges faced by these organisms include contact with pathogenic microorganisms, spatial competition with other benthic life, and predation [30-32]. In order to deal with these challenges, many sponges utilize complex chemical armaments which display an array of bioactivities towards targets such as pathogens, fouling organisms and tumoral cells [33-35]. While a majority of these bioactivities have been attributed to small molecules, some larger proteinaceous toxins have also been identified, such as suberitine from *Suberites domuncula*, halilectin-3 from *Haliclona caerulea*, and chondrosin from *Chondrosia reniformis* [36-38]. In addition, sponges have also been shown to be a source of cytolytic pore-forming proteins, one of which is an antibacterial, perforin-like protein from *S. domuncula*, which was found to be upregulated upon exposure to lipopolysaccharide [39,40]. No ALPs have been isolated and characterized from this phylum, but a recent phylogenetic study indicated that genes encoding for these proteins are present in the genome of the species *Oscarella pearsei* (then *Oscarella carmela*, when its genome was sequenced) [21,41,42]. Little was reported on this ALP other than it being phylogenetically distant from both anthozoan APs, hydrozoan ALPs and mollusc ALPs.

While analysing the transcriptome assemblies of native Irish *Haliclona* species, the presence of numerous genes encoding for proteins with the Pfam domain PF06369, representing sea anemone cytotoxic proteins, was noticed (Chapter 2). The presence of ALPs in sponges of both the classes Demospongiae and Homoscleromorpha prompted the questions of whether

these proteins are widely distributed throughout the phylum and how similar they are to known APs. As discussed previously, such a quality would expand the possible biotechnological and therapeutic applications of sponges. To date, these organisms have been the subject of numerous transcriptomic and genomic studies, resulting in a wealth of public data with which to carry out such an inquiry [43]. Based on this information, several objectives were pursued. First, to identify ALPs in all available sponge transcriptomes and genomes. Second, to assess the structural similarity of sponge ALPs to APs of cnidarians. Third, to predict the functional capability of one ALP with specific attention towards the core processes of membrane binding, oligomerization, and pore formation. Fourth, to estimate whether the identified sponge ALPs can instead be classified as APs, similar to those from plants, based on phylogenetic analysis [22].

Materials and Methods

Transcriptome Assemblies

Previously detailed transcriptome assemblies and annotations of *Haliclona cinerea*, *Haliclona indistincta*, *Haliclona oculata*, *Haliclona simulans* and *Haliclona viscosa* were used for analyses in this chapter (Chapter 2). Sequential analyses were performed with an account at the Leibniz Supercomputing Centre.

Identification of Novel Actinoporin-like Proteins from Sponges

While analysing the output of the homology search against the Pfam database (Chapter 2), it was noticed that numerous translated protein sequences from the *Haliclona* transcriptomes possessed the PF06369 domain representing sea anemone cytotoxic proteins such as actinoporins. Due to the biotechnological potential of actinoporins, this prompted the screening of all publicly available sponge genomes [41,44-49] and transcriptomes [42,45, 50-68] to see whether other members of the phylum also encoded ALPs in their genome. If transcriptome assemblies were not provided in the original publication, the data were assembled in the same manner as the Irish *Haliclona* as previously detailed (Chapter 2). The *Haliclona* ALPs were used as queries in a tblastn search against the other sponge assemblies with an E-value cut-off of 1e-4. The longest open reading frames were then extracted from the hits using TransDecoder version 5.5.0 [69]. These protein sequences were screened against the Pfam database (accessed on 14 January 2020) with HMMER version 3.2.1 [70,71] and only those indicated as possessing domain PF06369 were retained for further analysis. These identified sponge ALPs were also screened against the NCBI Conserved Domain Database to confirm that PF06369 was the primary conserved domain [72]. To explore the possible evolutionary origin of ALPs in animals, the genomes of the choanoflagellates *Monosiga brevicollis* and *Salpingoeca rosetta*, the ctenophores *Mnemiopsis leidyi* and *Pleurobrachia bachei*, and the placozoans *Trichoplax adhaerens* and *Hoilungia hongkongensis* were also screened using EqT-II as a query [73-78]. Furthermore, nineteen choanoflagellate transcriptomes were also screened in a similar manner [79].

Sequence Analysis and Structural Prediction

All identified sponge ALPs were used as a query against the NCBI non-redundant protein database to identify the closest homologous sequence [80]. SignalP 5.0 was used to identify the presence of signal peptides in sponge ALPs contained within a complete ORF [81]. The isoelectric point and molecular weight of complete, mature sponge ALPs were determined using the compute pI/Mw tool of ExPASy [82]. Protein structure prediction of all sponge ALPs was performed using Phyre2 Suite version 5.1 [83]. The quality of the predicted protein structure for Hi2 was assessed with ProQ3D and ModFOLD8 [84,85]. Structural alignment of sponge ALPs upon the crystal structure of chain A from EqT-II (1iaz) was performed with TM-Align [86]. Protein structures were visualized using UCSF Chimera version 1.15 [87]. A protein topology plot was created using Pro-origami [88]. A multiple sequence alignment of Hi2, EqT-II (P61914), Fra-C (B9W5G6), and Stn-II (P07845) was performed with MAFFT v7.490, with the L-INS-i alignment method using default settings [89]. The multiple sequence alignment was then visualized using Jalview version 2.11.1.4 [90]. Analysis of the N-terminal α-helix of Hi2 and generation of an Edmundson wheel projection were accomplished using HeliQuest and NetWheels [91,92].

Phylogenetic Analysis of Actinoporin-like Proteins from Sponges

Sponge ALPs derived from complete ORFs were chosen for multiple sequence alignment and phylogenetic analysis. All alignments were performed using MAFFT 7.490, with the L-INS-i alignment method using default settings [89]. APs were represented by well-characterized actiniarian proteins, such as the aforementioned EqT-II, Fra-C, and Stn-II, as well as those from stony and soft corals. Cnidarian ALPs were represented by the series of HALT proteins from *H. magnipapillata*. In general, the sponge ALPs most consistently aligned with sequences from sea anemones, fungi of the class Glomeromycota, and teleost fish, in that order. To get more sequences for the phylogenetic tree, all complete sponge ALPs were queried against the NCBI nr database against these three taxonomic groups, as well as against molluscs and invertebrates which did not fall under these aforementioned phyla. The top hit from each category for each sponge sequence was then retrieved. This resulted in seven groups of sequences: glomeromycete fungi, sponges, anthozoan cnidarians, hydrozoan cnidarians, molluscs, miscellaneous invertebrates, and teleost fish. All signal peptides were removed prior to alignments using SignalP version 5.0 [81]. Each of these groups was separately aligned to the mature sequences of EqT-II, Fra-C, and Stn-II. The individual alignments were then trimmed corresponding to the boundaries of EqT-II, Fra-C, and Stn-II using Jalview version 2.11.1.4 [90]. Several sponge ALPs, while complete, were excluded due to being excessively truncated compared to EqT-II or introducing significant gaps. All trimmed sequences were then pooled together and once more aligned. Maximum likelihood trees were constructed in IQ-TREE version 2.1.4 with 1000 bootstrap pseudoreplicates with the intention of visualizing the degree of similarity the sponge ALPs had to APs and ALPs

from other phyla [93]. The resulting phylogenetic trees were modified using the Interactive Tree of Life (iTOL) v6 [94].

Data Availability

Transcriptome assemblies, sponge ALP sequences, predicted protein structures, multiple sequence alignments, and maximum likelihood trees are available at Mendeley [96].

Results

A total of 68 unique open reading frames encoding for proteins with the Pfam domain PF06369 or exhibiting significant alignments towards APs or ALPs was identified [96,97]. Out of the 63 poriferan species screened, 20 were found to be the source of ALP genes (Appendix 5) [96,97]. Of these, 38 ORFs were determined to be complete by TransDecoder and derived from 10 species: *Cliona orientalis* (Co), *Cliona varians* (Cv), *Dysidea avara* (Da), *Geodia barretti* (Gb), *H. indistincta* (Hi), *H. viscosa* (Hv), *O. pearsei* (Op), a *Scolapina* sp. (S), *Spongia officinalis* (So), and *Tethya wilhelma* (Tw). As none of these sponge-derived proteins have been biochemically characterized, they are hence referred to as ALPs instead of APs. ALPs were absent from all freshwater sponge data. Their presence in an order did not necessarily translate to this being an absolute feature of said order. This is particularly exemplified by the order Haplosclerida, in which numerous paralogs were detected in the sister species *H. indistincta* and *H. viscosa*, a few in *H. amboinensis* and *H. cinerea*, and none in *H. oculata*, *H. simulans*, *Haliclona tubifera*, or *Amphimedon queenslandica*. Only *C. varians* and *D. avara* possessed the same degree of parology as *H. indistincta* and *H. viscosa*. Of the sponge ALPs, only two from *T. wilhelma* were identified as having a signal peptide. All the complete sponge ALPs aligned most closely with cnidarian actinoporins when these were used in a blastp query against the NCBI non-redundant database [97]. Specifically, one ALP from *H. indistincta* named Hi2 exhibited the highest sequence similarity with an actinoporin from *Haloclava producta* at 59.06% identity. It was also observed that several ALPs from incomplete ORFs most closely aligned with pore- forming proteins from other phyla, such as coluporins and tereporins from the Mollusca. No ALPs were detected in the screened genomes or transcriptomes of choanoflagellates, ctenophores, or placozoans. The theoretical isoelectric point of complete sponge ALPs ranged from 4.66 to 9.46, whereas the average molecular weight ranged from 14,153.27 to 33,419.71 Da [97].

In contrast to the low-to-modest sequence similarity many sponge ALPs showed in relation to actiniarian APs (27.70–59.06%), homology modelling with Phyre2 indicated that the aligned predicted structure of all complete sponge ALPs was highly similar to that of Stn-II and EqT-II, with confidence values ranging from 97.3% to 100%. Quantification of the similarity between the predicted sponge ALP models and the crystal structure of EqT-II *via* TM-align showed that all produced models had a TM-align value above 0.5, indicating that the structural similarity was not random (Figure 5.1), whereas the RMSD values ranged from 0.54 to 1.74 (Å) [96].

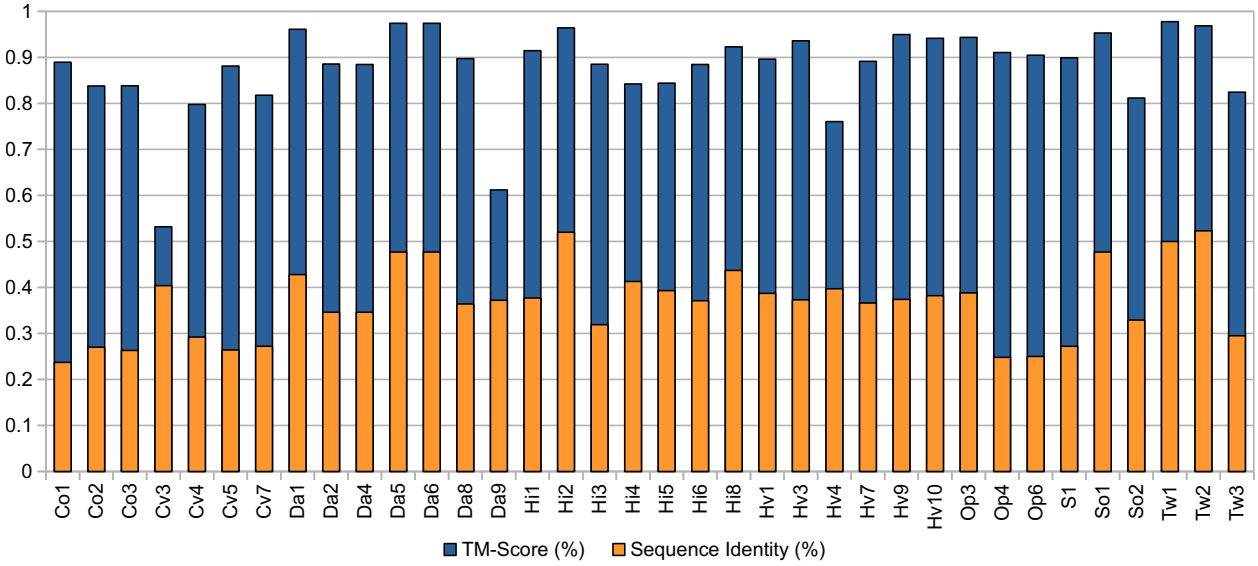


Figure 5.1 Comparison of complete sponge ALPs with EqT-II. Blue represents the TM-score of all predicted sponge ALP structures *via* Phyre2 with the crystal structure of EqT-II (1IAZ). Orange represents the sequence identity of the aligned region of the two structures. The orange bar representing percent identity is overlaid upon the blue bar representing TM-score. Abbreviations are as follows: Co, *C. orientalis*; Cv, *C. varians*; Da, *Dysidea avara*; Hi, *H. indistincta*; Hv, *H. viscosa*; Op, *O. pearsei*; S, *Scopalina* sp.; So, *S. officinalis*; Tw, *T. wilhelma*. The protein Sc1 from *S. carteri* is excluded due to having unknown residues. Sequence identity is expressed on a scale of 0–1 rather than as a percentage.

Several sponge ALPs exhibited structural inconsistencies with the expected AP skeleton, such as the lack of an N-terminal α-helix. Hi2 exhibited one of the highest TM-scores at 0.96406. The quality of the predicted model for Hi2 was further supported with a ProQ3D S-score of 0.697 (0.5–1.0 representing a good model; Arne Elofsson, personal communication) and a ModFOLD8 p-value of 3.772 × 10 − 6 (less than a 1/1000 chance that the model is incorrect). As can be seen by the structure generated by Phyre2, Hi2 shares many characteristics typical of cnidarian actinoporins, such as comprising a β-sandwich flanked by two α-helices (Figure 5.2A,D). Furthermore, the localization of the interfacial binding site can also be observed (Figure 5.2B). In general, the predicted structure of Hi2 appears to overlay well with the crystal structure of EqT-II chain A (1IAZ) (Figure 5.2C).

Due to its high sequence similarity with cnidarian actinoporins and its confirmed presence in the *H. indistincta* genome (Chapter 3), Hi2 was chosen for further in-depth analyses to determine whether it exhibited the same membrane- binding and pore-forming activities. A multiple sequence alignment of Hi2 with the final mature peptides of the well-studied equinatoxin II (EqT-II; P61914), fragaceatoxin C (Fra- C; B9W5G6), and stichyolysin II (Stn-II; P07845) indicated a percent identity of 50.56%, 50.00%, and 51.41%, respectively (Figure 5.3). Despite these modest values, the alignment illustrated a high degree of conservation regarding residues and motifs critical for the functional activities of actinoporins [15]. For example, a majority of the residues associated with the interfacial

binding site in Hi2 are consistent with those of EqT-II, Fra-C, and Stn-II [6]. In addition, Hi2 possesses the conserved residue Tyr112, which is critical for SM recognition; however, a substitution of Leu for Trp at residue 111 is also observed. Furthermore, the presence of Ser53, Val86, Ser104, Pro106, Trp115, Tyre132, Tyr136, and Tyr137 are consistent with the POC binding site found in cnidarian APs [4,6,15], and the conserved P-[WYF]-D binding motif found in this region of APs is also present in Hi2 at residues 106–108 [22]. Oligomerization of actinoporin monomers upon the cell membrane is another crucial step towards pore formation and is known to be influenced by an Arg–Gly–Asp motif. Hi2 shows inconsistency with this motif, as it instead possesses Lys142, Gly143, and Glu144. However, Hi2 possesses the residue Lys76, which is consistent with similar residues associated with oligomerization in other APs [95]. The presence of Ile59 and Trp147 are also partially consistent with residues of Fra-C, associated with oligomerization and protein–protein interaction between protomers; the observed substitution of Ile for Val at this site can be seen in Stn-II [13,98]. Unlike most cnidarian APs, Hi2 exhibits the presence of cysteine at residue 141, but such a characteristic has previously been reported in these proteins [21].

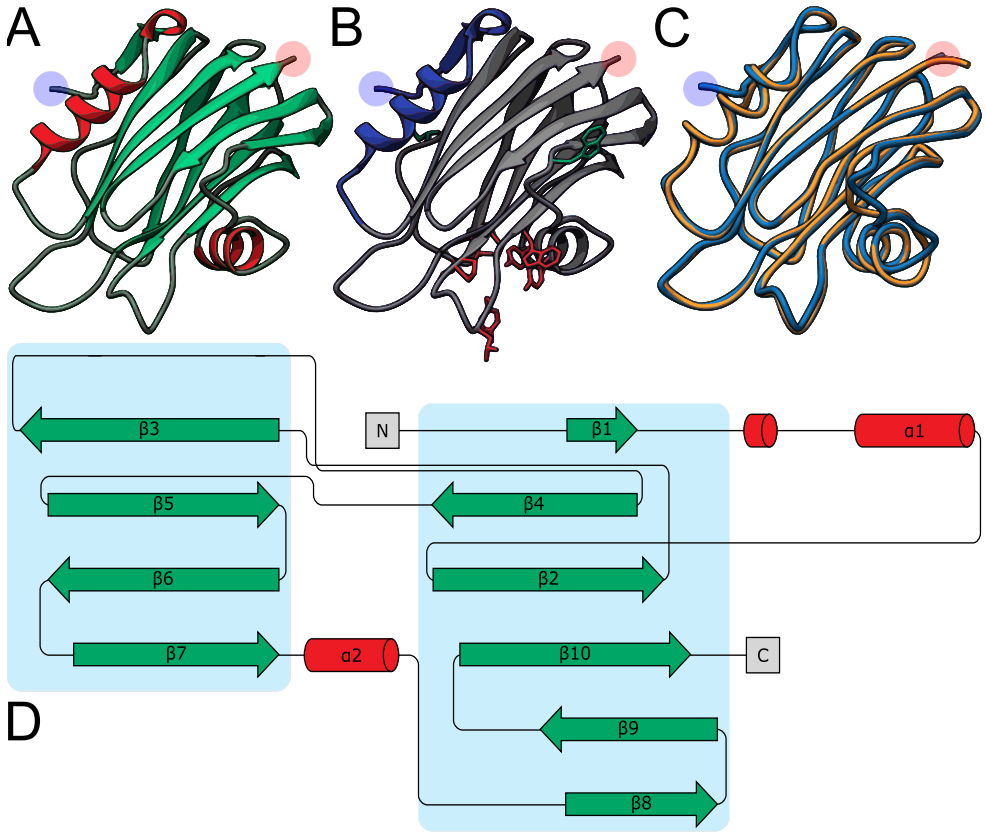


Figure 5.2 (**A**) Predicted structure of Hi2 by Phyre2. Red represents α-helices. Green represents β-sheets. (**B**) Significant functional residues of Hi2. Blue represents the N-terminal α-helix associated with pore formation. Red represents the residues of the interfacial binding site. Green represents the residues associated with oligomerization. (**C**) Structural alignment of Hi2 in blue upon EqT-II Chain A (1IAZ) in orange. For all predicted structures blue and red highlights represent the N- and C-terminus, respectively. (**D**) Protein topology plot of Hi2 with the same colour scheme as (**A**).

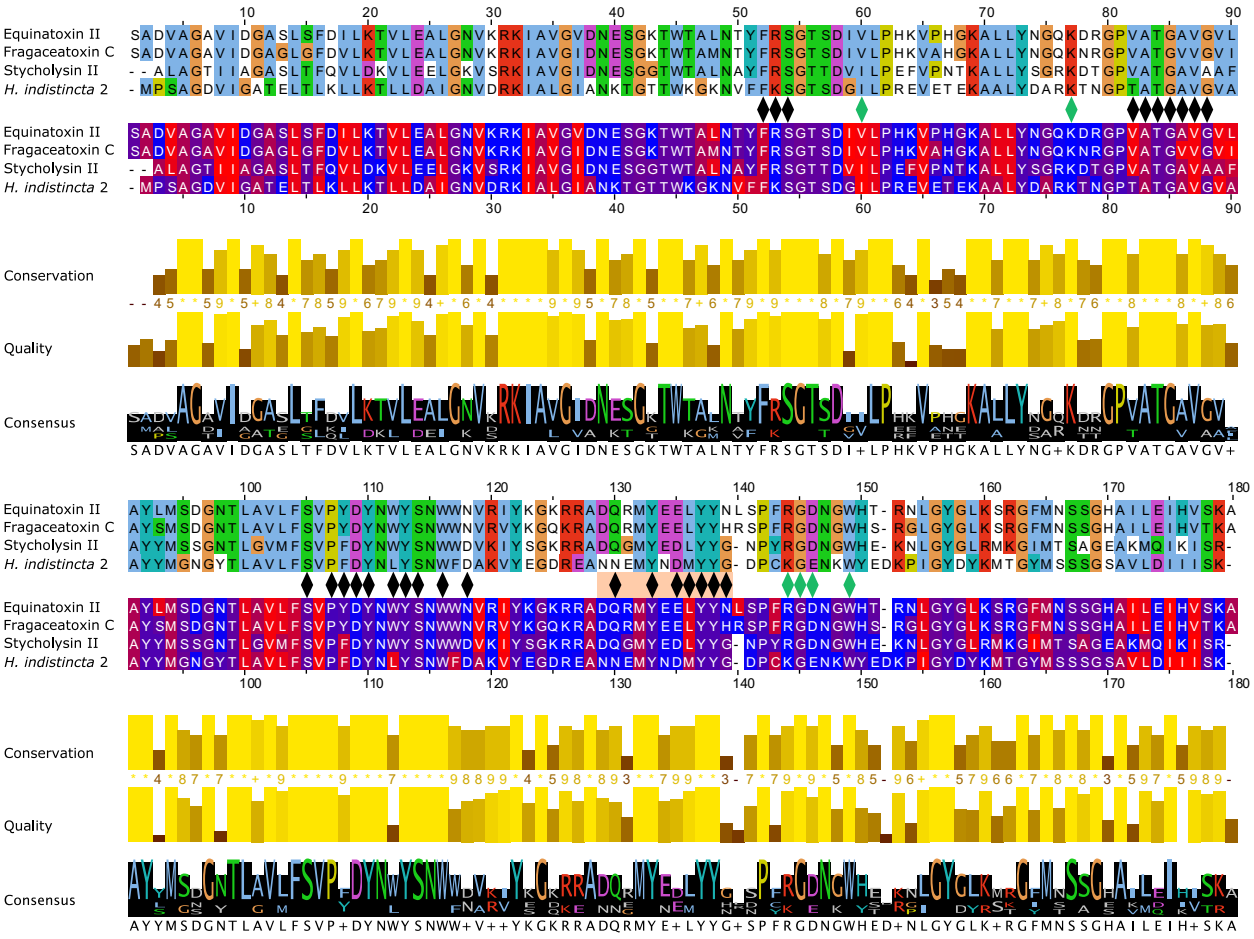


Figure 5.3 Multiple sequence alignment of Hi2 with EqT-II, Fra-C, and Stn-II. The top alignment represents Clustalx colour coding. The bottom alignment represents hydrophobicity colour coding. Between these two alignments, black markers represent residues important for membrane binding, green markers represent residues important for oligomerization, and orange rectangles represent the α-helices of EqT-II (the final length of the N-terminal α-helix after a conformational change and insertion into the target membrane is presented [10,12,13,15,95, 98].

Membrane penetration and pore formation by oligomerized APs is achieved by their respective amphipathic N-terminal α-helices. In EqT-II, the N-terminal region undergoes a conformational change, producing an α-helix comprising residues 6-28 which is capable of spanning a target membrane [10]. These residues correspond to a conserved N-terminal glycine and C-terminal asparagine of the α-helix, which are also present in Fra-C and Hi2. Within this region, Hi2 also showed consistency with several previously determined highly conserved hydrophobic residues (Val7, Ile8, Leu13, Leu18, Leu22, and Ile25), as well as Arg30, which is associated with the insertion of the α-helix into the target membrane [20]. Using these sequence boundaries allowed for the construction of an Edmundson peptide helical wheel of the predicted Hi2 N-terminal α-helix after a hypothetical conformational change (Figure 5.4A). Consistent with the amphipathic nature of the N-terminal α-helix of EqT-II, Fra-C, and Stn-II, a side comprising a majority of polar amino acids opposite another

comprising a majority of nonpolar amino acids can be seen in that of Hi2. Furthermore, the hydrophobic moment of Hi2, a measure of helix amphipathicity, was calculated to be 0.384 μH, which is comparable to that of EqT-II at 0.337 μH. The N-terminal μH of Hi2 exhibits hydrophobicity of 0.607 and a net charge of 0. The two faces of the N-terminal α-helix prior to a hypothetical conformational change also display a hydrophobic and hydrophilic side, which are, respectively, oriented towards and away from the rest of the protein (Figure 5.4B).

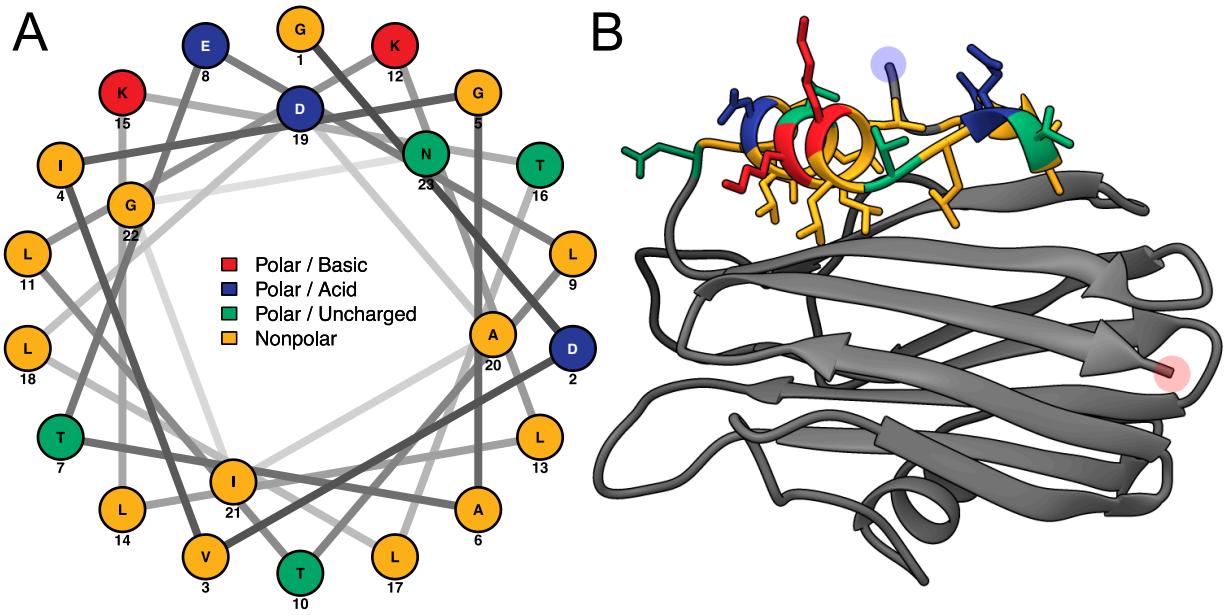


Figure 5.4 (**A**) Edmundson peptide helical wheel projection of residues 5-28 from Hi2. (**B**) Residues 5-28 of the Phyre2 predicted structure of Hi2 coloured in the same manner as (**A**). Blue and red highlights represent the N- and C-terminus, respectively.

The relatedness between AP/ALP sequences from the two *Haliclon*a species in which they were present, AP sequences from cnidarians, and ALP sequences from other taxa was visualized with an initial maximum likelihood tree (Figure 5.5). Here we find that the sequence Hi2 from *H. indistincta* nestles well within the AP clade from cnidarians, but no sequence of this type was found in its close sister species, *H. viscosa*. Additional (ALP) sequences from *H. indistincta* and its sister species *H. viscosa* form a distinct and highly supported clade well outside that of the cnidarians, indicating radiation from an additional AP/ALP copy in the ancestor of that species group. This clade is distinct from the other animal ALPs, which also form a monophyletic grouping supported by 91 BP.

When all available AP/ALP sequences from Porifera are added to the analysis, the high number of paralogs present in the dataset obscure the phylogenetic signal and reduce confidence in generating an accurate phylogeny, However, the reconstructed maximum likelihood tree provides insight into the sequence similarity and potential relatedness of the sponge ALPs, as well as similar proteins from other phyla. Four distinct groups with strong

bootstrap support were produced: (1) ALPs from fungi of the class Glomeromycota (used as outgroup), (2) the majority of the sequences from the genus *Haliclona*, (3) anthozoan APs with other sponge ALPs, and (4) ALPs from other invertebrates and chordates (Figure 5.6). The strong grouping of the *Haliclona* ALPs, independent of all other sequences and placed as the sister group to all other APs and ALPs, was a consistently observed phenomenon while testing other alignments for the reconstruction of a final tree (data not shown). The remaining APs and ALPs form a monophyletic grouping within which there are two clades; one, supported by 100 BP, consisting of the freshwater Hydra and the bilaterians, the other containing the marine cnidarians and sponges (75 BP). The presence of a large number of other sponge sequences has the effect of pulling the Hi2 sequence outside the cnidarian clade, which itself is no longer supported by bootstrapping. Relationships between the other sponge sequences are unclear, as indicated by the low bootstrap support of internal nodes. This is particularly exemplified by the likely spurious placement of Op3 within the clade of anthozoan APs. Frequently, it was observed that additions or subtractions of sequences in the alignment would result in this protein—along with Hi2, Sc1, So1, Tw1, and Tw2—being shifted in and out of the cnidarian AP group. The main exception to this is a second strongly supported clade consisting of ALPs from the genera *Cliona*, *Geodia*, *Scopalina*, and *Tethya*. While this clade may move relative to other sponge sequences, the clade remained intact and distinct from the anthozoan APs. Two groups of *D. avara* sequences were present, both highly supported, but not always remaining together on trees, depending on the comparative sequences included. They were also always distinct from anthozoan APs. Despite having a high sequence similarity to both anthozoan APs and sponge ALPs, those derived from plants, such as bryoporin, were only found to introduce additional noise into the data without significantly changing tree topology and were thus excluded as they are hypothesized to be the result of a horizontal gene transfer event (data not shown) [27].

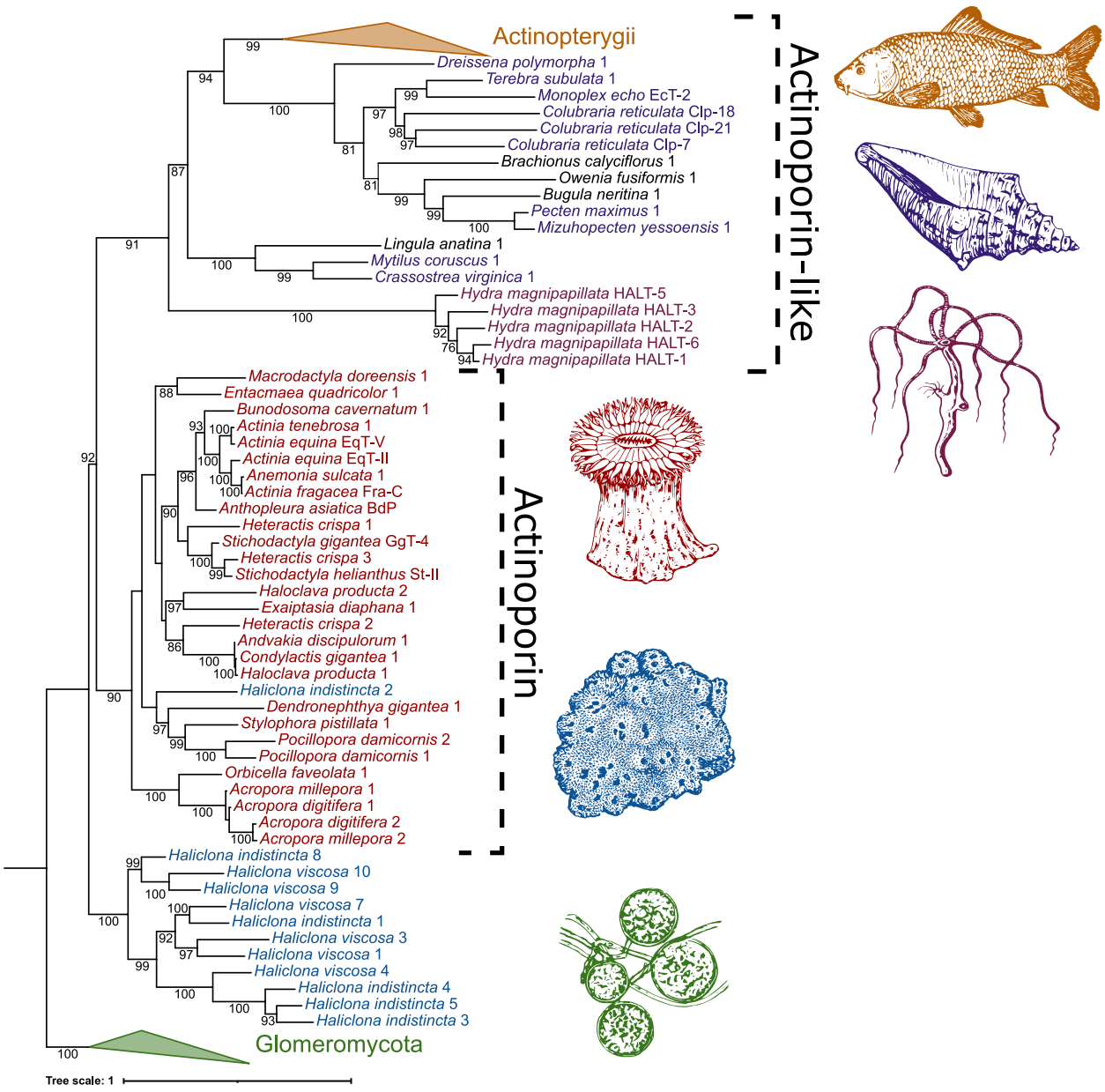


Figure 5.5 Phylogenetic tree of ALPs from the genus *Haliclona* and similar proteins. Numbers on nodes represent maximum likelihood bootstrap values above 75%. The colour scheme is as follows: green represents ALPs from fungi of the class Glomeromycota; blue represents ALPs from sponges of the genus *Haliclona*; red represents APs from cnidarians of the class Anthozoa; reddish-purple represents ALPs from cnidarians of the class Hydrozoa; purple represents ALPs from the phylum Mollusca; orange represents ALPs from fish of the class Actinopterygii; black represents ALPs from miscellaneous phyla. Abbreviations are as follows: EqT, equinatoxin; Fra, fragaceatoxin; St, sticholysin; BdP, bandaporin; GgT, gigantoxin; HALT, hydra actinoporin-like toxin; EcT, echotoxin; Clp, coluporin. The glomeromycete vector image is a reproduced copy of the high-resolution image 4341-2_p1 of *Rhizophagus irregularis* (Błaszk., Wubet, Renker, and Buscot) C. Walker and A. Schüßler derived from AACF/CCAMF (Accessed 26 November 2021). The sponge, anemone, seashell, and carp vector images were derived from those of Pearson Scott Foresman under a CC0 1.0 license (Accessed on 26 November 2021). The *Hydra vulgaris* vector image was derived from the Freshwater and Marine Image Bank at the University of Washington under a CC0 1.0 license (accessed on 26 November 2021).

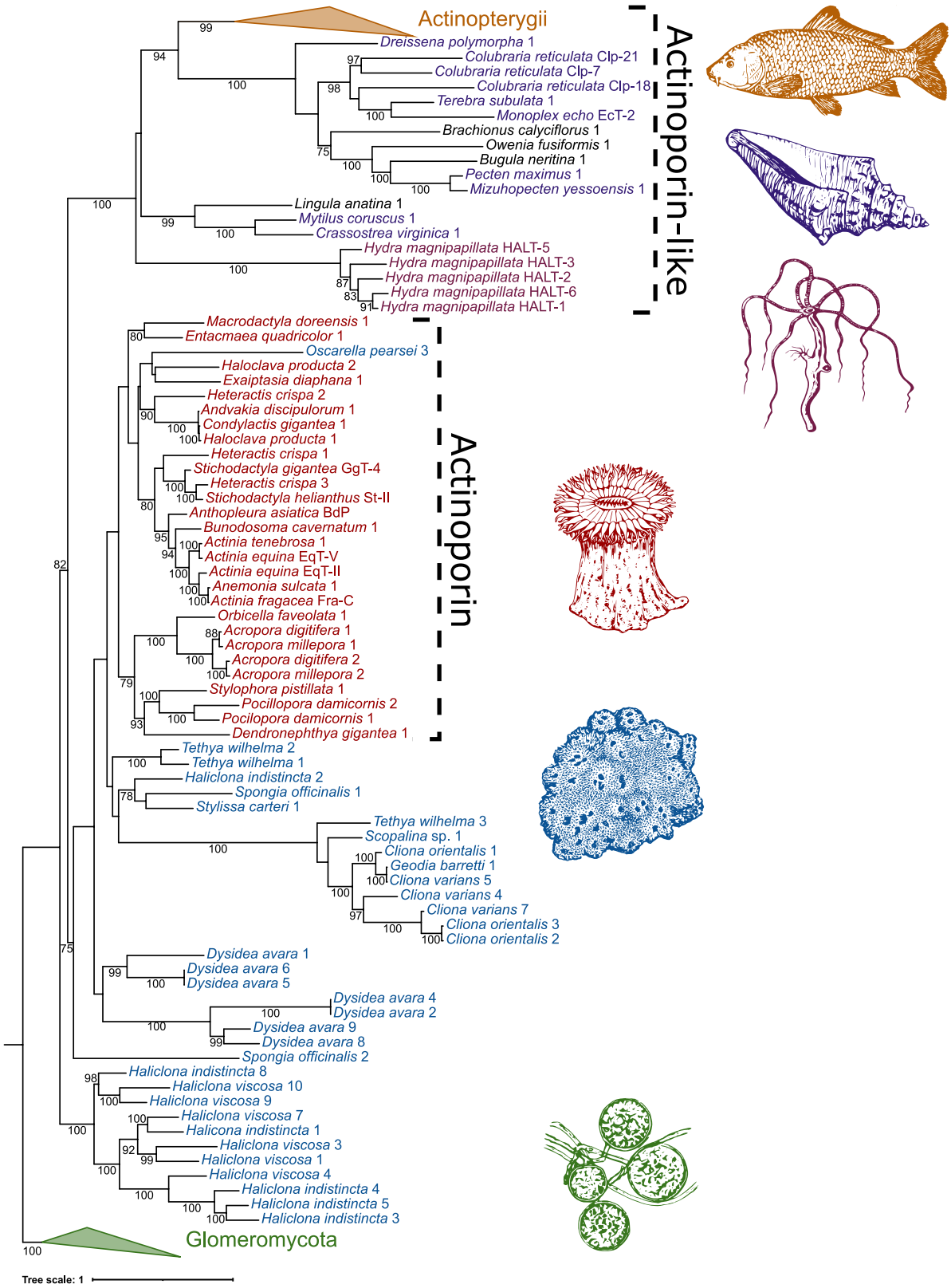


Figure 5.6 Phylogenetic tree of ALPs from sponges and similar proteins. Numbers on nodes represent maximum likelihood bootstrap values above 75%. Blue represents sponge AP/ALPs. All other colour schemes as well as image acknowledgements are identical to those in the prior ALP phylogenetic tree (Figure 5.5).

Discussion

Cytolytic pore-forming toxins are widely distributed throughout prokaryotic and eukaryotic life and often function as immunological defences in the latter [99,100]. The proposed AF superfamily includes α-pore-forming toxins derived from diverse eukaryotic lineages with similar predicted protein structure despite low sequence similarity [22]. Members of this family include anthozoan APs, plant APs, hydrozoan ALPs, ALPs from other animals, and fungal fruit-body lectins. Herein, new additions from the phylum Porifera are proposed to belong to the AF superfamily, as members of the classes Demospongiae, Homoscleromorpha, and Calcarea have been shown to possess genes encoding for ALPs. Many functionally characterized AF proteins appear to serve a role in envenomation as they have been localized in the nematocysts of cnidarians and the salivary glands of predatory molluscs [25,26,29,101]. However, AF proteins appear to also have functions other than envenomation, as indicated by their presence in the mesenteric filaments of cnidarians, as well as organisms which do not perform this process, such as bivalves [28,101]. With this in consideration, it is not entirely unusual to observe the presence of these proteins in non-venomous animals, such as sponges, in which they may serve a different ecological function *via* a similar molecular mechanism.

When considering the taxonomic distribution of ALPs in sponges, no clear pattern can be discerned. The majority of the identified ALPs were derived from the Demospongiae and more specifically from the orders Axinellida, Bubarida, Clionaida, Dendroceratida, Dictyoceratida, Haplosclerida, Poecilosclerida, Suberitida, Tethyida, and Tetractinellida, but not Chondrillida, Spongillida, or Verongiida. However, further analysis at the genus level indicated that these proteins appear to have been lost in species closely related to those which possess ALPs. This is particularly exemplified by the genera *Haliclona* and *Geodia*, in which the transcriptomes of numerous species have been reported [43,50]. Furthermore, ALPs appear to be distributed throughout the phylum Porifera, as several were identified in the classes Homoscleromorpha and Calcarea, but not in Hexactinellida. However, two caveats should be considered regarding these observations. The first is that the high prevalence of these proteins in the Demospongiae is most likely due to sampling bias, as transcriptomes of this class have been disproportionately sequenced compared to the other three. Second, as most of the data analysed in this study are derived from transcriptomes, a lack of gene expression or sequencing depth cannot be ruled out as an explanation for these genes not having been identified in a species. That said, the absence of ALPs in the *A. queenslandica* genome does appear to indicate that the loss of ALP genes has occurred in the order Haplosclerida, which may be an explanation for their absence in other species [49].

The predicted structure of most identified ALPs from sponges exhibited a high degree of aligned structural similarity to anthozoan APs, despite a low to moderate sequence similarity. Such an observation is common in studies of AF proteins from other organisms

[22,25]. That said, the observed high aligned structural similarity does not necessarily equate to these sponge ALPs having the same membrane-binding and pore-forming capabilities. This is particularly exemplified by the observation that the N-terminal region of sponge ALPs appears to vary greatly. For example, this region is fully present in Hi2, truncated in Hi3, and completely absent in Hi4. Such a quality is not unique to sponge ALPs, as can be seen by analysing those of hydrozoans [29]. Similarly, an incomplete N-terminal α-helix was also observed on the ALP Dr1 from *D. rerio* and was hypothesized to influence the lack of pore-forming capabilities and specificity towards SM exhibited by the protein [22]. With this in mind, many of the identified ALPs from sponges may serve functions other than those associated with well-characterized APs, such as EqT-II. In contrast, the higher similarity of Hi2 to anthozoan APs at both a sequence and structural level could indicate that it is capable of SM recognition and pore formation, which prompted a further analysis on this concept.

A multiple sequence alignment of Hi2 with the model APs EqT-II, Fra-C, and Stn-II indicated that this predicted protein shares numerous conserved residues associated with the fundamental processes of lipid recognition and membrane binding, insertion of the N-terminal α-helix into the target membrane, and oligomerization allowing for pore formation. The presence of a patch of aromatic residues (Tyr114, Trp117, Tyr134, Tyr138, and Tyr139) in Hi2 is consistent with the IBS observed in the model APs EqT-II, Fra-C, and Stn-II. The major observed deviation is that a substitution of leucine for tryptophan occurs at residue 113 of Hi2. The importance of the equivalent residue in EqT-II, Trp112, in membrane binding and SM recognition has been exemplified by studies in which this residue was mutated to phenylalanine or subject to ^{19}F NMR studies [102,103]. This substitution also appears to be prevalent in nature, as it exists in numerous anemone APs, as well as the ALP Dr1 from *D. rerio* [20,22]. Furthermore, a mutant of EqT-II containing this substitution exhibited similar SM specificity to the wild-type protein [6]. For these reasons, this substitution in Hi2 is not expected to significantly inhibit any possible membrane-binding and SM-recognition capabilities of the protein. While at a lower level of conservation the N-terminal region of Hi2 also appears to form an α-helix of similar length to EqT-II, Fra-C, and Stn-II. The notion that this α-helix is capable of inserting itself into a target membrane is supported by the amphipathic nature of the predicted helical wheel and its hydrophobic moment comparable to previously analysed AP N-terminal α-helices [20]. Finally, residues associated with oligomerization were somewhat consistent between Hi2 and the model actinoporins. While Hi2 showed conservation at residues Lys76, Ile59, and Trp147, whose equivalents in anthozoan APs are associated with oligomerization, its RGD motif was another site of substitution [12,13,45,46]. Hi2 instead shows a substitution of Lys for Arg and Glu for Asp. That said, Lys and Glu are of a similar charge and hydrophobicity to Arg and Asp, respectively, and may allow for the retention of the oligomerization function [12]. Furthermore, this Lys substitution has been observed in natural actinoporins [20]. Hi2 and

many other sponge ALPs also exhibited an acidic pI (Supplementary Table), which is uncharacteristic of the typically basic anthozoan APs [15,104]. This observation is not entirely unusual, however, as there is a precedent for acidic APs derived from anthozoans [20, 104,105].

The identification of ALPs from glomeromycete fungi with high sequence similarity to known APs supports the previous notion of a pre-metazoan origin of these proteins [22]. The presence of these genes in fungi, sponges, and cnidarians, and their absence in choanoflagellates, ctenophores, and placozoans, could possibly be explained by a series of gene losses occurring throughout their history. However, additional assemblies from these organisms should be assessed prior to making this conclusion. Furthermore, the observation that many species of sponge were the source of numerous ALP isoforms may be an indication that duplication of this gene is a common event in this phylum, similar to the situation in cnidarians and molluscs [25,104]. This observed diversification paired with the fact that many sponge ALPs were derived from transcriptomes indicates that these are not simply genomic relics but do play some sort of functional role in sponges. In the phylogenetic analysis, the sponge ALPs were not found to cluster together in a clear monophyletic group, which appears to further exemplify the notion of a high degree of divergence of these proteins in the Porifera. However, based on the strong signal that numerous sponge ALPs share with anthozoan APs, to the point of being grouped together, and considering the higher sequence similarity many sponge ALPs have with cnidarian APs, several of these proteins from sponges may instead be classifiable as APs.

References

1. Kem, W.R., 1988. structure and action. The Biology of Nematocysts 375–405.
2. Anderluh, G., Maček, P., 2002. Cytolytic peptide and protein toxins from sea anemones (Anthozoa: Actiniaria). Toxicon 40, 111–124. https://doi.org/10.1016/S0041-0101(01)00191-X
3. Athanasiadis, A., Anderluh, G., Maček, P., Turk, D., 2001. Crystal structure of the soluble form of equinatoxin II, a pore-forming toxin from the sea anemone *Actinia equina*. Structure 9, 341–346. https://doi.org/10.1016/S0969-2126(01)00592-5
4. Mancheño, J.M., Martín-Benito, J., Martínez-Ripoll, M., Gavilanes, J.G., Hermoso, J.A., 2003. Crystal and electron microscopy structures of sticholysin II actinoporin reveal insights into the mechanism of membrane pore formation. Structure 11, 1319–1328. https://doi.org/10.1016/j.str.2003.09.019
5. Mechaly, A.E., Bellomio, A., Gil-Cartón, D., Morante, K., Valle, M., González-Mañas, J.M., Guérin, D.M.A., 2011. Structural insights into the oligomerization and architecture of eukaryotic membrane pore-forming toxins. Structure 19, 181–191. https://doi.org/10.1016/j.str.2010.11.013
6. Bakrač, B., Gutiérrez-Aguirre, I., Podlesek, Z., Sonnen, A.F.-P., Gilbert, R.J.C., Maček, P., Lakey, J.H., Anderluh, G., 2008. Molecular determinants of sphingomyelin specificity of a eukaryotic pore-forming toxin. Journal of Biological Chemistry 283, 18665–18677. https://doi.org/10.1074/jbc.M708747200
7. Bakrač, B., Kladnik, A., Maček, P., McHaffie, G., Werner, A., Lakey, J.H., Anderluh, G., 2010. A Toxin-based probe reveals cytoplasmic exposure of golgi sphingomyelin. Journal of Biological Chemistry 285, 22186–22195. https://doi.org/10.1074/jbc.M110.105122
8. Malovrh, P., Viero, G., Serra, M.D., Podlesek, Z., Lakey, J.H., Maček, P., Menestrina, G., Anderluh, G., 2003. A novel mechanism of pore formation: membrane penetration by the N-terminal amphipathic of equinatoxin. Journal of Biological Chemistry 278, 22678–22685. https://doi.org/10.1074/jbc.M300622200
9. Rojko, N., Kristan, K.Č., Viero, G., Žerovnik, E., Maček, P., Serra, M.D., Anderluh, G., 2013. Membrane damage by an α-helical pore-forming protein, equinatoxin II, proceeds through a succession of ordered steps. Journal of Biological Chemistry 288, 23704–23715. https://doi.org/10.1074/jbc.M113.481572
10. Drechsler, A., Potrich, C., Sabo, J.K., Frisanco, M., Guella, G., Dalla Serra, M., Anderluh, G., Separovic, F., Norton, R.S., 2006. Structure and activity of the N-terminal region of the eukaryotic cytolysin equinatoxin II. Biochemistry 45, 1818–1828. https://doi.org/10.1021/bi052166o
11. Lam, Y.H., Hung, A., Norton, R.S., Separovic, F., Watts, A., 2010. Solid-state NMR and simulation studies of equinatoxin II N-terminus interaction with lipid bilayers. Proteins: Structure, Function, and Bioinformatics 78, 858–872. https://doi.org/10.1002/prot.22612
12. García-Linares, S., Richmond, R., García-Mayoral, M.F., Bustamante, N., Bruix, M., Gavilanes, J.G., Martínez-del-Pozo, Á., 2014. The sea anemone actinoporin (Arg-Gly-Asp) conserved motif is involved in maintaining the competent oligomerization state of these pore-forming toxins. The Federation of European Biochemical Societies Journal 281, 1465–1478. https://doi.org/10.1111/febs.12717
13. Tanaka, K., Caaveiro, J.M.M., Morante, K., González-Mañas, J.M., Tsumoto, K., 2015. Structural basis for self-assembly of a cytolytic pore lined by protein and lipid. Nature Communications 6, 6337. https://doi.org/10.1038/ncomms7337
14. Črnigoj Kristan, K., Viero, G., Dalla Serra, M., Maček, P., Anderluh, G., 2009. Molecular mechanism of pore formation by actinoporins. Toxicon, Cnidarian Toxins and Venoms 54, 1125–1134. https://doi.org/10.1016/j.toxicon.2009.02.026
15. Rojko, N., Dalla Serra, M., Maček, P., Anderluh, G., 2016. Pore formation by actinoporins, cytolysins from sea anemones. Biochimica et Biophysica Acta (BBA) - Biomembranes, Pore-forming Toxins: cellular effects and biotech applications 1858, 446–456. https://doi.org/10.1016/j.bbamem.2015.09.007
16. Mutter, N.L., Soskine, M., Huang, G., Albuquerque, I.S., Bernardes, G.J.L., Maglia, G., 2018. Modular Pore-forming immunotoxins with caged cytotoxicity tailored by directed evolution. The Journal of the American Chemical Society, Chemical Biology. 13, 3153–3160. https://doi.org/10.1021/acschembio.8b00720

17. Wloka, C., Mutter, N.L., Soskine, M., Maglia, G., 2016. Alpha-helical fragaceatoxin C nanopore engineered for double-stranded and single-stranded nucleic acid analysis. Angewandte Chemie International Edition 55, 12494–12498. https://doi.org/10.1002/anie.201606742
18. Makino, A., Abe, M., Murate, M., Inaba, T., Yilmaz, N., Hullin-Matsuda, F., Kishimoto, T., Schieber, N.L., Taguchi, T., Arai, H., Anderluh, G., Parton, R.G., Kobayashi, T., 2015. Visualization of the heterogeneous membrane distribution of sphingomyelin associated with cytokinesis, cell polarity, and sphingolipidosis. The FASEB Journal 29, 477–493. https://doi.org/10.1096/fj.13-247585
19. Laborde, R.J., Sanchez-Ferras, O., Luzardo, M.C., Cruz-Leal, Y., Fernández, A., Mesa, C., Oliver, L., Canet, L., Abreu-Butin, L., Nogueira, C.V., Tejuca, M., Pazos, F., Álvarez, C., Alonso, M.E., Longo-Maugéri, I.M., Starnbach, M.N., Higgins, D.E., Fernández, L.E., Lanio, M.E., 2017. Novel adjuvant based on the pore-forming protein sticholysin II encapsulated into liposomes effectively enhances the antigen-specific CTL-mediated immune response. The Journal of Immunology 198, 2772–2784. https://doi.org/10.4049/jimmunol.1600310
20. Macrander, J., Daly, M., 2016. Evolution of the cytolytic pore-forming proteins (actinoporins) in sea anemones. Toxins 8, 368. https://doi.org/10.3390/toxins8120368
21. Ben-Ari, H., Paz, M., Sher, D., 2018. The chemical armament of reef-building corals: inter- and intra-specific variation and the identification of an unusual actinoporin in *Stylophora pistilata*. Scientific Reports 8, 251. https://doi.org/10.1038/s41598-017-18355-1
22. Gutiérrez-Aguirre, I., Trontelj, P., Maček, P., Lakey, J.H., Anderluh, G., 2006. Membrane binding of zebrafish actinoporin-like protein: AF domains, a novel superfamily of cell membrane binding domains. Biochemical Journal 398, 381–392. https://doi.org/10.1042/BJ20060206
23. von Reumont, B.M., Campbell, L.I., Richter, S., Hering, L., Sykes, D., Hetmank, J., Jenner, R.A., Bleidorn, C., 2014. A polychaete's powerful punch: venom gland transcriptomics of *Glycera* reveals a complex cocktail of toxin homologs. Genome Biology and Evolution 6, 2406–2423. https://doi.org/10.1093/gbe/evu190
24. Gerdol, M., Luo, Y.-J., Satoh, N., Pallavicini, A., 2018. Genetic and molecular basis of the immune system in the brachiopod *Lingula anatina*. Developmental & Comparative Immunology 82, 7–30. https://doi.org/10.1016/j.dci.2017.12.021
25. Gerdol, M., Cervelli, M., Oliverio, M., Modica, M.V., 2018. Piercing Fishes: Porin expansion and adaptation to hematophagy in the vampire snail *Cumia reticulata*. Molecular Biology and Evolution 35, 2654–2668. https://doi.org/10.1093/molbev/msy156
26. Kawashima, Y., Nagai, H., Ishida, M., Nagashima, Y., Shiomi, K., 2003. Primary structure of echotoxin 2, an actinoporin-like hemolytic toxin from the salivary gland of the marine gastropod *Monoplex echo*. Toxicon 42, 491–497. https://doi.org/10.1016/S0041-0101(03)00226-5
27. Hoang, Q.T., Cho, S.H., McDaniel, S.F., Ok, S.H., Quatrano, R.S., Shin, J.S., 2009. An actinoporin plays a key role in water stress in the moss *Physcomitrella patens*. New Phytologist 184, 502–510. https://doi.org/10.1111/j.1469-8137.2009.02975.x
28. Takara, T., Nakagawa, T., Isobe, M., Okino, N., Ichinose, S., Omori, A., Ito, M., 2011. Purification, molecular cloning, and application of a novel sphingomyelin-binding protein (clamlysin) from the brackishwater clam, *Corbicula japonica*. Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids 1811, 323–332. https://doi.org/10.1016/j.bbalip.2011.02.004
29. Glasser, E., Rachamim, T., Aharonovich, D., Sher, D., 2014. Hydra actinoporin-like toxin-1, an unusual hemolysin from the nematocyst venom of *Hydra magnipapillata* which belongs to an extended gene family. Toxicon, special issue: Freshwater and marine toxins 91, 103–113. https://doi.org/10.1016/j.toxicon.2014.04.004
30. Luter, H.M., Bannister, R.J., Whalan, S., Kutti, T., Pineda, M.-C., Webster, N.S., 2017. Microbiome analysis of a disease affecting the deep-sea sponge *Geodia barretti*. FEMS Microbiology Ecology 93. https://doi.org/10.1093/femsec/fix074

31. Syue, S.-T., Hsu, C.-H., Soong, K., 2021. Testing of how and why the *Terpios hoshinota* sponge kills stony corals. Scientific Reports 11, 7661. https://doi.org/10.1038/s41598-021-87350-4
32. C. J., P.V., J. L., C., M. L., C., 2010. A Qualitative Assessment of Sponge-Feeding Organisms from the Mexican Pacific Coast. The Open Marine Biology Journal 4.
33. Laport, M.S., Santos, O.C.S., Muricy, G., 2009. Marine sponges: potential sources of new antimicrobial drugs. Current Pharmaceutical Biotechnology 10, 86–105. https://doi.org/10.2174/138920109787048625
34. Qi, S.-H., Ma, X., 2017. Antifouling compounds from marine invertebrates. Marine Drugs 15, 263. https://doi.org/10.3390/md15090263
35. Calcabrini, C., Catanzaro, E., Bishayee, A., Turrini, E., Fimognari, C., 2017. Marine sponge natural products with anticancer potential: an updated review. Marine Drugs 15, 310. https://doi.org/10.3390/md15100310
36. Müller, W.E.G., Wang, X., Binder, M., Lintig, J. von, Wiens, M., Schröder, H.C., 2012. Differential expression of the demosponge (*Suberites domuncula*) carotenoid oxygenases in response to light: protection mechanism against the self-produced toxic protein (suberitine). Marine Drugs 10, 177–199. https://doi.org/10.3390/md10010177
37. do Nascimento-Neto, L.G., Cabral, M.G., Carneiro, R.F., Silva, Z., Arruda, F.V.S., Nagano, C.S., Fernandes, A.R., Sampaio, A.H., Teixeira, E.H., Videira, P.A., 2018. Halilectin-3, a lectin from the marine sponge *Haliclona caerulea*, induces apoptosis and autophagy in human breast cancer MCF7 cells through caspase-9 pathway and LC3-II protein expression. Anti-Cancer Agents in Medicinal Chemistry 18, 521–528. https://doi.org/10.2174/1871520617666171114094847
38. Scarfì, S., Pozzolini, M., Oliveri, C., Mirata, S., Salis, A., Damonte, G., Fenoglio, D., Altosole, T., Ilan, M., Bertolino, M., Giovine, M., 2020. Identification, purification and molecular characterization of chondrosin, a new protein with anti-tumoral activity from the marine sponge *Chondrosia Reniformis* Nardo 1847. Marine Drugs 18, 409. https://doi.org/10.3390/md18080409
39. Mangel, A., Leitão, J.M., Batel, R., Zimmermann, H., Müller, W.E.G., Schröder, H.C., 1992. Purification and characterization of a pore-forming protein from the marine sponge *Tethya lyncurium*. European Journal of Biochemistry 210, 499–507. https://doi.org/10.1111/j.1432-1033.1992.tb17448.x
40. Wiens, M., Korzhev, M., Krasko, A., Thakur, N.L., Perović-Ottstadt, S., Breter, H.J., Ushijima, H., Diehl-Seifert, B., Müller, I.M., Müller, W.E.G., 2005. Innate immune defense of the sponge *Suberites domuncula* against bacteria involves a MyD88-dependent signaling pathway: induction of a perforin-like molecule. Journal of Biological Chemistry 280, 27949–27959. https://doi.org/10.1074/jbc.M504049200
41. Nichols, S.A., Roberts, B.W., Richter, D.J., Fairclough, S.R., King, N., 2012. Origin of metazoan cadherin diversity and the antiquity of the classical cadherin/β-catenin complex. The Proceedings of the National Academy of Sciences 109, 13046–13051. https://doi.org/10.1073/pnas.1120685109
42. Ereskovsky, A.V., Richter, D.J., Lavrov, D.V., Schippers, K.J., Nichols, S.A., 2017. Transcriptome sequencing and delimitation of sympatric *Oscarella* species (*O. carmela* and *O. pearsei* sp. nov) from California, USA. PLOS ONE 12, e0183002. https://doi.org/10.1371/journal.pone.0183002
43. Aguilar-Camacho, J.M., Doonan, L., McCormack, G.P., 2019. Evolution of the main skeleton-forming genes in sponges (phylum Porifera) with special focus on the marine Haplosclerida (class Demospongiae). Molecular Phylogenetics and Evolution 131, 245–253. https://doi.org/10.1016/j.ympev.2018.11.015
44. Kenny, N.J., Francis, W.R., Rivera-Vicéns, R.E., Juravel, K., de Mendoza, A., Díez-Vives, C., Lister, R., Bezares-Calderón, L.A., Grombacher, L., Roller, M., Barlow, L.D., Camilli, S., Ryan, J.F., Wörheide, G., Hill, A.L., Riesgo, A., Leys, S.P., 2020. Tracing animal genomic evolution with the chromosomal-level assembly of the freshwater sponge *Ephydatia muelleri*. Nature Communications 11, 3676. https://doi.org/10.1038/s41467-020-17397-w
45. Kenny, N.J., Plese, B., Riesgo, A., Itskovich, V.B., 2019. Symbiosis, selection, and novelty: freshwater adaptation in the unique sponges of Lake Baikal. Molecular Biology and Evolution 36, 2462–2480. https://doi.org/10.1093/molbev/msz151

46. Francis, W.R., Eitel, M., Vargas, S., Adamski, M., Haddock, S.H.D., Krebs, S., Blum, H., Erpenbeck, D., Wörheide, G., 2017. The genome of the contractile demosponge *Tethya wilhel*ma and the evolution of metazoan neural signalling pathways. bioRxiv 120998. https://doi.org/10.1101/120998
47. Ryu, T., Seridi, L., Moitinho-Silva, L., Oates, M., Liew, Y.J., Mavromatis, C., Wang, X., Haywood, A., Lafi, F.F., Kupresanin, M., Sougrat, R., Alzahrani, M.A., Giles, E., Ghosheh, Y., Schunter, C., Baumgarten, S., Berumen, M.L., Gao, X., Aranda, M., Foret, S., Gough, J., Voolstra, C.R., Hentschel, U., Ravasi, T., 2016. Hologenome analysis of two marine sponges with different microbiomes. BioMed Central Genomics 17, 158. https://doi.org/10.1186/s12864-016-2501-0
48. Fortunato, S.A.V., Adamski, M., Ramos, O.M., Leininger, S., Liu, J., Ferrier, D.E.K., Adamska, M., 2014. Calcisponges have a ParaHox gene and dynamic expression of dispersed NK homeobox genes. Nature 514, 620–623. https://doi.org/10.1038/nature13881
49. Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M.E.A., Mitros, T., Richards, G.S., Conaco, C., Dacre, M., Hellsten, U., Larroux, C., Putnam, N.H., Stanke, M., Adamska, M., Darling, A., Degnan, S.M., Oakley, T.H., Plachetzki, D.C., Zhai, Y., Adamski, M., Calcino, A., Cummins, S.F., Goodstein, D.M., Harris, C., Jackson, D.J., Leys, S.P., Shu, S., Woodcroft, B.J., Vervoort, M., Kosik, K.S., Manning, G., Degnan, B.M., Rokhsar, D.S., 2010. The *Amphimedon queenslandica* genome and the evolution of animal complexity. Nature 466, 720–726. https://doi.org/10.1038/nature09201
50. Koutsouveli, V., Cárdenas, P., Santodomingo, N., Marina, A., Morato, E., Rapp, H.T., Riesgo, A., 2020. The molecular machinery of gametogenesis in *Geodia* demosponges (Porifera): evolutionary origins of a conserved toolkit across animals. Molecular Biology and Evolution 37, 3485–3506. https://doi.org/10.1093/molbev/msaa183
51. Finoshin, A.D., Adameyko, K.I., Mikhailov, K.V., Kravchuk, O.I., Georgiev, A.A., Gornostaev, N.G., Kosevich, I.A., Mikhailov, V.S., Gazizova, G.R., Shagimardanova, E.I., Gusev, O.A., Lyupina, Y.V., 2020. Iron metabolic pathways in the processes of sponge plasticity. PLOS ONE 15, e0228722. https://doi.org/10.1371/journal.pone.0228722
52. Manousaki, T., Koutsouveli, V., Lagnel, J., Kollias, S., Tsigenopoulos, C.S., Arvanitidis, C., Magoulas, A., Dounas, C., Dailianis, T., 2019. A *de novo* transcriptome assembly for the bath sponge *Spongia officinalis*, adjusting for microsymbionts. BioMed Central Research Notes 12, 813. https://doi.org/10.1186/s13104-019-4843-6
53. González-Aravena, M., Kenny, N.J., Osorio, M., Font, A., Riesgo, A., Cárdenas, C.A., 2019. Warm temperatures, cool sponges: the effect of increased temperatures on the Antarctic sponge *Isodictya* sp. PeerJ 7, e8088. https://doi.org/10.7717/peerj.8088
54. Leiva, C., Taboada, S., Kenny, N.J., Combosch, D., Giribet, G., Jombart, T., Riesgo, A., 2019. Population substructure and signals of divergent adaptive selection despite admixture in the sponge *Dendrilla antarctica* from shallow waters surrounding the Antarctic Peninsula. Molecular Ecology 28, 3151–3170. https://doi.org/10.1111/mec.15135
55. Pita, L., Hoeppner, M.P., Ribes, M., Hentschel, U., 2018. Differential expression of immune receptors in two marine sponges upon exposure to microbial-associated molecular patterns. Scientific Reports 8, 16081. https://doi.org/10.1038/s41598-018-34330-w
56. Revilla-i-Domingo, R., Schmidt, C., Zifko, C., Raible, F., 2018. Establishment of transgenesis in the demosponge *Suberites domuncula*. Genetics 210, 435–443. https://doi.org/10.1534/genetics.118.301121
57. Kenny, N.J., de Goeij, J.M., de Bakker, D.M., Whalen, C.G., Berezikov, E., Riesgo, A., 2018. Towards the identification of ancestrally shared regenerative mechanisms across the Metazoa: a transcriptomic case study in the Demosponge *Halisarca caerulea*. Marine Genomics 37, 135–147. https://doi.org/10.1016/j.margen.2017.11.001
58. Leys, S., 2017. *Ephydatia muelleri* Trinity transcriptome [WWW Document]. Education & Research Archive. https://doi.org/10.7939/R3WH2DV20

59. Leys, S., 2017. *Eunapius fragilis* Trinity transcriptome [WWW Document]. Education & Research Archive. https://doi.org/10.7939/R3794177K
60. Simion, P., Philippe, H., Baurain, D., Jager, M., Richter, D.J., Di Franco, A., Roure, B., Satoh, N., Quéinnec, É., Ereskovsky, A., Lapébie, P., Corre, E., Delsuc, F., King, N., Wörheide, G., Manuel, M., 2017. A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. Current Biology 27, 958–967. https://doi.org/10.1016/j.cub.2017.02.031
61. Díez-Vives, C., Moitinho-Silva, L., Nielsen, S., Reynolds, D., Thomas, T., 2017. Expression of eukaryotic-like protein in the microbiome of sponges. Molecular Ecology 26, 1432–1451. https://doi.org/10.1111/mec.14003
62. Borisenko, I., Adamski, M., Ereskovsky, A., Adamska, M., 2016. Surprisingly rich repertoire of Wnt genes in the demosponge *Halisarca dujardini*. BioMed Central Evolutionary Biology 16, 123. https://doi.org/10.1186/s12862-016-0700-6
63. Guzman, C., Conaco, C., 2016. Comparative transcriptome analysis reveals insights into the streamlined genomes of haplosclerid demosponges. Scientific Reports 6, 18774. https://doi.org/10.1038/srep18774
64. Alié, A., Hayashi, T., Sugimura, I., Manuel, M., Sugano, W., Mano, A., Satoh, N., Agata, K., Funayama, N., 2015. The ancestral gene repertoire of animal stem cells. The Proceedings of the National Academy of Sciences USA 112, E7093–E7100. https://doi.org/10.1073/pnas.1514789112
65. Qiu, F., Ding, S., Ou, H., Wang, D., Chen, J., Miyamoto, M.M., 2015. Transcriptome changes during the life cycle of the red sponge, *Mycale phyllophila* (Porifera, Demospongiae, Poecilosclerida). Genes 6, 1023–1052. https://doi.org/10.3390/genes6041023
66. Whelan, N.V., Kocot, K.M., Moroz, L.L., Halanych, K.M., 2015. Error, signal, and the placement of Ctenophora sister to all other animals. The Proceedings of the National Academy of Sciences 112, 5773–5778. https://doi.org/10.1073/pnas.1503453112
67. Riesgo, A., Peterson, K., Richardson, C., Heist, T., Strehlow, B., McCauley, M., Cotman, C., Hill, M., Hill, A., 2014. Transcriptomic analysis of differential host gene expression upon uptake of symbionts: a case study with *Symbiodinium* and the major bioeroding sponge *Cliona varians*. BioMed Central Genomics 15, 376. https://doi.org/10.1186/1471-2164-15-376
68. Riesgo, A., Farrar, N., Windsor, P.J., Giribet, G., Leys, S.P., 2014. The analysis of eight transcriptomes from all poriferan classes reveals surprising genetic complexity in sponges. Molecular Biology and Evolution 31, 1102–1120. https://doi.org/10.1093/molbev/msu057
69. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A., 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 8, 1494–1512. https://doi.org/10.1038/nprot.2013.084
70. Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Research 39, W29–W37. https://doi.org/10.1093/nar/gkr367
71. Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J., Punta, M., 2014. Pfam: the protein families database. Nucleic Acids Research 42, D222–D230. https://doi.org/10.1093/nar/gkt1223
72. Lu, S., Wang, J., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Marchler, G.H., Song, J.S., Thanki, N., Yamashita, R.A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C.J., Marchler-Bauer, A., 2020. CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Research 48, D265–D268. https://doi.org/10.1093/nar/gkz991
73. King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I., Marr, M., Pincus, D., Putnam, N., Rokas, A., Wright, K.J., Zuzow, R., Dirks, W., Good, M., Goodstein, D., Lemons, D., Li, W., Lyons, J.B., Morris, A., Nichols, S., Richter, D.J., Salamov, A., Sequencing, J., Bork, P., Lim, W.A., Manning, G., Miller, W.T., McGinnis, W., Shapiro, H., Tjian, R., Grigoriev, I.V., Rokhsar, D., 2008. The

genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. Nature 451, 783–788. https://doi.org/10.1038/nature06617

74. Fairclough, S.R., Chen, Z., Kramer, E., Zeng, Q., Young, S., Robertson, H.M., Begovic, E., Richter, D.J., Russ, C., Westbrook, M.J., Manning, G., Lang, B.F., Haas, B., Nusbaum, C., King, N., 2013. Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate *Salpingoeca rosetta*. Genome Biology 14, R15. https://doi.org/10.1186/gb-2013-14-2-r15
75. Ryan, J.F., Pang, K., Schnitzler, C.E., Nguyen, A.-D., Moreland, R.T., Simmons, D.K., Koch, B.J., Francis, W.R., Havlak, P., Smith, S.A., Putnam, N.H., Haddock, S.H.D., Dunn, C.W., Wolfsberg, T.G., Mullikin, J.C., Martindale, M.Q., Baxevanis, A.D., 2013. The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. Science 342. https://doi.org/10.1126/science.1242592
76. Moroz, L.L., Kocot, K.M., Citarella, M.R., Dosung, S., Norekian, T.P., Povolotskaya, I.S., Grigorenko, A.P., Dailey, C., Berezikov, E., Buckley, K.M., Ptitsyn, A., Reshetov, D., Mukherjee, K., Moroz, T.P., Bobkova, Y., Yu, F., Kapitonov, V.V., Jurka, J., Bobkov, Y.V., Swore, J.J., Girardo, D.O., Fodor, A., Gusev, F., Sanford, R., Bruders, R., Kittler, E., Mills, C.E., Rast, J.P., Derelle, R., Solovyev, V.V., Kondrashov, F.A., Swalla, B.J., Sweedler, J.V., Rogaev, E.I., Halanych, K.M., Kohn, A.B., 2014. The ctenophore genome and the evolutionary origins of neural systems. Nature 510, 109–114. https://doi.org/10.1038/nature13400
77. Srivastava, M., Begovic, E., Chapman, J., Putnam, N.H., Hellsten, U., Kawashima, T., Kuo, A., Mitros, T., Salamov, A., Carpenter, M.L., Signorovitch, A.Y., Moreno, M.A., Kamm, K., Grimwood, J., Schmutz, J., Shapiro, H., Grigoriev, I.V., Buss, L.W., Schierwater, B., Dellaporta, S.L., Rokhsar, D.S., 2008. The *Trichoplax* genome and the nature of placozoans. Nature 454, 955–960. https://doi.org/10.1038/nature07191
78. Eitel, M., Francis, W.R., Varoqueaux, F., Daraspe, J., Osigus, H.-J., Krebs, S., Vargas, S., Blum, H., Williams, G.A., Schierwater, B., Wörheide, G., 2018. Comparative genomics and the nature of placozoan species. PLOS Biology 16, e2005359. https://doi.org/10.1371/journal.pbio.2005359
79. Richter, D.J., Fozouni, P., Eisen, M.B., King, N., 2018. Gene family innovation, conservation and loss on the animal stem lineage. eLife 7, e34226. https://doi.org/10.7554/eLife.34226
80. Pruitt, K.D., Tatusova, T., Maglott, D.R., 2005. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Research 33, D501–D504. https://doi.org/10.1093/nar/gki025
81. Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., Nielsen, H., 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nature Biotechnology 37, 420–423. https://doi.org/10.1038/s41587-019-0036-z
82. Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., Stockinger, H., 2012. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Research 40, W597–W603. https://doi.org/10.1093/nar/gks400
83. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols 10, 845–858. https://doi.org/10.1038/nprot.2015.053
84. Uziela, K., Menéndez Hurtado, D., Shu, N., Wallner, B., Elofsson, A., 2017. ProQ3D: improved model quality assessments using deep learning. Bioinformatics 33, 1578–1580. https://doi.org/10.1093/bioinformatics/btw819
85. McGuffin, L.J., Aldowsari, F.M.F., Alharbi, S.M.A., Adiyaman, R., 2021. ModFOLD8: accurate global and local quality estimates for 3D protein models. Nucleic Acids Research 49, W425–W430. https://doi.org/10.1093/nar/gkab321
86. Zhang, Y., Skolnick, J., 2005. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Research 33, 2302–2309. https://doi.org/10.1093/nar/gki524

87. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera—a visualization system for exploratory research and analysis. Journal of Computational Chemistry 25, 1605–1612. https://doi.org/10.1002/jcc.20084
88. Stivala, A., Wybrow, M., Wirth, A., Whisstock, J.C., Stuckey, P.J., 2011. Automatic generation of protein structure cartoons with Pro-origami. Bioinformatics 27, 3315–3316. https://doi.org/10.1093/bioinformatics/btr575
89. Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 30, 772–780. https://doi.org/10.1093/molbev/mst010
90. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J., 2009. Jalview version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191. https://doi.org/10.1093/bioinformatics/btp033
91. Gautier, R., Douguet, D., Antonny, B., Drin, G., 2008. HELIQUEST: a web server to screen sequences with specific α-helical properties. Bioinformatics 24, 2101–2102. https://doi.org/10.1093/bioinformatics/btn392
92. Mól, A.R., Castro, M.S., Fontes, W., 2018. NetWheels: a web application to create high quality peptide helical wheel and net projections. bioRxiv 416347. https://doi.org/10.1101/416347
93. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015
94. Letunic, I., Bork, P., 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Research 49, W293–W296. https://doi.org/10.1093/nar/gkab301
95. Anderluh, G., Barlič, A., Potrich, C., Maček, P., Menestrina, G., 2000. Lysine 77 is a key residue in aggregation of equinatoxin II, a Pore-forming toxin from sea anemone *Actinia equina*. The Journal of Membrane Biology 173, 47–55. https://doi.org/10.1007/s002320001006
96. Sandoval, K., McCormack, G. 2022. Transcriptome assemblies from marine haplosclerid species (phylum Porifera, class Demospongia) collected in Ireland, with additional data from analyzing actinoporin-like proteins from these and other poriferan genome/transcriptome assemblies. Mendeley Data. https://doi.org/10.17632/w9t6zsjjb7.1
97. Sandoval, K., McCormack, G.P., 2022. Actinoporin-like proteins are widely distributed in the phylum Porifera. Marine Drugs 20, 74. https://doi.org/10.3390/md20010074
98. Morante, K., Caaveiro, J.M.M., Viguera, A.R., Tsumoto, K., González-Mañas, J.M., 2015. Functional characterization of Val60, a key residue involved in the membrane-oligomerization of fragaceatoxin C, an actinoporin from *Actinia fragacea*. FEBS Letters 589, 1840–1846. https://doi.org/10.1016/j.febslet.2015.06.012
99. Peraro, M.D., van der Goot, F.G., 2016. Pore-forming toxins: ancient, but never really out of fashion. Nature Reviews Microbiology 14, 77–92. https://doi.org/10.1038/nrmicro.2015.3
100. Spicer, B.A., Conroy, P.J., Law, R.H.P., Voskoboinik, I., Whisstock, J.C., 2017. Perforin—a key (shaped) weapon in the immunological arsenal. Seminars in Cell & Developmental Biology, Skeletal muscle development on the 30th anniversary of MyoD 72, 117–123. https://doi.org/10.1016/j.semcdb.2017.07.033
101. Basulto, A., Pérez, V.M., Noa, Y., Varela, C., Otero, A.J., Pico, M.C., 2006. Immunohistochemical targeting of sea anemone cytolysins on tentacles, mesenteric filaments and isolated nematocysts of *Stichodactyla helianthus*. Journal of Experimental Zoology Part A: Comparative Experimental Biology 305A, 253–258. https://doi.org/10.1002/jez.a.256
102. Hong, Q., Gutiérrez-Aguirre, I., Barlič, A., Malovrh, P., Kristan, K., Podlesek, Z., Maček, P., Turk, D., González-Mañas, J.M., Lakey, J.H., Anderluh, G., 2002. Two-step membrane binding by equinatoxin II, a

pore-forming toxin from the sea anemone, involves an exposed aromatic cluster and a flexible helix. Journal of Biological Chemistry 277, 41916–41924. https://doi.org/10.1074/jbc.M204625200

103. Anderluh, G., Razpotnik, A., Podlesek, Z., Maček, P., Separovic, F., Norton, R.S., 2005. Interaction of the eukaryotic pore-forming cytolysin equinatoxin II with model membranes: ^{19}F NMR studies. Journal of Molecular Biology 347, 27–39. https://doi.org/10.1016/j.jmb.2004.12.058

104. Valle, A., Alvarado-Mesén, J., Lanio, M.E., Álvarez, C., Barbosa, J.A.R.G., Pazos, I.F., 2015. The multigene families of actinoporins (part I): Isoforms and genetic structure. Toxicon 103, 176–187. https://doi.org/10.1016/j.toxicon.2015.06.028

105. Jiang, X., Chen, H., Yang, W., Liu, Y., Liu, W., Wei, J., Tu, H., Xie, X., Wang, L., Xu, A., 2003. Functional expression and characterization of an acidic actinoporin from sea anemone *Sagartia rosea*. Biochemical and Biophysical Research Communications 312, 562–570. https://doi.org/10.1016/j.bbrc.2003.10.159

Chapter 6

General Discussion

Part of this chapter contributed to the as of yet unpublished manuscript:
Sandoval, K., Thomas, O.P., McCormack, G.P., 2022. From Monomers to Manzamines: Observed Patterns, Biosynthetic Considerations and Chemical Ecology of 3-alkylpyridine Alkaloids from the Order Haplosclerida

Discovery and characterization of Novel Poriferan Biosynthetic Pathways *via* Next-Generation Sequencing

A focused, main objective of this thesis was the identification of a biosynthetic origin of 3-alkylpyridine alkaloids (3-APs) from the Order Haplosclerida *via* next-generation sequencing and genome mining techniques. Specifically, I sought to test the hypothesis that the common monomeric units of these alkaloids are produced by a polyketide synthase (PKS) which accepts nicotinic acid (NTA) as a starter unit as such genes should be readily detectable by available genome mining software [1]. Based on this hypothesis, two features were expected to be identified in a single open reading frame (ORF) or grouped together in a biosynthetic gene cluster (BGC). First, an enzyme or enzymatic domain that could activate NTA to be incorporated into the biosynthetic pathway akin to what is seen with several polyketide natural products [2-5]. Second, a multidomain enzyme which contains the necessary PKS domains to catalyse a Claisen condensation with NTA and an extender unit as well as sequentially reduce the ketone backbone [5]. Based on the lack of evidence for common microbial symbionts between two Irish species that are known sources of 3-APs, *Haliclona indistincta* and *Haliclona viscosa*, a possible animal origin of these compounds was also considered [6]. Furthermore, chemical extracts from these two species have displayed selective activity towards various tumoral cell lines which makes them attractive targets for biodiscovery purposes (Grace McCormack, personal communication). While 3-APs from these species are the likely source of such activity, it could also be that other undetected compounds with promising bioactivity may also be isolatable. Based on this, a broader, secondary objective was the identification of additional genes that could have relevance to the production of other bioactive small molecules and proteins.

As NTA and an alkyl chain, the latter likely derived from a polyketide or fatty acid, are the signature motifs of 3-APs I first sought to determine if *H. indistincta* and *H. viscosa* could synthesize these building blocks. I approached this problem by mapping protein-coding genes to known metabolic pathways associated with the production of these metabolites (Chapter 2) [7]. While the sequenced *Haliclona* transcriptomes likely do not represent every protein-coding gene found in the genomes of these two species, it was expected that the genes responsible for 3-APs would have been expressed at the time of sampling and flash-freezing. This is due to the detergent-like substance associated with these two species, which was copiously exuded during the dissection process prior to sample preservation, being hypothesized to be comprised of 3-AP polymers (Grace McCormack and Olivier Thomas, personal communication). This analysis produced two key findings. First, congruent with isotope-labelling studies, both *H. indistincta* and *H. viscosa* have a possible pathway for the biosynthesis of NTA from tryptophan [8]. Second, no discernible pathway accounting for the biosynthesis of the 3-AP alkyl chain could be identified, although a complete degradation pathway of fatty acids was present. This latter finding, while surprising, is consistent with the lack of a complete pathway for the biosynthesis of fatty acids in the *Amphimedon*

queenslandica genome [9]. The implications of this observation may be that if the 3-APs are produced by the sponge then the alkyl chain could be derived from a dietary source rather than *de novo* biosynthesis [10]. That is, following consumption dietary fatty acids may be degraded to the appropriate length for 3-AP biosynthesis. Such an observation may explain why numerous 3-APs which are nearly identical in structure except for the length of their alkyl chain can be identified from a single specimen [11]. While the utilized transcriptomic methods were insightful on the biosynthesis of potential 3-AP building blocks by *H. indistincta* and *H. viscosa*, no identifiable enzymes which would carry out the actual biosynthesis of 3-APs could be identified. This could be due to several possible factors such as a lack of sequencing depth, the fragmented nature of the transcriptomes, or a bacterial origin. Due to these limitations, genome sequencing, which would overcome such problems, was deemed necessary.

Long-read sequencing with Pacific Biosciences technology produced a highly contiguous genome of *H. indistincta*, although it appears that the very 3-APs which were the focus of this project may have been co-purified and interfered with sequencing as is evident by the disparity between the length of the isolated DNA fragments versus the length of the sequenced reads (Chapter 3). In truth, the specialized metabolites of this species likely hampered other attempts at producing next-generation sequencing data for the assembly of a reference-quality genome. Chromosome conformation capture was also attempted on samples of *H. indistincta*, although it was evident that the utilized Arima-HiC Kit was designed for model organisms as the characteristic detergent-like slime was consistently co-purified (personal observation) [12]. This likely translated to the erroneous Hi-C sequencing data which did not map back properly to the genome assembly (data not shown). Still, the draft genome assembly was sufficient for testing the PKS hypothesis on 3-AP biosynthesis. Specifically, only a single gene which fit said hypothesis, that for the hybrid nonribosomal peptide synthetase-polyketide synthase (NRPS-PKS), was identified on a contig likely derived from the sponge chromosome. Surprisingly, no PKS genes were detected in the metagenome-assembled genomes (MAGs) despite bacteria associated with sponges being a consistent source of these genes [13]. Two main features of the sponge NRPS-PKS partially satisfy the PKS hypothesis. First, its two adenylation domains can be considered a type of AMP-binding enzymatic domain. The presence of these adenylation domains has relevance to 3-AP biosynthesis as a standalone adenylation domain has been previously shown to activate NTA for incorporation into a polyketide [3]. Second, its acyltransferase and ketosynthase domains would be expected to activate a PKS extender unit, such as malonyl-CoA, and catalyse a Claisen condensation reaction respectively. This reaction is used to produce the molluscan haminols which are structurally similar to monomeric 3-APs and is the foundation for the PKS hypothesis of 3-APs [14,15]. However, several characteristics of this NRPS-PKS make connecting it with 3-AP biosynthesis difficult. Regarding the nonribosomal peptide synthetase (NRPS) portion, attempts to predict the substrate

specificity of the adenylation domains, particularly whether relevant features have consistency with those which activate NTA for polyketide biosynthesis, were inconclusive [2-5,16-18]. Furthermore, the presence of a condensation domain strongly suggested that a peptide bond is formed during the biosynthetic process of this NRPS-PKS and such a feature is absent in all reported 3-APs [19]. However, it should be noted that condensation domains are not limited to this reaction with some catalysing the conjugation of an activated fatty acid with the amine of an amino acid [20]. Such a reaction bears some relevance to the carbon-nitrogen bond which oligomerized 3-APs possess. Moving to the PKS portion, as the 3-AP polymer of *H. indistincta* exhibits a fully reduced alkyl chain it would be expected that a responsible enzyme would contain ketoreductase, dehydratase and enoylreductase domains akin to that which produces pyridonopyrone A [5]. However, such domains are absent in the NRPS-PKS which necessitates alternative hypotheses to explain how it may produce a 3-AP monomeric unit. One possible method is that extender units derived from malonyl-CoA are sequentially added upon a NTA starter unit in a typical PKS fashion. Following this, the elongated ketone backbone could be fully reduced by separate t*rans*-acting enzymes akin to the enoylreductase involved in the biosynthesis of tenellin and the ketoreductase for the polyketide dimer SIA7248 [21,22]. Alternatively, a mature, fully reduced fatty acid could be conjugated to NTA *via* a Claisen condensation akin to the biosynthesis of norsolorinic acid anthrone which involves a ketosynthase conjugating a fatty acid to a polyketide [23].

Based on the results of genome sequencing, it can be concluded that 3-APs are likely not produced *via* a PKS which contains ketoreductase, dehydratase and enoylreductase enzymatic domains as has been previously hypothesized (Chapter 3) [1]. However, the high degree of sequence novelty which rendered substrate prediction for the NRPS-PKS of *H. indistincta* inconclusive does not necessarily mean that this enzyme is disproven to being involved in 3-AP biosynthesis. In truth, while *in silico* predictive methods are powerful tools for estimating the function of an enzyme related to the biosynthesis of a specialized metabolite, often these predictive methods are based on bacterial and fungal data [16,24]. As such, applying these methods to metazoan biosynthetic enzymes such as the *H. indistincta* NRPS-PKS likely resulted in the low predictive confidence. To truly determine what biochemical reactions an enzyme catalyses *in vitro* experimentation is necessary. For this reason I pursued the heterologous expression of the NRPS-PKS gene in *Saccharomyces cerevisiae* with the intention of functionally characterizing the produced enzyme akin to what has been done with molluscan and echinoderm PKS genes (Chapter 4) [25,26]. Despite the NRPS-PKS gene being accurately synthesized, cloned into the expression vector XW-55, and then transformed into *S. cerevisiae* strain BJ5464-NpgA no recombinant protein was detected. In turn, due to time constraints I was not able to further pursue this aspect of my project. Still, these results revealed important information as to the feasibility on attempting heterologous expression of the entire NRPS-PKS ORF. In turn, several paths to overcome this difficulty can be envisioned. For one, a deeper transcriptome is likely necessary to determine

whether introns are present within the NRPS-PKS gene similar to the previously mentioned molluscan PKS [25]. Additionally, alternative heterologous hosts with potentially better capabilities of splicing introns should be considered (Eric Schmidt, personal communication) [27]. Finally, the heterologous expression and biochemical characterization of individual enzymatic domains should also be considered as has been done with previous metazoan PKS genes [28]. If the adenylation domains were individually expressed their ability to activate NTA could be assessed by means such as the malachite green assay; such an analysis would easily confirm or disconfirm the role of this NRPS-PKS gene in 3-AP biosynthesis [29].

Overall while I have identified a candidate gene for the biosynthesis of 3-APs, I was not able to successfully determine if it is responsible for these compounds *via in silico* and *in vitro* methods (Chapter 2, Chapter 3, Chapter 4). If the NRPS-PKS is not responsible for the 3-AP polymers of *H. indistincta*, the notion that common monomeric building blocks are produced by any sort of megasynthase can be considered false. In turn, this will instead imply that a multitude of smaller enzymes are likely responsible for catalysing the necessary biosynthetic reactions in cohesion. Such a scenario presents a far more difficult pathway towards determining the biosynthetic origin of these compounds as no other obvious BGCs were identified in the *H. indistincta* genome or microbial MAGs (Chapter 3). Instead, it may be that the 3-AP genes are not clustered in one region of a chromosome as is observed with those linked to nemamide biosynthesis which thus renders genome mining for BGCs ineffective [30]. Besides genome mining, a second approach to identifying biosynthetic enzymes of uncharacterized pathways is through differential transcriptomics. For example, the accountable genes may be locally expressed in certain tissues of an organism as seen with the conazolium compounds from cone snails [31]. Alternatively, culture conditions can result in the overexpression of biosynthetic genes as was used to identify the BGC associated with domoic acid production in diatoms [32]. The latter approach presents an interesting possibility to identify 3-AP biosynthetic genes as sustained cell cultures of the species *Amphimedon compressa* and *Amphimedon erina*, both of which are known sources of these compounds, has been achieved [33]. Considering the role NTA plays in the biosynthesis of 3-APs it could represent a potential variable to induce differential expression. For example, the production of 3-APs may be a method of detoxifying NTA from the sponge while producing a bioactive specialized metabolite as is done with certain plant alkaloids [34]. However, considering the difficulty of isolating 3-AP polymers, a different species to *H. indistincta* should be considered as these are the only 3-AP compounds known to be derived from it (Olivier Thomas, personal communication). Based on the reported success of culturing *A. compressa* cells as well as it being the source of the much smaller 8,8′-dienecyclostellettamine this species could represent a promising target for this goal [35].

Moving from 3-APs, the presence of similar NRPS-PKS genes in the transcriptomes and genomes of other demosponges as well as NRPS genes in those from the class

Homoscleromorpha is a significant finding which expands upon what is known about specialized metabolism in not only the phylum Porifera but all metazoans (Chapter 3). To date, the only published metazoan hybrid megasynthases genes have been from nematodes whereas no modular metazoan NRPS genes have been published [36,37]. Genomics and transcriptomics have indicated that sponges possess the genes for the biosynthesis of bioactive natural products such as sterols, ovothiols and peptides [38-40]. However, until now no evidence of megasynthase genes have been reported from this phylum despite numerous sponge-derived polyketides, nonribosomal peptides, and hybrids of the two being reported [41-43]. Rather, most sponge-derived natural products have been credited to associated microorganisms [44,45]. Now, I have shown that a diverse array of sponges can be considered likely sources of nitrogenous specialized metabolites based on their possession of NRPS and NRPS-PKS genes. However, like that of *H. indistincta* these other sponge megasynthases exhibit a lack of sequence similarity to anything in public databases. In turn, this results in functional prediction of their enzymatic domains also being inconclusive even when focusing on conserved residues [17,18]. If the NRPS-PKS genes were responsible for the biosynthesis of the same compound, one would expect similar natural products to be isolated from the species which possess said genes as is observed with other metazoan natural products [25]. However, such a comparison cannot currently be made due to a lack of chemical analysis on most species which possess these genes. Other than *H. indistincta*, the only species which possesses a NRPS-PKS and has been shown to yield bioactive specialized metabolites is *Stylissa carteri* which is a consistent source of alkaloids, albeit not 3-APs [46].

While these sponge megasynthase genes are found on the chromosome of the animal, it is difficult to say whether or not they are the result of horizontal gene transfer from bacteria. Such an event has been previously reported as the origin of one sponge gene and is likely due to the close association these animals often have with bacteria [47]. Unlike previously characterized metazoan PKS genes, these sponge NRPS-PKSs do not exhibit a strong degree of sequence similarity to those of other metazoans [25,26]. Rather, they consistently display the highest, albeit still low, sequence similarity to bacterial megasynthases in the NCBI nonredundant database which would be expected of a horizontally transferred gene. Phylogenetic analysis of individual PKS enzymatic domains from nematodes has been used to estimate whether they are the result of horizontal gene transfer or gene duplication from a fatty acid synthase; unfortunately the lack of fatty acid synthase genes from not only *H. indistincta* but other sponges as well renders this method inapplicable for the NRPS-PKS genes of this phylum [36]. Still, a maximum likelihood tree of the sponge NRPS-PKS ketosynthase enzymatic domains indicates a well-resolved clade with statistical support separating it from any of the closest bacterial or eukaryotic sequences (Chapter 3). Initially, the topology indicates a closer phylogenetic signal to bacterial sequences, but a lack of statistical support for any of the connecting nodes renders said topology unreliable. As no

similar genes were found to be present in choanoflagellates or other non-metazoan eukaryotes, inheritance *via* a eukaryotic ancestor appears unlikely unless a series of gene losses occurred. Based on this as well as a lack of evidence for a gene duplication event from a similar megasynthase, horizontal transfer remains the most likely explanation for the specific presence of this gene in diverse sponge genomes. Further analysis on the other enzymatic domains, such as the well conserved adenylation and condensation domains, may provide better insight on the origin of these genes.

A secondary objective of mine was to identify genes associated with undiscovered natural products from *H. indistincta* and *H. viscosa* alternative to the polymeric 3-APs. This is because while these compounds exhibit strong cytotoxicity, they are often nonselective which may hamper their development into clinically approved drugs (Grace McCormack, personal communication) [48]. Sponges are a well-known source of sterols with unique skeletons not seen in any other phylum [49]. As structurally unique sterols have been isolated from previously analysed *Haliclona* species I reasoned that similar compounds may be derivable from *H. indistincta* and *H. viscosa* (Chapter 2) [50-51]. Consistent with a prior analysis of the haplosclerid species *A. queenslandica*, *H. indistincta* and *H. viscosa* appear capable of creating common animal sterols as well as catalysing several reactions associated with plant sterol biosynthesis [9]. The lack of a complete plant sterol biosynthetic pathway may be an indication that the comprehensive sterol biosynthetic pathway of sponges cannot be neatly conformed to those typically associated with plants and other animals. Overall, it is not unreasonable to hypothesize that these species can be sources of cytotoxic sterols, although chemical analyses will be necessary to confirm such a presence.

Unexpectedly, *H. indistincta* and *H. viscosa* were also found to contain a variety of genes encoding for actinoporin-like proteins (ALPs) (Chapter 5). As these small proteinaceous toxins are known for their cytolytic activity, they too should be considered as possible pharmaceutical candidates [53-56]. Such a hypothesis is further supported by *in silico* analyses of conserved residues and secondary structure of one ALP from *H. indistincta*. Furthermore, the closer phylogenetic relationship between sponge ALPs and cnidarian actinoporins (APs) rather than to ALPs from other animals may also serve as an indicator that they are capable of a similar cytotoxic function. Extrapolating upon this, the presence of similar genes in the genomes and transcriptomes of species from the Classes Demospongiae, Heterosclermorpha and Calcarea shows an aspect of the Phylum which could be considered for biotechnological and therapeutic applications akin to APs [57]. This widespread distribution perhaps indicates a common purpose of the molecules to the phylum, although it is difficult to envision what such a purpose is. Often, those APs and ALPs associated with cnidarians and molluscs play a role in offensive envenomation which, barring two exceptions, is an absent process in sponges [53,56,58-61]. However, many non-cnidarian ALPs are associated with animals incapable of envenomation such as bivalves which may imply a defensive role in such organisms as well as sponges [55]. While not an ALP, a cytolytic

perforin-like protein has been identified from *Suberites domuncula* and was shown to be upregulated when the sponge is exposed to lipopolysaccharides [62]. As sponges are constantly exposed to microorganisms due to their filter feeding nature, it can be reasoned that these pore-forming proteins, such as the sponge ALPs, instead serve a purpose of eliminating antagonistic microorganisms. However, to truly determine the ecological role as well as applied potential of these sponge ALPs, cloning, heterologous expression and functional studies will be necessary. Fortunately, the small size of these proteins has allowed *Escherichia coli* to function as a heterologous host for cnidarian APs in the past which indicates feasibility in attempting the same methodology on the sponge ALPs. [63,64].

Concluding Remarks

While I have not been able to solve the mystery on what biosynthetic mechanisms are responsible for the production of 3-APs, the results of my work necessitate a reconsideration as to how these alkaloids are biosynthesized in contrast to what has been previously hypothesized. Specifically, the origin of these compounds appears to be more complex than a fully reducing PKS which accepts NTA as a starter unit. My efforts identified a single, novel megasynthase gene which possessed PKS and AMP-binding enzymatic domains that fit the PKS hypothesis. Although my attempts to clone and heterologously express this NRPS-PKS gene for functional characterization were unsuccessful, these efforts have laid the foundation for future biochemical work to determine the role of this enigmatic gene. Without functional characterization in the laboratory, it is impossible to completely rule out if this gene has a role in 3-AP biosynthesis. However, based on several previously discussed key qualities, it is my opinion that the *H. indistincta* NRPS-PKS is not responsible for the 3-APs and that alternative origins should be considered. Still, my identification of similar genes in the Demospongiae and Heteroscleromorpha indicates that sponges themselves should not be overlooked as a potential source of nitrogenous marine natural products in comparison to their associated microbiota. In addition, my identification of genes associated with sterol biosynthesis and ALPs from the analysed species has produced alternative avenues for the discovery of new natural products and proteinaceous toxins from Irish *Haliclona* species which may have applied uses. Overall, this endeavour has mainly been one of *in silico* analyses which produced a myriad of computational predictions related to the goal of biodiscovery. While this is a crucial portion of genomics-driven discovery of natural products, it is only one half of the process. To connect genes to natural products said predictions must be tested *in vitro* in the laboratory by means such as heterologous expression or differential transcriptomics. It is my hope that a continuation of my efforts will be pursued in this manner so that possible therapeutic or biotechnological applications may be derived from these challenging but fascinating Irish sponges.

References

1. Fontana, A., 2006. Biogenetic proposals and biosynthetic studies on secondary metabolites of opisthobranch molluscs, in: Cimino, G., Gavagnin, M. (Eds.), Molluscs: from chemo-ecological study to biotechnological application, Progress in molecular and subcellular biology. Springer, Berlin, Heidelberg, pp. 303–332. https://doi.org/10.1007/978-3-540-30880-5_14
2. Itoh, T., Tokunaga, K., Matsuda, Y., Fujii, I., Abe, I., Ebizuka, Y., Kushiro, T., 2010. Reconstitution of a fungal meroterpenoid biosynthesis reveals the involvement of a novel family of terpene cyclases. Nature Chemistry 2, 858–864. https://doi.org/10.1038/nchem.764
3. Ma, H.-M., Zhou, Q., Tang, Y.-M., Zhang, Z., Chen, Y.-S., He, H.-Y., Pan, H.-X., Tang, M.-C., Gao, J.-F., Zhao, S.-Y., Igarashi, Y., Tang, G.-L., 2013. Unconventional origin and hybrid system for construction of pyrrolopyrrole moiety in kosinostatin biosynthesis. Chemistry & Biology 20, 796–805. https://doi.org/10.1016/j.chembiol.2013.04.013
4. Yaegashi, J., Romsdahl, J., Chiang, Y.-M., Wang, C.C.C., 2015. Genome mining and molecular characterization of the biosynthetic gene cluster of a diterpenic meroterpenoid, 15-deoxyoxalicine B, in *Penicillium canescens*. Chemical Science 6, 6537–6544. https://doi.org/10.1039/C5SC01965F
5. Myronovskyi, M., Rosenkränzer, B., Nadmid, S., Pujic, P., Normand, P., Luzhetskyy, A., 2018. Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters. Metabolic Engineering 49, 316–324. https://doi.org/10.1016/j.ymben.2018.09.004
6. Marra, M.V., 2019. Investigation of biological factors that may contribute to bioactivity in *Haliclona* (Porifera, Haplosclerida).
7. Kanehisa, M., 2002. The KEGG Database, in: '*In silico*' simulation of biological processes. John Wiley & Sons, Ltd, pp. 91–103. https://doi.org/10.1002/0470857897.ch8
8. Tribalat, M.-A., 2016. Métabolismes spécialisés d'éponges méditerranéennes du genre *Haliclona* Grant, 1836 205.
9. Gold, D.A., O'Reilly, S.S., Watson, J., Degnan, B.M., Degnan, S.M., Krömer, J.O., Summons, R.E., 2017. Lipidomics of the sea sponge *Amphimedon queenslandica* and implication for biomarker geochemistry. Geobiology 15, 836–843. https://doi.org/10.1111/gbi.12253
10. Thurber, A.R., 2007. Diets of Antarctic sponges: links between the pelagic microbial loop and benthic metazoan food web. Marine Ecology Progress Series 351, 77–89. https://doi.org/10.3354/meps07122
11. Fusetani, N., Asai, N., Matsunaga, S., Honda, K., Yasumuro, K., 1994. Cyclostellettamines A-F, pyridine alkaloids which inhibit binding of methyl quinuclidinyl benzilate (QNB) to muscarinic acetylcholine receptors, from the marine sponge, *Stelletta maxima*. Tetrahedron Letters 35, 3967–3970. https://doi.org/10.1016/S0040-4039(00)76715-3
12. McCord, R.P., Kaplan, N., Giorgetti, L., 2020. Chromosome conformation capture and beyond: toward an integrative view of chromosome structure and function. Molecular Cell 77, 688–708. https://doi.org/10.1016/j.molcel.2019.12.021
13. Trindade-Silva, A.E., Rua, C.P.J., Andrade, B.G.N., Vicente, A.C.P., Silva, G.G.Z., Berlinck, R.G.S., Thompson, F.L., 2013. Polyketide synthase gene diversity within the microbiome of the sponge *Arenosclera brasiliensis*, endemic to the southern Atlantic Ocean. Applied and Environmental Microbiology 79, 1598–1605. https://doi.org/10.1128/AEM.03354-12
14. Cutignano, A., Tramice, A., De Caro, S., Villani, G., Cimino, G., Fontana, A., 2003. Biogenesis of 3-alkylpyridine alkaloids in the marine mollusc *Haminoea orbignyana*. Angewandte Chemie 115, 2737–2740. https://doi.org/10.1002/ange.200250642
15. Cutignano, A., Cimino, G., Giordano, A., d'Ippolito, G., Fontana, A., 2004. Polyketide origin of 3-alkylpyridines in the marine mollusc *Haminoea orbignyana*. Tetrahedron Letters 45, 2627–2629. https://doi.org/10.1016/j.tetlet.2004.01.138

16. Robinson, S.L., Terlouw, B.R., Smith, M.D., Pidot, S.J., Stinear, T.P., Medema, M.H., Wackett, L.P., 2020. Global analysis of adenylate-forming enzymes reveals β-lactone biosynthesis pathway in pathogenic *Nocardia*. Journal of Biological Chemistry 295, 14826–14839. https://doi.org/10.1074/jbc.RA120.013528
17. Stachelhaus, T., Mootz, H.D., Marahiel, M.A., 1999. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chemistry & Biology 6, 493–505. https://doi.org/10.1016/S1074-5521(99)80082-9
18. Kudo, F., Miyanaga, A., Eguchi, T., 2019. Structural basis of the nonribosomal codes for nonproteinogenic amino acid selective adenylation enzymes in the biosynthesis of natural products. Journal of Industrial Microbiology and Biotechnology 46, 515–536. https://doi.org/10.1007/s10295-018-2084-7
19. Tribalat, M.-A., Marra, M.V., McCormack, G.P., Thomas, O.P., 2016. Does the chemical diversity of the order Haplosclerida (phylum Porifera: class Demospongia) fit with current taxonomic classification? Planta Medica 82, 843–856. https://doi.org/10.1055/s-0042-105879
20. Dekimpe, S., Masschelein, J., 2021. Beyond peptide bond formation: the versatile role of condensation domains in natural product biosynthesis. Natural Products Reports 38, 1910–1937. https://doi.org/10.1039/D0NP00098A
21. Halo, L.M., Marshall, J.W., Yakasai, A.A., Song, Z., Butts, C.P., Crump, M.P., Heneghan, M., Bailey, A.M., Simpson, T.J., Lazarus, Colin.M., Cox, R.J., 2008. Authentic heterologous expression of the tenellin iterative polyketide synthase nonribosomal peptide synthetase requires coexpression with an enoyl reductase. ChemBioChem 9, 585–594. https://doi.org/10.1002/cbic.200700390
22. Zou, Y., Yin, H., Kong, D., Deng, Z., Lin, S., 2013. A *trans*-acting ketoreductase in biosynthesis of a symmetric polyketide dimer SIA7248. ChemBioChem 14, 679–683. https://doi.org/10.1002/cbic.201300068
23. Korman, T.P., Crawford, J.M., Labonte, J.W., Newman, A.G., Wong, J., Townsend, C.A., Tsai, S.-C., 2010. Structure and function of an iterative polyketide synthase thioesterase domain catalyzing Claisen cyclization in aflatoxin biosynthesis. Proceedings of the National Academy of Sciences 107, 6246–6251. https://doi.org/10.1073/pnas.0913531107
24. Ziemert, N., Podell, S., Penn, K., Badger, J.H., Allen, E., Jensen, P.R., 2012. The Natural Product Domain Seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. PLOS ONE 7, e34064. https://doi.org/10.1371/journal.pone.0034064
25. Torres, J.P., Lin, Z., Winter, J.M., Krug, P.J., Schmidt, E.W., 2020. Animal biosynthesis of complex polyketides in a photosynthetic partnership. Nature Communications 11, 2882. https://doi.org/10.1038/s41467-020-16376-5
26. Li, F., Lin, Z., Torres, J.P., Hill, E.A., Li, D., Townsend, C.A., Schmidt, E.W., 2022. Sea urchin polyketide synthase SpPks1 produces the naphthalene precursor to echinoderm pigments. The Journal of the American Chemical Society 144, 9363–9371. https://doi.org/10.1021/jacs.2c01416
27. Kakule, T.B., Jadulco, R.C., Koch, M., Janso, J.E., Barrows, L.R., Schmidt, E.W., 2015. Native promoter strategy for high-yielding synthesis and engineering of fungal secondary metabolites. The Journal of American Chemical Society Synthetic Biology 4, 625–633. https://doi.org/10.1021/sb500296p
28. Sabatini, M., Comba, S., Altabe, S., Recio-Balsells, A.I., Labadie, G.R., Takano, E., Gramajo, H., Arabolaza, A., 2018. Biochemical characterization of the minimal domains of an iterative eukaryotic polyketide synthase. The Federation of European Biochemical Societies Journal 285, 4494–4511. https://doi.org/10.1111/febs.14675
29. McQuade, T.J., Shallop, A.D., Sheoran, A., DelProposto, J.E., Tsodikov, O.V., Garneau-Tsodikova, S., 2009. A nonradioactive high-throughput assay for screening and characterization of adenylation domains for nonribosomal peptide combinatorial biosynthesis. Analytical Biochemistry 386, 244–250. https://doi.org/10.1016/j.ab.2008.12.014
30. Feng, L., Gordon, M.T., Liu, Y., Basso, K.B., Butcher, R.A., 2021. Mapping the biosynthetic pathway of a hybrid polyketide-nonribosomal peptide in a metazoan. Nature Communications 12, 4912. https://doi.org/10.1038/s41467-021-24682-9

31. Torres, J.P., Lin, Z., Watkins, M., Salcedo, P.F., Baskin, R.P., Elhabian, S., Safavi-Hemami, H., Taylor, D., Tun, J., Concepcion, G.P., Saguil, N., Yanagihara, A.A., Fang, Y., McArthur, J.R., Tae, H.-S., Finol-Urdaneta, R.K., Özpolat, B.D., Olivera, B.M., Schmidt, E.W., 2021. Small-molecule mimicry hunting strategy in the imperial cone snail, *Conus imperialis*. Science Advances 7, eabf2704. https://doi.org/10.1126/sciadv.abf2704
32. Brunson, J.K., McKinnie, S.M.K., Chekan, J.R., McCrow, J.P., Miles, Z.D., Bertrand, E.M., Bielinski, V.A., Luhavaya, H., Oborník, M., Smith, G.J., Hutchins, D.A., Allen, A.E., Moore, B.S., 2018. Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom. Science 361, 1356–1358. https://doi.org/10.1126/science.aau0382
33. Conkling, M., Hesp, K., Munroe, S., Sandoval, K., Martens, D.E., Sipkema, D., Wijffels, R.H., Pomponi, S.A., 2019. Breakthrough in marine invertebrate cell culture: sponge cells divide rapidly in improved nutrient medium. Scientific Reports 9, 17321. https://doi.org/10.1038/s41598-019-53643-y
34. Kajikawa, M., Hirai, N., Hashimoto, T., 2008. A PIP-family protein is required for biosynthesis of tobacco alkaloids. Plant Molecular Biology 69, 287. https://doi.org/10.1007/s11103-008-9424-3
35. Xu, N.J., Sun, X., Yan, X.J., 2007. A new cyclostellettamine from sponge *Amphimedon compressa*. Chinese Chemical Letters 18, 947–950. https://doi.org/10.1016/j.cclet.2007.06.006
36. O'Brien, R.V., Davis, R.W., Khosla, C., Hillenmeyer, M.E., 2014. Computational identification and analysis of orphan assembly-line polyketide synthases. The Journal of Antibiotics 67, 89–97. https://doi.org/10.1038/ja.2013.125
37. Torres, J.P., Schmidt, E.W., 2019. The biosynthetic diversity of the animal world. Journal of Biological Chemistry 294, 17684–17692. https://doi.org/10.1074/jbc.REV119.006130
38. Brown, M., McShea, H., Olagunju, B., Giner, J., Welander, P., 2021. Testing the sponge biomarker hypothesis through identification of 24-isopropenylcholesterol biosynthesis enzymes. European Association of Geoscientists & Engineers, pp. 1–2. https://doi.org/10.3997/2214-4609.202134237
39. Gerdol, M., Sollitto, M., Pallavicini, A., Castellano, I., 2019. The complex evolutionary history of sulfoxide synthase in ovothiol biosynthesis. Proceedings of the Royal Society B: Biological Sciences 286, 20191812. https://doi.org/10.1098/rspb.2019.1812
40. Steffen, K., Laborde, Q., Gunasekera, S., Payne, C.D., Rosengren, K.J., Riesgo, A., Göransson, U., Cárdenas, P., 2021. Barrettides: a peptide family specifically produced by the deep-sea sponge *Geodia barretti*. Journal of Natural Products 84, 3138–3146. https://doi.org/10.1021/acs.jnatprod.1c00938
41. Piao, S.-J., Song, Y.-L., Jiao, W.-H., Yang, F., Liu, X.-F., Chen, W.-S., Han, B.-N., Lin, H.-W., 2013. Hippolachnin A, a new antifungal polyketide from the South China Sea sponge *Hippospongia lachne*. Organic Letters 15, 3526–3529. https://doi.org/10.1021/ol400933x
42. Teta, R., Irollo, E., Della Sala, G., Pirozzi, G., Mangoni, A., Costantino, V., 2013. Smenamides A and B, chlorinated peptide/polyketide hybrids containing a dolapyrrolidinone unit from the Caribbean sponge *Smenospongia aurea*. Evaluation of Their Role as Leads in Antitumor Drug Research. Marine Drugs 11, 4451–4463. https://doi.org/10.3390/md11114451
43. Shin, H.J., Rashid, M.A., Cartner, L.K., Bokesch, H.R., Wilson, J.A., McMahon, J.B., Gustafson, K.R., 2015. Stellettapeptins A and B, HIV-inhibitory cyclic depsipeptides from the marine sponge *Stelletta* sp. Tetrahedron Letters 56, 4215–4219. https://doi.org/10.1016/j.tetlet.2015.05.058
44. Wilson, M.C., Mori, T., Rückert, C., Uria, A.R., Helf, M.J., Takada, K., Gernert, C., Steffens, U.A.E., Heycke, N., Schmitt, S., Rinke, C., Helfrich, E.J.N., Brachmann, A.O., Gurgui, C., Wakimoto, T., Kracht, M., Crüsemann, M., Hentschel, U., Abe, I., Matsunaga, S., Kalinowski, J., Takeyama, H., Piel, J., 2014. An environmental bacterial taxon with a large and distinct metabolic repertoire. Nature 506, 58–62. https://doi.org/10.1038/nature12959
45. Tianero, M.D., Balaich, J.N., Donia, M.S., 2019. Localized production of defence chemicals by intracellular symbionts of *Haliclona* sponges. Nature Microbiology 4, 1149–1159. https://doi.org/10.1038/s41564-019-0415-8

46. Hamed, A.N.E., Schmitz, R., Bergermann, A., Totzke, F., Kubbutat, M., Müller, W.E.G., Youssef, D.T.A., Bishr, M.M., Kamel, M.S., Edrada-Ebel, R., Wätjen, W., Proksch, P., 2018. Bioactive pyrrole alkaloids isolated from the Red Sea: marine sponge *Stylissa carteri*. Zeitschrift für Naturforschung C 73, 199–210. https://doi.org/10.1515/znc-2017-0161
47. Jackson, D.J., Macis, L., Reitner, J., Wörheide, G., 2011. A horizontal gene transfer supported the evolution of an early metazoan biomineralization strategy. BioMed Central Evolutionary Biology 11, 238. https://doi.org/10.1186/1471-2148-11-238
48. Bugni, T.S., Harper, M.K., McCulloch, M.W.B., Reppart, J., Ireland, C.M., 2008. Fractionated marine invertebrate extract libraries for drug discovery. Molecules 13, 1372–1383. https://doi.org/10.3390/molecules13061372
49. Aiello, A., Fattorusso, E., Menna, M., 1999. Steroids from sponges: recent reports. Steroids 64, 687–714. https://doi.org/10.1016/S0039-128X(99)00032-X
50. Elenkov, I., Dragova, B., Andreev, S., Popov, S., 1997. 4α-Methyl Sterols from the Sponges *Haliclona cinerea* and *Haliclona flavescens*. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 118, 155–157. https://doi.org/10.1016/S0305-0491(97)00029-1
51. Findlay, J.A., Patil, A.D., 1985. Novel sterols from the finger sponge *Haliclona oculata*. Canadian Journal of Chemistry 63, 2406–2410. https://doi.org/10.1139/v85-398
52. Viegelmann, C., Parker, J., Ooi, T., Clements, C., Abbott, G., Young, L., Kennedy, J., Dobson, A.D.W., Edrada-Ebel, R., 2014. Isolation and identification of antitrypanosomal and antimycobacterial active steroids from the sponge *Haliclona simulans*. Marine Drugs 12, 2937–2952. https://doi.org/10.3390/md12052937
53. Kawashima, Y., Nagai, H., Ishida, M., Nagashima, Y., Shiomi, K., 2003. Primary structure of echotoxin 2, an actinoporin-like hemolytic toxin from the salivary gland of the marine gastropod *Monoplex echo*. Toxicon 42, 491–497. https://doi.org/10.1016/S0041-0101(03)00226-5
54. Hoang, Q.T., Cho, S.H., McDaniel, S.F., Ok, S.H., Quatrano, R.S., Shin, J.S., 2009. An actinoporin plays a key role in water stress in the moss *Physcomitrella patens*. New Phytologist 184, 502–510. https://doi.org/10.1111/j.1469-8137.2009.02975.x
55. Takara, T., Nakagawa, T., Isobe, M., Okino, N., Ichinose, S., Omori, A., Ito, M., 2011. Purification, molecular cloning, and application of a novel sphingomyelin-binding protein (clamlysin) from the brackishwater clam, *Corbicula japonica*. Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids 1811, 323–332. https://doi.org/10.1016/j.bbalip.2011.02.004
56. Glasser, E., Rachamim, T., Aharonovich, D., Sher, D., 2014. Hydra actinoporin-like toxin-1, an unusual hemolysin from the nematocyst venom of *Hydra magnipapillata* which belongs to an extended gene family. Toxicon, Special Issue: Freshwater and Marine Toxins 91, 103–113. https://doi.org/10.1016/j.toxicon.2014.04.004
57. Ramírez-Carreto, S., Miranda-Zaragoza, B., Rodríguez-Almazán, C., 2020. Actinoporins: from the structure and function to the generation of biotechnological and therapeutic tools. Biomolecules 10, 539. https://doi.org/10.3390/biom10040539
58. Gerdol, M., Cervelli, M., Oliverio, M., Modica, M.V., 2018. Piercing fishes: porin expansion and adaptation to hematophagy in the vampire snail *Cumia reticulata*. Molecular Biology and Evolution 35, 2654–2668. https://doi.org/10.1093/molbev/msy156
59. Basulto, A., Pérez, V.M., Noa, Y., Varela, C., Otero, A.J., Pico, M.C., 2006. Immunohistochemical targeting of sea anemone cytolysins on tentacles, mesenteric filaments and isolated nematocysts of *Stichodactyla helianthus*. Journal of Experimental Zoology Part A: Comparative Experimental Biology 305A, 253–258. https://doi.org/10.1002/jez.a.256
60. Russell, B., Degnan, B., Garson, M., Skilleter, G., 2003. Distribution of a nematocyst-bearing sponge in relation to potential coral donors. Coral Reefs 22, 11–16. https://doi.org/10.1007/s00338-002-0271-4

61. Schellenberg, J., Reichert, J., Hardt, M., Schmidtberg, H., Kämpfer, P., Glaeser, S.P., Schubert, P., Wilke, T., 2019. The precursor hypothesis of sponge kleptocnidism: development of nematocysts in *Haliclona cnidata* sp. nov. (Porifera, Demospongiae, Haplosclerida). Frontiers in Marine Science 5.
62. Wiens, M., Korzhev, M., Krasko, A., Thakur, N.L., Perović-Ottstadt, S., Breter, H.J., Ushijima, H., Diehl-Seifert, B., Müller, I.M., Müller, W.E.G., 2005. Innate immune defense of the sponge *Suberites domuncula* against bacteria involves a MyD88-dependent signaling pathway: induction of a perforin-like molecule. Journal of Biological Chemistry 280, 27949–27959. https://doi.org/10.1074/jbc.M504049200
63. Alegre-Cebollada, J., Clementi, G., Cunietti, M., Porres, C., Oñaderra, M., Gavilanes, J.G., Pozo, Á.M. del, 2007. Silent mutations at the 5′-end of the cDNA of actinoporins from the sea anemone *Stichodactyla helianthus* allow their heterologous overproduction in *Escherichia coli*. Journal of Biotechnology 127, 211–221. https://doi.org/10.1016/j.jbiotec.2006.07.006
64. Jiang, X., Chen, H., Yang, W., Liu, Y., Liu, W., Wei, J., Tu, H., Xie, X., Wang, L., Xu, A., 2003. Functional expression and characterization of an acidic actinoporin from sea anemone *Sagartia rosea*. Biochemical and Biophysical Research Communications 312, 562–570. https://doi.org/10.1016/j.bbrc.2003.10.159

Appendices

Appendix 1 UniProt sequences used to manually screen *Haliclona* transcriptomes for missing KEGG genes.

Enzyme	**Pathway**	**Species**	**UniProt Accession**
6.3.4.21	Nicotinate Metabolism	*H. sapiens*	Q6XQN6
3.5.1.19	Nicotinate Metabolism	*S. cerevisiae*	P53184
2.3.1.85 - FASN	Fatty Acid Biosynthesis	*H. sapiens*	P49327
2.3.1.86 - FAS1	Fatty Acid Biosynthesis	*S. cerevisiae*	P07149
2.3.1.86 - FAS2	Fatty Acid Biosynthesis	*S. cerevisiae*	P19097
2.3.1.- - Fas	Fatty Acid Biosynthesis	*M. tuberculosis* H37Rv	P95029
2.3.1.41 - FabB	Fatty Acid Biosynthesis	*E. coli* K12	P0A953
1.1.1.100 - FabG	Fatty Acid Biosynthesis	*E. coli* K12	P0AEK2
1.1.1.100 - FabG	Fatty Acid Biosynthesis	*A. thaliana*	P33207
1.1.1.- - CBR4	Fatty Acid Biosynthesis	*H. sapiens*	Q8N4T8
4.2.1.59 - FabA	Fatty Acid Biosynthesis	*E. coli*	P0A6Q3
4.2.1.59 - FabZ	Fatty Acid Biosynthesis	*E. coli* K12	P0A6Q6
4.2.1.59 - FabZ	Fatty Acid Biosynthesis	*A. thaliana*	Q9SIE3
4.2.1.- - HTD2	Fatty Acid Biosynthesis	*H. sapiens*	P86397
1.3.1.9 - FabI	Fatty Acid Biosynthesis	*E. coli* K12	P0AEK4
1.3.1.9 - FabI	Fatty Acid Biosynthesis	*A. thaliana*	Q9SLAB
1.3.1.9 - FabK	Fatty Acid Biosynthesis	*S. pneumoniae*	Q9FBC5
1.3.1.104 - FabL	Fatty Acid Biosynthesis	*B. subtilis* 168	P71079
1.3.1.9 - FabV	Fatty Acid Biosynthesis	*P. aeruginosa*	Q9HZP8
1.14.14.17	Sterol Biosynthesis - Animals	*A. queenslandica*	A0A1X7V870
5.4.99.7	Sterol Biosynthesis - Animals	*A. queenslandica*	A0A1X7VQB7
1.3.1.72 - DHCR24	Sterol Biosynthesis - Animals	*H. sapiens*	Q15392
5.4.99.8	Sterol Biosynthesis - Plants	*A. thaliana*	P38605
2.1.1.41 - SMT1	Sterol Biosynthesis - Plants	*A. queenslandica*	A0A1X7ULF8
1.14.18.10	Sterol Biosynthesis - Plants	*A. thaliana*	Q8L7W5
1.1.1.418	Sterol Biosynthesis - Plants	*A. thaliana*	Q9FX01
1.1.1.270	Sterol Biosynthesis - Plants	*A. queenslandica*	A0A1X7V922
1.14.14.154, 1.14.15.36 - CYP51G1	Sterol Biosynthesis - Plants	*A. thaliana*	Q9SAA9
1.3.1.70 - FK	Sterol Biosynthesis - Plants	*A. thaliana*	Q9LDR4
5.3.3.5 - HYD1	Sterol Biosynthesis - Plants	*A. thaliana*	O48962
2.1.1.143 - SMT2	Sterol Biosynthesis - Plants	*A. thaliana*	Q39227
1.14.18.11	Sterol Biosynthesis - Plants	*A. thaliana*	Q8VWZ8
1.14.19.20 - STE1	Sterol Biosynthesis - Plants	*A. thaliana*	Q39208
1.3.1.21 - DWF5	Sterol Biosynthesis - Plants	*A. thaliana*	Q9LDU6
1.3.1.72 - DWF1	Sterol Biosynthesis - Plants	*A. thaliana*	Q39085
1.14.19.41 - CYP710A	Sterol Biosynthesis - Plants	*A. thaliana*	Q9ZV29

Appendix 2 Hybrid ketosynthase sequences used for the construction of a phylogenetic tree.

Fungi	
Sequence	**Accession**
Cetraspora pellucida 1	CAG8479329.1
Dentiscutata erythropus 1	CAG8465812.1
Dentiscutata erythropus 2	CAG8557733.1
Dentiscutata erythropus 3	CAG8562942.1
Dentiscutata erythropus 4	CAG8743934.1
Dentiscutata heterogama 1	CAG8528616.1
Dentiscutata heterogama 1	CAG8580022.1
Gigaspora margarita 1	CAG8484623.1
Gigaspora margarita 2	CAG8718178.1
Gigaspora margarita 3	CAG8723241.1
Gigaspora margarita 4	CAG8732995.1
Gigaspora margarita 5	KAF0367615.1
Gigaspora margarita 6	KAF0480591.1
Gigaspora margarita 7	KAF0497156.1
Gigaspora margarita 8	KAF0524901.1
Gigaspora rosea 1	CAG8454249.1
Gigaspora rosea 2	RIB03237.1
Gigaspora rosea 3	RIB15045.1
Gigaspora rosea 4	RIB28562.1
Racocetra fulgida 1	CAG8462630.1
Racocetra persica 1	CAG8560316.1

Rotifers	
Species	**Accession**
Didymodactylos carnosus 1	CAF0936322.1
Adineta ricciae 1	CAF1215328.1
Rotaria sp. Silwood1 1	CAF1288588.1
Rotaria sp. Silwood2 1	CAF2828736.1
Rotaria socialis 1	CAF3353722.1
Rotaria sordida 1	CAF3621424.1
Rotaria magnacalcarata 1	CAF3978175.1

Nematodes	
Species	**Accession**
Haemonchus contortus 1	CDJ83277.1
Ancylostoma ceylanicum 1	EPB66696.1

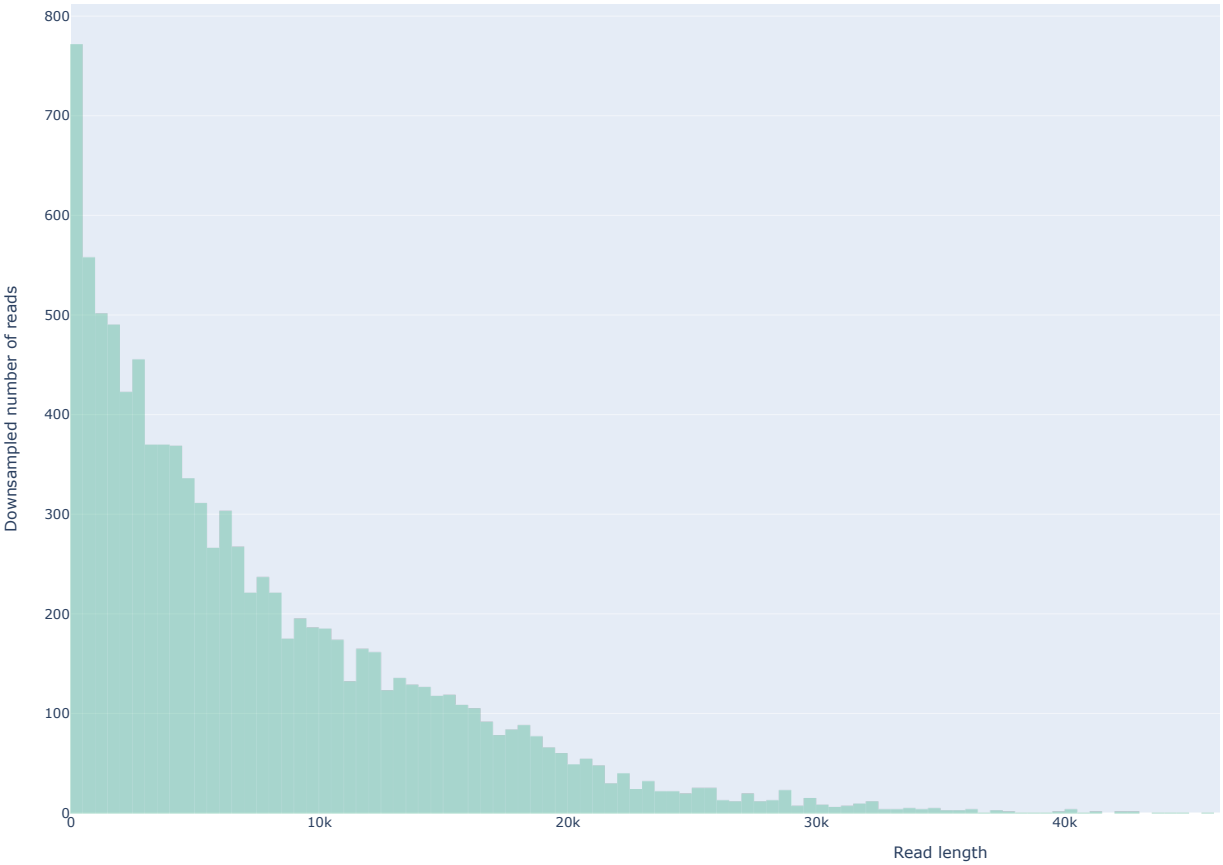
Bacteria	
Sequence	**Accession**
Nostoc sp. GSV224 nosB	AAF15892.2
Lyngbya majuscula 1	AAS98784.1
Lyngbya majuscula CurG	AAT70102.1
Cylindrospermopsis raciborskii AWT205 cyrB	ABX60161.1
Nostoc sp. 'Peltigera membranacea cyanobiont' 1	ADA69241.1
Nostoc sp. CENA543 ndaC	ATP76242.1
Okeania hirsuta MgcK	AZH23793.1
Candidatus Entotheonella sp. kasB	BAW32323.1
Streptomyces avermitilis ptxC	BBA21069.1
Kitasatospora griseola 1	KIQ64709.1
Unidentified bacterium 1	MCP4653886.1
Unidentified bacterium 2	MCP4653886.1
Streptomyces halstedii 1	NEA20538.1
Acidobacteriia bacterium AA117 1	PYP93107.1
Chloroflexi bacterium 1	RMF30773.1
Salinispora arenicola CNR107 1	WP_019032754.1
Streptomyces kasugaensis 1	WP_131122990.1
Cystobacter gracilis 1	WP_224245440.1
Streptomyces sp. MA3 2.13 1	WP_228078005.1
Streptomyces sp. RY43-2 1	WP_252427324.1

Algae	
Sequence	**Accession**
Symbiodinium sp. CCMP2592 1	CAE7473882.1
Symbiodinium sp. CCMP2456 1	CAE7507918.1
Polarella glacialis 1	CAE8667626.1
Diacronema lutheri 1	KAG8458283.1
Chrysochromulina tobinii 1	KOO23785.1
Gambierdiscus polynesiensis 1	QJU71784.1

Sponges	
Species	**Accession**
Raspailiidae sp. 1	AAX62314.1

Angiostrongylus cantonensis 1	KAE9414965.1
Oesophagostomum dentatum 1	KHJ99846.1
Ancylostoma duodenale 1	KIH69030.1
Ancylostoma caninum 1	RCN45771.1
Nippostrongylus brasiliensis 1	VDL74159.1
Heligmosomoides polygyrus 1	VDP18007.1

Appendix 3 Histogram of read length from Pacific Biosciences Sequel ii single-molecule real-time sequencing of MIIG1388.



Appendix 4 Primers used to attempt amplification of regions of the *H. indistincta* NRPS-PKS gene.

Primer	**Sequence**	**Fragment Size (bp)**
Fragment 1 Forward	ATGTCGTCAGAATCTCAATC	1843
Fragment 1 Reverse	GGCTACTGAATTCATCAATG	
Fragment 2 Forward	ATTCAGTAGCCATACATACC	1377
Fragment 2 Reverse	ACTACAGGTACTATAAGAGC	
Fragment 3 Forward	ATTCCAGCTCTTATAGTACC	2615
Fragment 3 Reverse	GTGTCTTCATCACTAACATC	
Fragment 4 Forward	TGTTAGTGATGAAGACACTG	2189
Fragment 4 Reverse	AACTGAGCTCTAATACACTC	
Fragment 5 Forward	AGTGTATTAGAGCTCAGTTC	2075
Fragment 5 Reverse	GATACAGAGAGCATACAACC	
Fragment 6 Forward	GTATGCTCTCTGTATCATTG	1859
Fragment 6 Reverse	CATACATGTGAATTCTCCAC	

Appendix 5 Presence and absence of ALP genes in the assembled sponge genomes and transcriptomes.

Order	Family	Genus	Species	# Non-redundant ALPs
Amphidiscosida	Hyalonematidae	*Hyalonema*	*populiferum*	0
Axinellida	Axinellidae	*Cymbastela*	*concentrica*	0
Axinellida	Axinellidae	*Cymbastela*	*stipitata*	0
Axinellida	Axinellida	*Stylissa*	*carteri*	3
Baerida	Baeriidae	*Leuconia*	*nivea*	0
Bubarida	Dictyonellidae	*Scopalina*	*sp.*	3
Chondrillida	Chondrillidae	*Chondrilla*	*nucula*	0
Clathrinida	Clathrinidae	*Clathrina*	*coriacea*	0
Clionaida	Clionaidae	*Cliona*	*orientalis*	3
Clionaida	Clionaidae	*Cliona*	*varians*	7
Dendroceratida	Darwinellidae	*Dendrilla*	*antarctica*	0
Dendroceratida	Halisarcidae	*Halisarca*	*caerulea*	0
Dendroceratida	Halisarcidae	*Halisarca*	*dujardini*	0
Dendroceratida	Darwinellidae	*Pleraplysilla*	*spinifera*	1
Dictyoceratida	Thorectidae	*Carteriospongia*	*foliascens*	0
Dictyoceratida	Dysideidae	*Dysidea*	*avara*	9
Dictyoceratida	Irciniidae	*Ircinia*	*fasciculata*	0
Dictyoceratida	Spongiidae	*Spongia*	*officinalis*	2
Haplosclerida	Niphatidae	*Amphimedon*	*queenslandica*	0
Haplosclerida	Chalinidae	*Haliclona*	*amboinensis*	1
Haplosclerida	Chalinidae	*Haliclona*	*cinerea*	1
Haplosclerida	Chalinidae	*Haliclona*	*indistincta*	8
Haplosclerida	Chalinidae	*Haliclona*	*oculata*	0
Haplosclerida	Chalinidae	*Haliclona*	*simulans*	0
Haplosclerida	Chalinidae	*Haliclona*	*tubifera*	0
Haplosclerida	Chalinidae	*Haliclona*	*viscosa*	10
Haplosclerida	Petrosiidae	*Petrosia*	*ficiformis*	0
Haplosclerida	Petrosiidae	*Xestospongia*	*testudinaria*	0
Hexactinosida	Sceptrulophora	*Aphrocallistes*	*vastus*	0
Homosclerophorida	Plakinidae	*Corticium*	*candelabrum*	1
Homosclerophorida	Oscarellidae	*Oscarella*	*carmela*	0
Homosclerophorida	Oscarellidae	*Oscarella*	*pearsei*	3
Homosclerophorida	Plakinidae	*Plakina*	*jani*	2
Leucosolenida	Grantiidae	*Grantia*	*compressa*	0
Leucosolenida	Leucosoleniidae	*Leucosolenia*	*complicata*	1
Leucosolenida	Sycettidae	*Sycon*	*coactum*	0
Leucosolenida	Sycettidae	*Sycon*	*ciliatum*	0
Lyssacinosida	Rossellidae	*Rossella*	*fibulata*	0
Lyssacinosida	Rossellidae	*Sympagella*	*nux*	0
Poecilosclerida	Crellidae	*Crella*	*elegans*	0
Poecilosclerida	Isodictyidae	*Isodictya*	*sp.*	0
Poecilosclerida	Hymedesmiidae	*Kirkpatrickia*	*variolosa*	0
Poecilosclerida	Latrunculiidae	*Latrunculia*	*apicalis*	0

Poecilosclerida	Mycalidae	*Mycale*	*phyllophila*	3
Poecilosclerida	Tedaniidae	*Tedania*	*anhelans*	1
Spongillida	Lubomirskiidae	*Baikalospongia*	*bacillifera*	0
Spongillida	Spongillidae	*Ephydatia*	*fluviatilis*	0
Spongillida	Spongillidae	*Ephydatia*	*muelleri*	0
Spongillida	Spongillidae	*Ephydatia*	*muelleri*	0
Spongillida	Spongillidae	*Eunapius*	*fragilis*	0
Spongillida	Lubomirskiidae	*Lubomirskia*	*abietina*	0
Spongillida	Lubomirskiidae	*Lubomirskia*	*baikalensis*	0
Spongillida	Spongillidae	*Spongilla*	*lacustris*	0
Suberitida	Halichondriidae	*Halichondria*	*panicea*	3
Suberitida	Suberitidae	*Pseudospongosorites*	*suberitoides*	0
Suberitida	Suberitidae	*Suberites*	*domuncula*	0
Tethyida	Tethyidae	*Tethya*	*wilhelma*	3
Tetractinellida	Astrophorina	*Geodia*	*atlantica*	0
Tetractinellida	Astrophorina	*Geodia*	*barretti*	1
Tetractinellida	Astrophorina	*Geodia*	*hentscheli*	0
Tetractinellida	Astrophorina	*Geodia*	*macandrewii*	0
Tetractinellida	Astrophorina	*Geodia*	*phlegraei*	0
Verongiida	Aplysinidae	*Aplysina*	*aerophoba*	0

Publications

marine drugs

MDPI

Article

Actinoporin-like Proteins Are Widely Distributed in the Phylum Porifera

Kenneth Sandoval and Grace P. McCormack *

Molecular Evolution and Systematics Laboratory, Zoology, Ryan Institute & School of Natural Sciences, National University of Ireland Galway, 23 University Rd., H91 R8EC Galway, Ireland; k.sandoval1@nuigalway.ie
* Correspondence: grace.mccormack@nuigalway.ie

Abstract: Actinoporins are proteinaceous toxins known for their ability to bind to and create pores in cellular membranes. This quality has generated interest in their potential use as new tools, such as therapeutic immunotoxins. Isolated historically from sea anemones, genes encoding for similar actinoporin-like proteins have since been found in a small number of other animal phyla. Sequencing and *de novo* assembly of Irish *Haliclona* transcriptomes indicated that sponges also possess similar genes. An exhaustive analysis of publicly available sequencing data from other sponges showed that this is a potentially widespread feature of the Porifera. While many sponge proteins possess a sequence similarity of 27.70–59.06% to actinoporins, they show consistency in predicted structure. One gene copy from *H. indistincta* has significant sequence similarity to sea anemone actinoporins and possesses conserved residues associated with the fundamental roles of sphingomyelin recognition, membrane attachment, oligomerization, and pore formation, indicating that it may be an actinoporin. Phylogenetic analyses indicate frequent gene duplication, no distinct clade for sponge-derived proteins, and a stronger signal towards actinoporins than similar proteins from other phyla. Overall, this study provides evidence that a diverse array of Porifera represents a novel source of actinoporin-like proteins which may have biotechnological and pharmaceutical applications.

Keywords: Porifera; marine sponge; *Haliclona*; transcriptomics; actinoporins; pore-forming toxins

Citation: Sandoval, K.; McCormack, G.P. Actinoporin-like Proteins Are Widely Distributed in the Phylum Porifera. *Mar. Drugs* **2022**, *20*, 74. https://doi.org/10.3390/md20010074

Academic Editors: Micha Ilan, Shmuel Carmeli, Michelle Kelly and Mark T. Hamann

Received: 15 December 2021
Accepted: 10 January 2022
Published: 15 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Actinoporins (APs) are proteinaceous α-pore-forming toxins originally isolated from and named after sea anemones [1]. This group of toxins typically exhibit several common characteristics, such as a common absence of cysteine residues, a high isoelectric point (>8.8), and a small size (~20 kDa) [2]. Furthermore, they comprise a compact β-sandwich flanked on each side by an α-helix, as indicated by the crystal structures of the well-studied equinatoxin II (EqT-II), stichyolysin II (Stn-II), and fragaceatoxin C (Fra-C) [3–5]. The molecular mechanism of cytolytic pore formation by APs has been extensively researched and appears to involve several steps, which are briefly summarized. First, lipid recognition and membrane binding are accomplished via the interfacial binding site (IBS), which features a cluster of prominent aromatic residues that bind to phosphocholine (POC) [4,6]. In particular, APs have an affinity for the POC group of sphingomyelin (SM) and are capable of discriminating between this target and other membrane lipids, such as phosphatidylcholine [6,7]. After binding to a target membrane, APs then undergo a conformational change in which the N-terminal region, containing one of the α-helices, is translocated to lie flat upon the membrane surface [8,9]. This N-terminal region is then inserted into the target membrane and undergoes further conformational change to increase the overall length of the amphipathic α-helix relative to its unbound state [10,11]. The pore is finally formed when oligomerization occurs via the recruitment of additional AP monomers, which undergo the same process in the same region of the membrane to bring about the death of targeted cells by osmotic shock [12,13]. For a more in-depth explanation on the

molecular mechanisms of pore formation by APs, the reader is referred to reviews which focus on this topic [13–15]. The qualities of APs which allow for their membrane-binding and pore-forming activity have attracted attention regarding potential biotechnological and therapeutic applications, such as the design of immunotoxins, nanopores, adjuvants, and SM-specific probes [16–19].

Historically, sea anemones have been the primary source of APs, although similar cytolytic proteins can be found in other anthozoans [20,21]. Indeed, an exhaustive bioinformatic analysis indicated that actinoporin-like proteins (ALPs) are distributed across multiple phyla with high structural similarity despite low sequence similarity [22]. In particular, APs and ALPs have been detected in chordates (primarily teleost fish), cnidarians, molluscs, mosses, and ferns. Furthermore, a structural similarity of APs and ALPs to fungal-fruit body lectins has also been determined. A phylogenetic analysis of these identified proteins revealed four distinct groups comprising ALPs primarily found in vertebrates, hydrozoan ALPs, APs from cnidarians and plants, and fungal fruit-body lectins, all of which were proposed to comprise the actinoporin-like proteins and fungal fruit-body lectins superfamily (AF). The presence of ALP genes in non-vertebrate bilaterians has been further illuminated in studies focused on polychaetes of the genus *Glycera*, the crustacean *Xibalbanus tulumensis*, the brachiopod *Lingula anatina*, and many molluscs of the classes Gastropoda and Bivalvia [23–25]. Several ALPs have been functionally characterized, indicating that they can possess similar qualities to APs regarding membrane binding and cytolytic activity.

The first published example of an ALP was echotoxin II, isolated from the salivary glands of the predatory mollusc *Monoplex echo*, which also has an amphipathic N-terminal α-helix, a patch of aromatic residues, and hemolytic activity, but a specificity to gangliosides rather than SM [26]. In further contrast to APs, an ALP from the zebrafish, *Danio rerio*, possessed no cytolytic activity and its membrane-binding activity was not specific to SM [22]. Yet bryoporin, from the moss *Physcomitrella patens*, showed consistency with its close phylogenetic grouping to APs in that it also exhibited specificity for SM as well as hemolytic activity, although its biological role appears to be related to dehydration stress [27]. Similarly, clamlysin B from the bivalve *Corbicula japonica* also exhibits SM-binding and cytolytic activity [28]. Finally, the ALP HALT-1 from the cnidarian *Hydra magnipapillata*, which is phylogenetically distinct from anthozoan APs, exhibits lower hemolytic activity, the creation of larger pores, and a lower affinity to SM in comparison to EqT-II [29]. Altogether, the observation that ALPs from non-anthozoans can possess similar biochemical properties to APs supports the notion that they may also be potential targets for the aforementioned biotechnological and therapeutic applications which have been investigated for EqT-II, Stn-II, and Fra-C.

Sea sponges of the phylum Porifera are benthic, filter-feeding animals which can be found in marine and freshwater environments throughout the world. Given the niche they fill, challenges faced by these organisms include contact with pathogenic microorganisms, spatial competition with other benthic life, and predation [30–32]. In order to deal with these challenges, many sponges utilize complex chemical armaments which display an array of bioactivities towards targets, such as pathogens, fouling organisms and cancerous cells [33–35]. While a majority of these bioactivities have been attributed to small molecules, some larger proteinaceous toxins have also been identified, such as suberitine from *Suberities domuncula*, halilectin-3 from *Haliclona caerulea*, and chondrosin from *Chondrosia reniformis* [36–38]. In addition, sea sponges have also been shown to be a source of cytolytic pore-forming proteins, one of which is an antibacterial, perforin-like protein from *S. domuncula*, which was found to be upregulated upon exposure to lipopolysaccharide [39,40]. No ALPs have been isolated and characterized from this phylum, but a recent phylogenetic study has indicated that genes encoding for these proteins are present in the genome of the species *Oscarella pearsei* (then *O. carmela*, when its genome was sequenced) [21,41,42]. Little was reported on this ALP other than it being phylogenetically distant from both anthozoan APs, hydrozoan ALPs and mollusc ALPs. Similarly, while

analyzing our transcriptome assemblies of native Irish *Haliclona* species, we noticed the presence of numerous genes encoding for proteins with the Pfam domain PF06369, representing sea anemone cytotoxic proteins. The presence of ALPs in sponges of both the classes Demospongiae and Homoscleromorpha prompted the questions of whether these proteins are widely distributed throughout the phylum and how similar they are to known APs. As discussed previously, such a quality would expand the possible biotechnological and therapeutic applications of sponges. To date, these organisms have been the subject of numerous transcriptomic and genomic studies, resulting in a wealth of public data with which to carry out such an inquiry [43]. To address these aforementioned questions and expand the knowledge of APs and ALPs, we herein present an exploration of the diversity, distribution, and predictive function of these proteins in the phylum Porifera.

2. Results

2.1. Transcriptome Sequencing and ALP Identification

After processing with fastp, approximately 472.68, 69.80, 257.97, 207.46, and 94.64 Mbp of data were acquired for *H. cinerea*, *H. indistincta*, *H. oculata*, *H. simulans*, and *H. viscosa*, respectively. Five separate transcriptomes were then assembled using the Trinity RNA-Seq assembler (Table 1). More data were available for *H. cinerea*, *H. oculata*, and *H. simulans*, which appeared to be reflected in the generally larger assembly size and higher number of true genes when compared to *H. indistincta* and *H. viscosa* (the latter two species belong to a separate species group that is phylogenetically distinct from the first three). Furthermore, this division was also apparent regarding GC content, in which the first three exhibited a value around 39%, while the latter exhibited a value of 44.5%. The total amount of translated open reading frames was largely reflected by the size of the assemblies, with *H. cinerea* yielding the most protein sequences, while *H. viscosa* yielded the least (Table 2). All five *Haliclona* transcriptomes exhibited very high completeness when assessed with the BUSCO eukaryote dataset (Table 3).

Table 1. Trinity assembly statistics for the five *Haliclona* transcriptomes.

Species	Total (Mbp)	Number of Contigs	Number of Trinity 'Genes' Excluding Isoforms	Contig N50 (Kbp)	GC (%)
H. cinerea	156.12	123,111	64,261	2.81	39.59
H. indistincta	106.87	101,413	48,788	2.03	44.10
H. oculata	142.18	122,855	70,008	2.46	38.56
H. simulans	104.12	106,366	55,501	1.89	39.87
H. viscosa	94.09	105,831	59,949	1.73	44.50

Table 2. TransDecoder open reading frame statistics for the five *Haliclona* transcriptomes.

Species	Total (aa)	Number of Complete ORFs	Number of 5′ Partial ORFs	Number of 3′ Partial ORFs
H. cinerea	34,243,765	58,606	14,476	6129
H. indistincta	26,553,534	29,232	15,794	7290
H. oculata	30,530,592	49,528	16,191	5687
H. simulans	24,643,377	31,081	16,112	7561
H. viscosa	22,299,218	24,435	12,776	7166

Table 3. BUSCO eukaryotic score for the five *Haliclona* transcriptomes.

Species	Complete (%)	Single (%)	Duplicate (%)	Fragmented (%)	Missing (%)
H. cinerea	97.6	23.9	73.7	0.8	1.6
H. indistincta	99.6	45.1	54.5	0.0	0.4
H. oculata	98.5	31.4	67.1	0.4	1.1
H. simulans	97.6	38.0	59.6	2.0	0.4
H. viscosa	95.3	55.3	40.0	4.3	0.4

A total of 66 unique open reading frames encoding for proteins with the pfam domain PF06369 were identified in the analyzed sponge NGS resources. These ORFs were derived from 20 of the total 63 species that were screened (Table S1) Of these, 38 ORFs were determined to be complete by TransDecoder. As none of these sponge-derived proteins have been experimentally characterized, they are hence referred to as ALPs. Specifically, these complete ALPs were found in *C. orientalis* (Co), *C. varians* (Cv), *D. avara* (Da), *G. barretti* (Gb), *H. indistincta* (Hi), *H. viscosa* (Hv), *O. pearsei* (Op), a *Scolapina* sp. (S), *S. officinalis* (So), and *T. wilhelma* (Tw). ALPs were absent from all freshwater sponge data. Their presence in an order did not necessarily translate to this being an absolute feature of said order. This is particularly exemplified by the order Haplosclerida, in which numerous paralogs were detected in the sister species *H. indistincta* and *H. viscosa*, a few in *H. amboinensis* and *H. cinerea*, and none in *H. oculata*, *H. simulans*, *H. tubifera*, or *A. queenslandica*. Only *C. varians* and *D. avara* possessed the same degree of parology as *H. indistincta* and *H. viscosa*. Of the sponge ALPs, only two from *T. wilhelma* were identified as having a signal peptide by SignalP [44]. All the complete sponge ALPs aligned most closely with cnidarian actinoporins when these were used in a blastp query against the NCBI non-redundant database (Table S2). In particular, one ALP from *H. indistincta* named Hi2 exhibited the highest sequence similarity with an actinoporin from *Haloclava producta* at 59.06% identity. It was also observed that several ALPs from incomplete ORFs most closely aligned with pore-forming proteins from other phyla, such as coluporins and tereporins from the Mollusca. No ALPs were detected in the screened genomes or transcriptomes of choanoflagellates, ctenophores, or placozoans. The theoretical isoelectric point of complete sponge ALPs ranged from 4.66 to 9.46, whereas the average molecular weight ranged from 14,153.27 to 33,419.71 Da (Table S3).

2.2. Structural Prediction

In contrast to the low-to-modest sequence similarity many sponge ALPs showed in relation to actiniarian APs (27.70–59.06%) (Table S2), homology modeling with Phyre2 indicated that the aligned predicted structure of all complete sponge ALPs was highly similar to that of Stn-II and EqT-II, with confidence values ranging from 97.3% to 100%. Quantification of the similarity between the predicted sponge ALP models and the crystal structure of EqT-II via TM-align showed that all produced models had a TM-align value above 0.5, indicating that the structural similarity was not random (Figure 1), whereas the RMSD values ranged from 0.54 to 1.74 (Å) (Table S4).

It must be noted, however, that several sea sponge ALPs exhibited structural inconsistencies with the expected AP skeleton, such as the lack of an N-terminal α-helix. Hi2 exhibited one of the highest TM-scores at 0.96406. The quality of the predicted model for Hi2 was further supported with a ProQ3D S-score of 0.697 (0.5–1.0 representing a good model; Arne Elofsson, personal communication) and a ModFOLD8 *p*-value of 3.772×10^{-6} (less than a 1/1000 chance that the model is incorrect). As can be seen by the structure generated by Phyre2, Hi2 shares many characteristics typical of cnidarian actinoporins, such as comprising a β-sandwich flanked by two α-helices (Figure 2a,d). Furthermore, the localization of the interfacial binding site can also be observed (Figure 2b). In general, the predicted structure of Hi2 appears to overlay well with the crystal structure of EqT-II chain A (1IAZ) (Figure 2c).

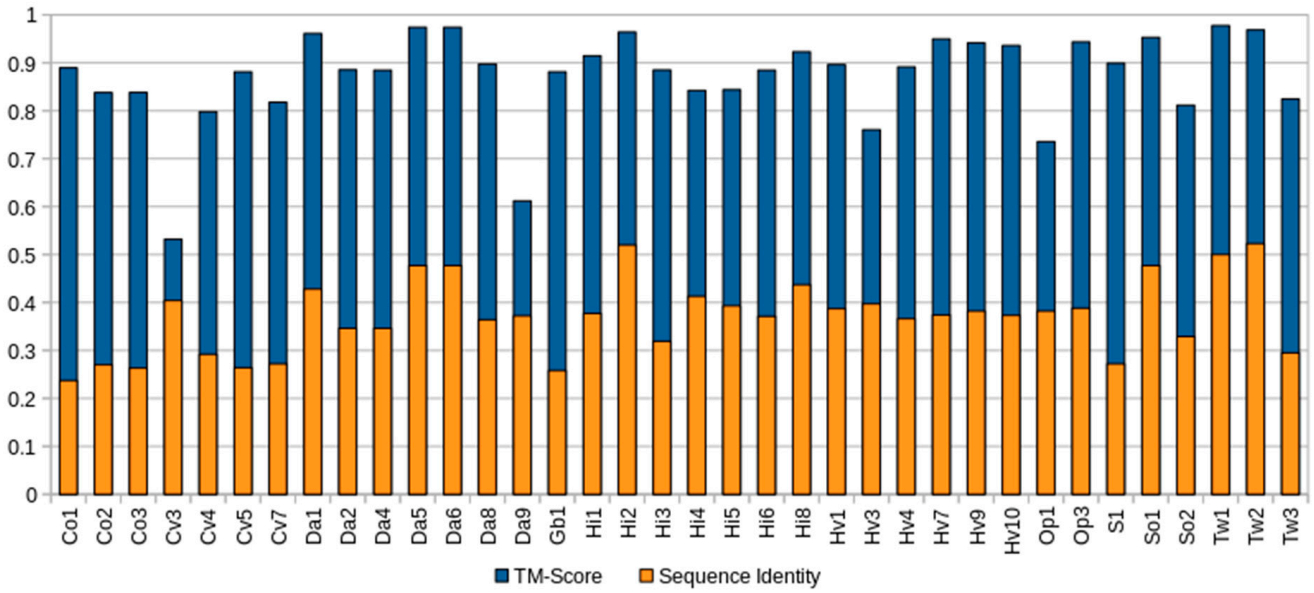


Figure 1. Comparison of complete sponge ALPs with EqT-II. Blue represents the TM-score of all predicted sponge ALP structures via Phyre2 with the crystal structure of EqT-II (1IAZ). Orange represents the sequence identity of the aligned region of the two structures. The orange bar representing percent identity is overlaid upon the blue bar representing TM-score. Abbreviations are as follows: Co, *C. orientalis*; Cv, *C. varians*; Da, *Dysidea avara*; Hi, *H. indistincta*; Hv, *H. viscosa*; Op, *O. pearsei*; *S, Scopalina* sp.; So, *S. officinalis*; Tw, *Tethya wilhelma*. The protein Sc1 from *S. carteri* is excluded due to having unknown residues. Sequence identity is expressed on a scale of 0–1 rather than as a percentage.

2.3. Multiple Sequence Alignment and Residue Analysis

Due to its high sequence similarity with cnidarian actinoporins, Hi2 was chosen for further in-depth analyses to determine whether it exhibited the same membrane-binding and pore-forming activities. A multiple sequence alignment of Hi2 with the final mature peptides of the well-studied equinatoxin II (EqT-II; P61914), fragaceatoxin C (Fra-C; B9W5G6), and stichyolysin II (Stn-II; P07845) indicated a percent identity of 50.56%, 50.00%, and 51.41%, respectively (Figure 3). Despite these modest values, the alignment illustrated a high degree of conservation regarding residues and motifs critical for the functional activities of actinoporins [15]. For example, a majority of the residues associated with the interfacial binding site in Hi2 are consistent with those of EqT-II, Fra-C, and Stn-II [6]. In addition, Hi2 possesses the conserved residue Tyr112, which is critical for SM recognition; however, a substitution of Leu for Trp at residue 111 is also observed. Furthermore, the presence of Ser53, Val86, Ser104, Pro106, Trp115, Tyre132, Tyr136, and Tyr137 are consistent with the POC binding site found in cnidarian APs [4,6,15], and the conserved P-[WYF]-D binding motif found in this region of APs is also present in Hi2 at residues 106–108 [22]. Oligomerization of actinoporin monomers upon the cell membrane is another crucial step towards pore formation and is known to be influenced by an Arg–Gly–Asp motif. Hi2 shows inconsistency with this motif, as it instead possesses Lys142, Gly143, and Glu144. However, Hi2 possesses the residue Lys76, which is consistent with similar residues associated with oligomerization in other APs [45]. The presence of Ile59 and Trp147 are also partially consistent with residues of Fra-C, associated with oligomerization and protein–protein interaction between protomers; the observed substitution of Ile for Val at this site can be seen in Stn-II [13,46]. Unlike most cnidarian APs, Hi2 exhibits the presence of cysteine at residue 141, but such a characteristic is not unheard of in these proteins [21].

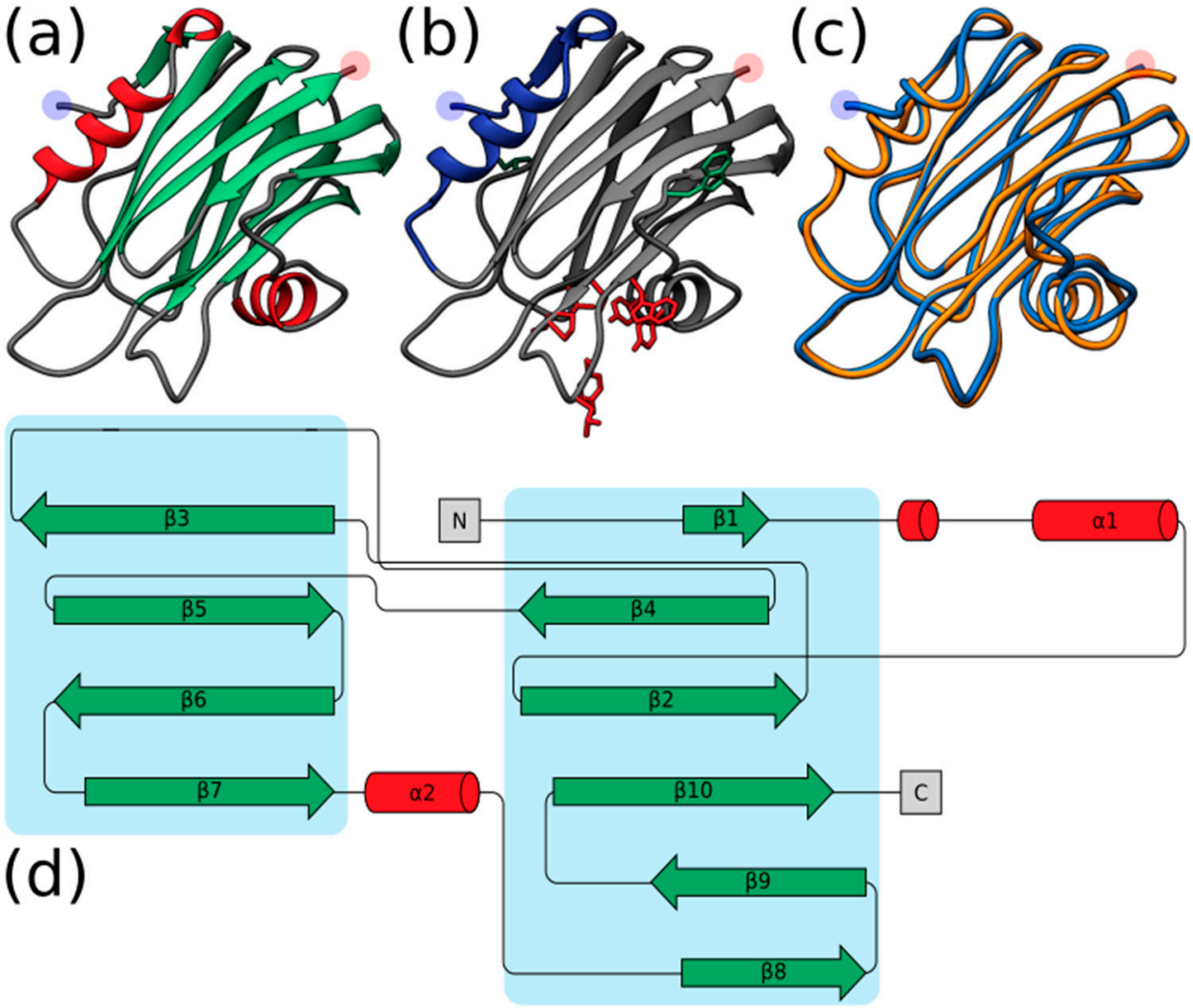


Figure 2. (**a**) Predicted structure of Hi2 by Phyre2. Red represents α-helices. Green represents β-sheets. (**b**) Significant functional residues of Hi2. Blue represents the N-terminal α-helix associated with pore formation. Red represents the residues of the interfacial binding site. Green represents the residues associated with oligomerization. (**c**) Structural alignment of Hi2 in blue upon EqT-II chain A (1IAZ) in orange. For all predicted structures blue and red highlights represent the N- and C-terminus, respectively. (**d**) Protein topology plot of Hi2 with the same color scheme as (**a**). The size of the topology plot was manually reduced to be more compact.

Membrane penetration and pore formation by oligomerized APs is achieved by their respective amphipathic N-terminal α-helices. In EqT-II, the N-terminal region undergoes a conformational change, producing an α-helix comprising residues 6–28 which is capable of spanning a target membrane [10]. These residues correspond to a conserved N-terminal glycine and C-terminal asparagine of the α-helix, which are also present in Fra-C and Hi2. Within this region, Hi2 also showed consistency with several previously determined highly conserved hydrophobic residues (Val7, Ile8, Leu13, Leu18, Leu22, and Ile25), as well as Arg30, which is associated with the insertion of the α-helix into the target membrane [20]. Using these sequence boundaries allowed for the construction of an Edmundson peptide helical wheel of the predicted Hi2 N-terminal α-helix after a hypothetical conformational change (Figure 4a). Consistent with the amphipathic nature of the N-terminal α-helix of EqT-II, Fra-C, and Stn-II, a side comprising a majority of polar amino acids opposite another comprising a majority of nonpolar amino acids can be seen in that of Hi2. Furthermore, the hydrophobic moment of Hi2, a measure of helix amphipathicity, was calculated to be

0.384 µH, which is comparable to that of EqT-II at 0.337 µH. The N-terminal α-helix of Hi2 exhibits hydrophobicity of 0.607 and a net charge of 0. The two faces of the N-terminal α-helix prior to a hypothetical conformational change also display a hydrophobic and hydrophilic side, which are, respectively, oriented towards and away from the rest of the protein (Figure 4b).

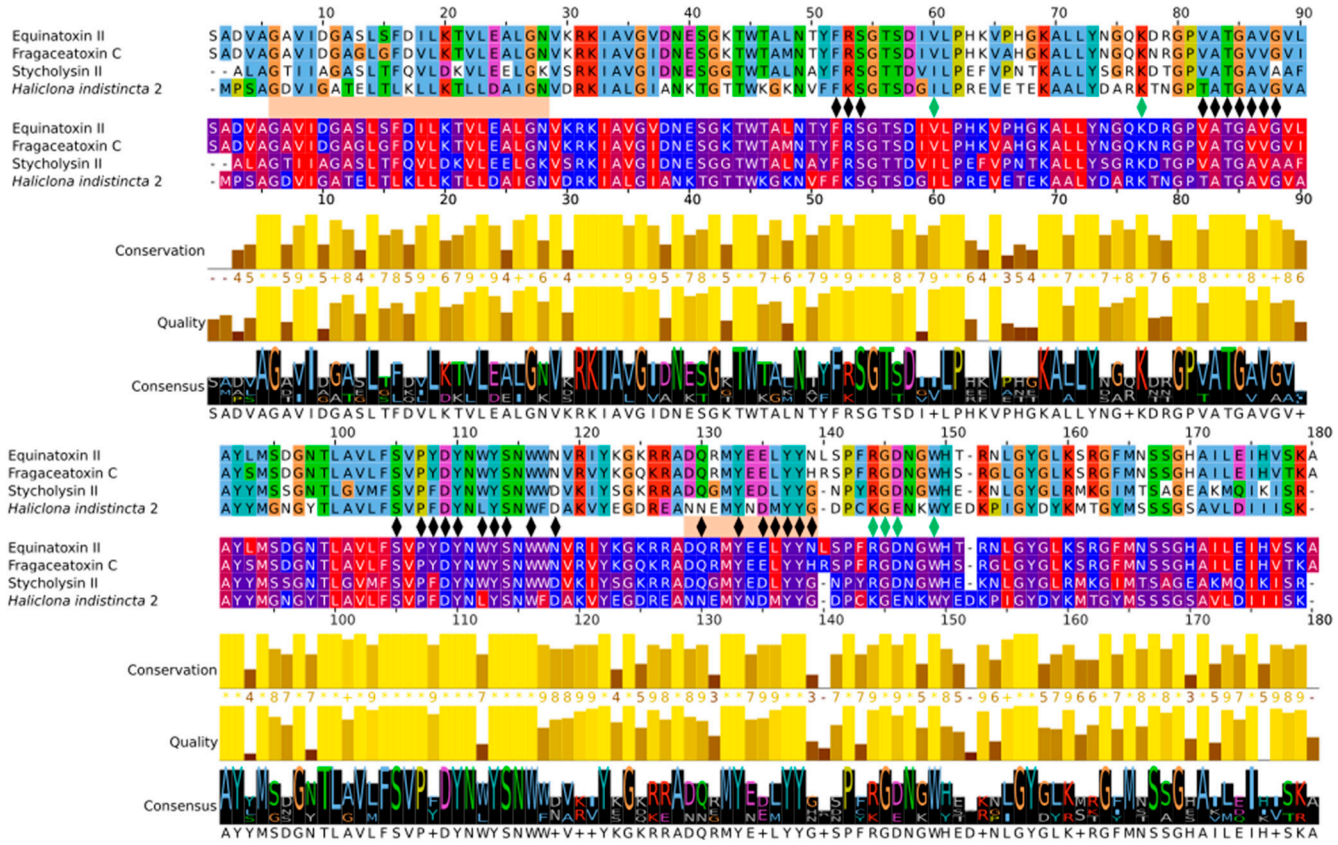


Figure 3. Multiple sequence alignment of Hi2 with EqT-II, Fra-C, and Stn-II. The top alignment represents Clustalx color coding. The bottom alignment represents hydrophobicity color coding. Between these two alignments, black markers represent residues important for membrane binding, green markers represent residues important for oligomerization, and orange rectangles represent the α-helices of EqT-II (the final length of the N-terminal α-helix after a conformational change and insertion into the target membrane is presented) [10,12,13,15,45,46].

2.4. Phylogenetic Analysis

The relatedness between AP/ALP sequences from the two *Haliclona* species in which they were present, AP sequences from cnidarians, and ALP sequences from other taxa was visualized with an initial maximum likelihood tree (Figure 5). Here we find that the sequence Hi2 from *H. indistincta* nestles well within the AP clade from cnidarians, but no sequence of this type was found in its close sister species, *H. viscosa*. Additional (ALP) sequences from *H. indistincta* and its sister species *H. viscosa* form a distinct and highly supported clade well outside that of the cnidarians, indicating radiation from an additional AP/ALP copy in the ancestor of that species group. This clade is distinct from the other animal ALPs, which also form a monophyletic grouping supported by 91 BP.

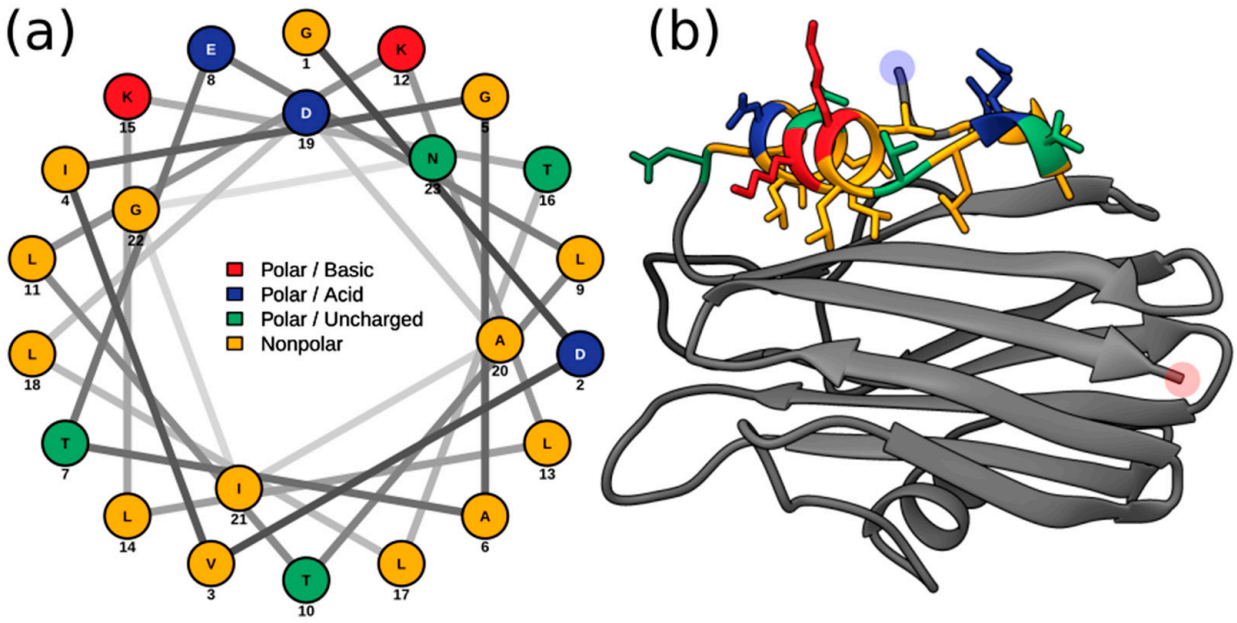


Figure 4. (**a**) Edmundson peptide helical wheel projection of residues 5–28 from Hi2. (**b**) Residues 5–28 of the Phyre2 predicted structure of Hi2 colored in the same manner as (**a**). Blue and red highlights represent the N- and C-terminus, respectively.

When all available AP/ALP sequences from Porifera are added to the analyses, the high number of paralogs present in the dataset obscure the phylogenetic signal and reduce confidence in generating an accurate phylogeny. However, the reconstructed maximum likelihood tree provides insight into the sequence similarity and potential relatedness of the sea sponge ALPs, as well as similar proteins from other phyla. Four distinct groups with strong bootstrap support were produced: (1) ALPs from fungi of the class Glomeromycota (used as outgroup), (2) the majority of sequences from the genus *Haliclona*, (3) anthozoan APs with other sponge ALPs, and (4) ALPs from other invertebrates and chordates (Figure 6). The strong grouping of the *Haliclona* ALPs, independent of all other sequences and sitting at the base of the tree, was a consistently observed phenomenon while testing other alignments for the reconstruction of a final tree (data not shown). These are a sister group to all other APs and ALPs in the dataset. The remaining APs and ALPs form a monophyletic grouping within which there are two clades; one, supported by 100 BP, consisting of the freshwater Hydra and the bilaterians, the other containing the marine cnidarians and sponges (75 BP). The presence of a large number of other sponge sequences has the effect of pulling the Hi2 sequence outside the cnidarian clade, which itself is no longer supported by bootstrapping. Relationships between the other sponge sequences are unclear, as indicated by the low bootstrap support of internal nodes. This is particularly exemplified by the likely spurious placement of Op3 within the clade of anthozoan APs. Frequently, it was observed that additions or subtractions of sequences in the alignment would result in this protein—along with Hi2, Sc1, So1, Tw1, and Tw2—being shifted in and out of the cnidarian AP group. The main exception to this is a second strongly supported clade consisting of ALPs from the genera *Cliona*, *Geodia*, *Scopalina*, and *Tethya*. While this clade may move relative to other sponge sequences, the clade remained intact and distinct from the anthozoan APs. Two groups of *D. avara* sequences were present, both highly supported, but not always remaining together on trees, depending on the comparative sequences included. They were also always distinct from anthozoan APs. Despite having a high sequence similarity to both anthozoan APs and sponge ALPs, those derived from plants, such as bryoporin, were only found to introduce additional noise into the data without significantly changing tree topology and were thus excluded, being hypothesized to be the result of a horizontal gene transfer event (data not shown) [27].

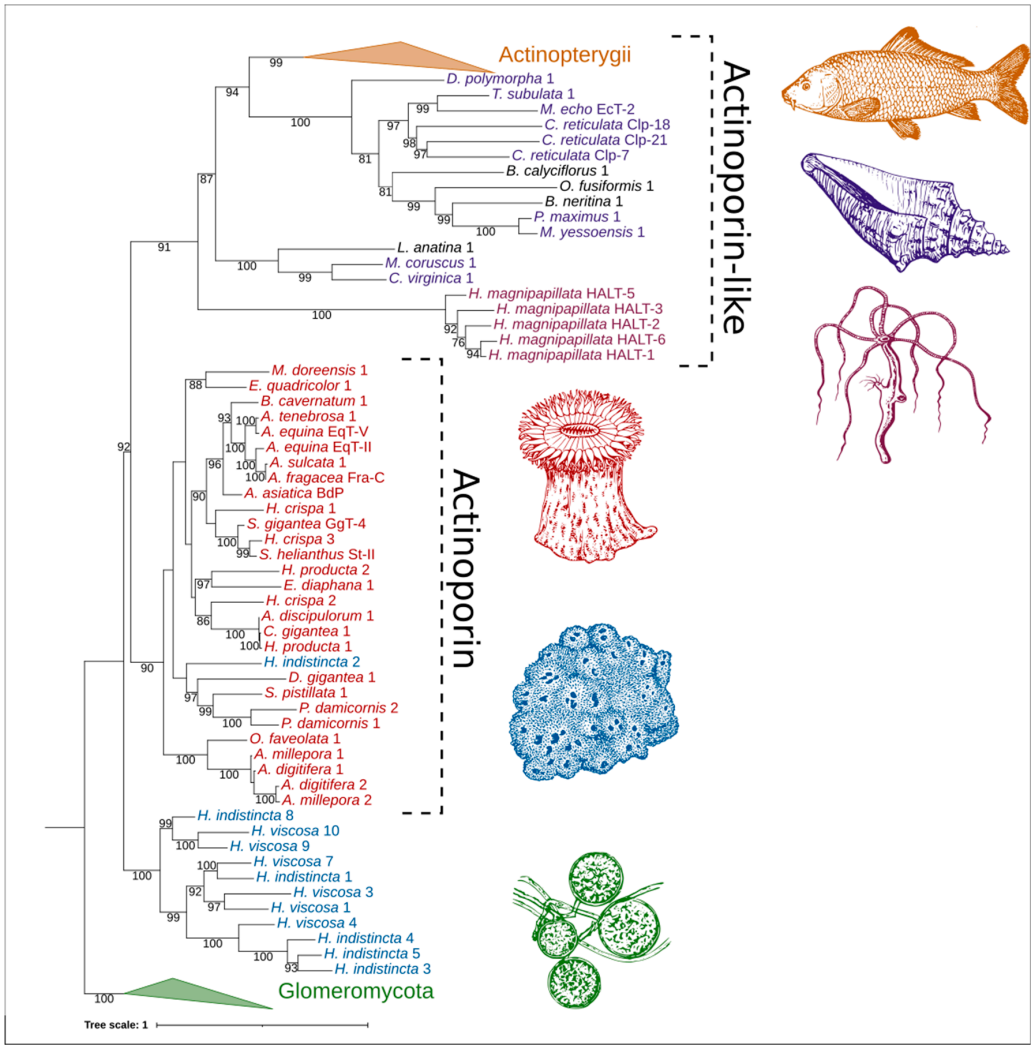


Figure 5. Phylogenetic tree of ALPs from the genus *Haliclona* and similar proteins. Numbers on nodes represent maximum likelihood bootstrap values above 75%. The color scheme is as follows: green represents ALPs from fungi of the class Glomeromycota; blue represents ALPs from *Haliclona* sp.; red represents APs from cnidarians of the class Anthozoa; reddish-purple represents ALPs from cnidarians of the class Hydrozoa; purple represents ALPs from the phylum Mollusca; orange represents ALPs from fish of the class Actinopterygii; black represents ALPs from miscellaneous phyla. Abbreviations are as follows: EqT, equinatoxin; Fra, fragaceatoxin; St, sticholysin; BdP, bandaporin; GgT, gigantoxin; HALT, hydra actinoporin-like toxin; EcT, echotoxin; Clp, coluporin. The glomeromycete vector image is a reproduced copy of the high-resolution image 4341-2_P1 of *Rhizophagus irregularis* (Błaszk., Wubet, Renker, and Buscot) C. Walker and A. Schüßler derived from AAFC/CCAMF (https://agriculture.canada.ca/en/agricultural-science-and-innovation/agriculture-and-agri-food-research-centres-and-collections/glomeromycota-vitro-collection-ginco/catalogue-arbuscular-mycorrhizal-fungi-strains-available-glomeromycetes-vitro-collection) (accessed on 25 November 2021). The sea sponge, sea anemone, seashell, and carp vector images were derived from those of Pearson Scott Foresman under a CC0 1.0 license (https://commons.wikimedia.org/wiki/File:Sponge_(PSF).png; https://commons.wikimedia.org/wiki/File:Anemone_2_(PSF).png; https://commons.wikimedia.org/wiki/File:Seashell_3_(PSF).png; https://commons.wikimedia.org/wiki/File:Carp_(PSF).jpg) (accessed on 26 November 2021). The *Hydra vulgaris* vector image was derived from the Freshwater and Marine Image Bank at the University of Washington under a CC0 1.0 license (https://commons.wikimedia.org/wiki/File:FMIB_50097_Hydra_vulgaris.jpeg) (accessed on 26 November 2021).

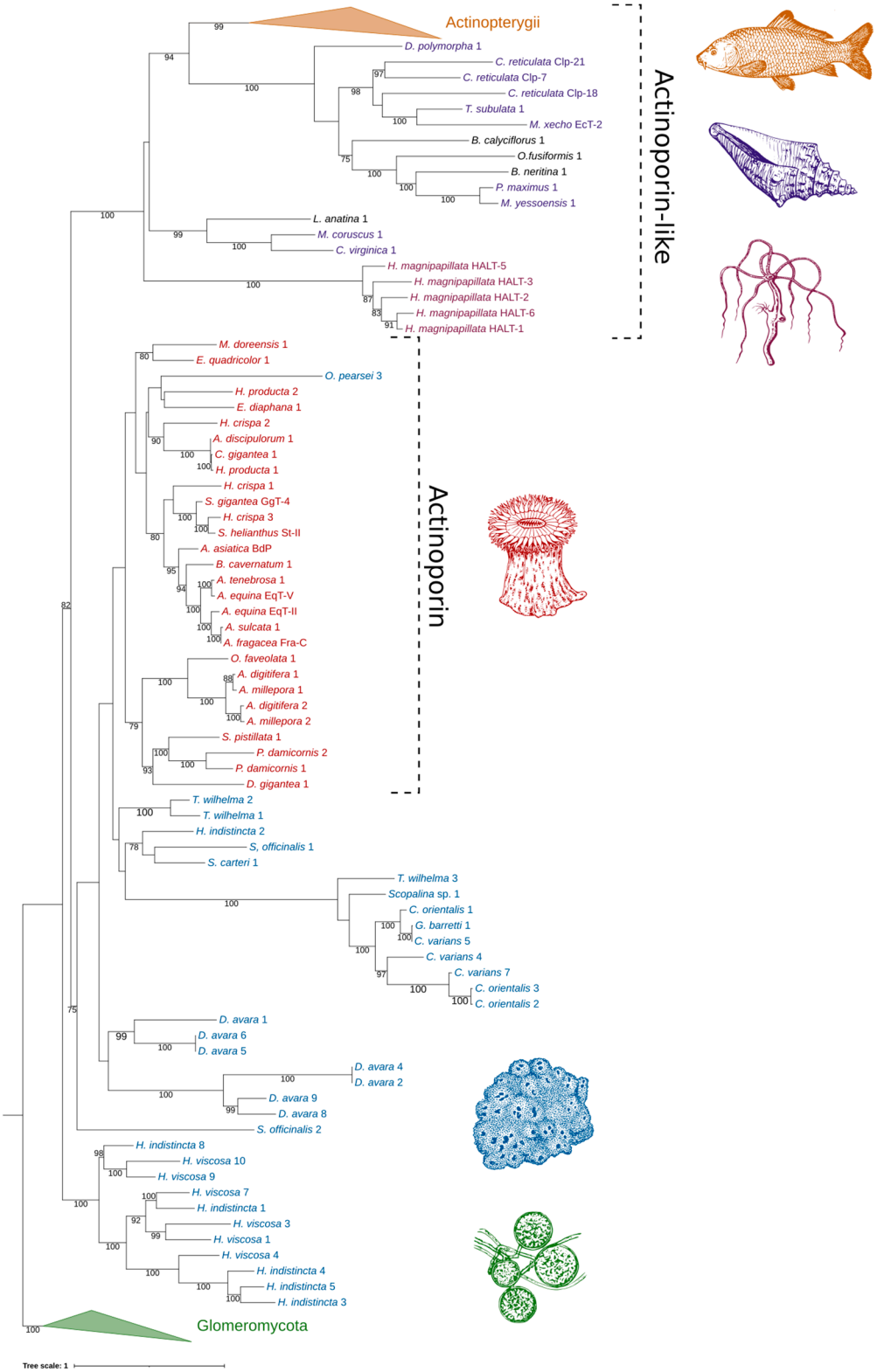


Figure 6. Phylogenetic tree of ALPs from sea sponges and similar proteins. Numbers on nodes represent maximum likelihood bootstrap values above 75%. Blue represents sea sponge AP/ALPs. All other color schemes as well as image acknowledgements are identical to those of Figure 5.

3. Discussion

Cytolytic pore-forming toxins are widely distributed throughout prokaryotic and eukaryotic life and often function as immunological defenses in the latter [47,48]. The proposed AF superfamily includes α-pore-forming toxins derived from diverse eukaryotic lineages with similar predicted protein structure despite low sequence similarity [22]. Members of this family include anthozoan APs, plant APs, hydrozoan ALPs, ALPs from other animals, and fungal fruit-body lectins. Herein, new additions from the phylum Porifera are proposed to belong to the AF superfamily, as members of the classes Demospongiae, Homoscleromorpha, and Calcarea have been shown to possess genes encoding for ALPs. Many functionally characterized AF proteins appear to serve a role in envenomation as they have been localized in the nematocysts of cnidarians and the salivary glands of predatory molluscs [25,26,29,49]. However, AF proteins appear to also have functions other than envenomation, as indicated by their presence in the mesenteric filaments of cnidarians, as well as organisms which do not perform this process, such as bivalves [28,49]. With this in consideration, it is not entirely unusual to observe the presence of these proteins in non-venomous animals, such as sponges, in which they may serve a different ecological function via a similar molecular mechanism.

When considering the taxonomic distribution of ALPs in sponges, no clear pattern can be discerned. The majority of the identified ALPs were derived from the Demospongiae and more specifically from the orders Axinellida, Bubarida, Clionaida, Dendroceratida, Dictyoceratida, Haplosclerida, Poecilosclerida, Suberitida, Tethyida, and Tetractinellida, but not Chondrillida, Spongillida, or Verongiida. However, further analysis at the genus level indicated that these proteins appear to have been lost in species closely related to those which possess ALPs. This is particularly exemplified by the genera *Haliclona* and *Geodia*, in which the transcriptomes of numerous species have been reported [43,50]. Furthermore, ALPs appear to be distributed throughout the phylum Porifera, as several were identified in the classes Homoscleromorpha and Calcarea, but not in Hexactinellida. However, two caveats should be considered regarding these observations. The first is that the high prevalence of these proteins in the Demospongiae is most likely due to sampling bias, as transcriptomes of this class have been disproportionately sequenced compared to the other three. Second, as most of the data analyzed in this study are derived from transcriptomes, a lack of gene expression or sequencing depth cannot be ruled out as an explanation for these genes not having been identified in a species. That said, the absence of ALPs in the *Amphimedon queenslandica* genome does appear to indicate that the loss of ALP genes has occurred in the order Haplosclerida, which may be an explanation for their absence in other species [51].

The predicted structure of most identified ALPs from sponges exhibited a high degree of aligned structural similarity to anthozoan APs, despite a low to moderate sequence similarity. Such an observation is common in studies of AF proteins from other organisms [22,25]. That said, the observed high aligned structural similarity does not necessarily equate to these sea sponge ALPs having the same membrane-binding and pore-forming capabilities. This is particularly exemplified by the observation that the N-terminal region of sea sponge ALPs appears to vary greatly. For example, this region is fully present in Hi2, truncated in Hi3, and completely absent in Hi4. Such a quality is not unique to sea sponge ALPs, as can be seen by analyzing those of hydrozoans [29]. Similarly, an incomplete N-terminal α-helix was also observed on the ALP Dr1 from *D. rerio* and was hypothesized to influence the lack of pore-forming capabilities and specificity towards SM exhibited by the protein [22]. With this in mind, many of the identified ALPs from sea sponges may serve functions other than those associated with well-characterized APs, such as EqT-II. In contrast, the higher similarity of Hi2 to anthozoan APs at both a sequence and structural level could indicate that it is capable of SM recognition and pore formation, which prompted a further analysis on this concept.

A multiple sequence alignment of Hi2 with the model APs EqT-II, Fra-C, and Stn-II indicated that this predicted protein shares numerous conserved residues associated with

the fundamental processes of lipid recognition and membrane binding, insertion of the N-terminal α-helix into the target membrane, and oligomerization allowing for pore formation. The presence of a patch of aromatic residues (Tyr114, Trp117, Tyr134, Tyr138, and Tyr139) in Hi2 is consistent with the IBS observed in the model APs EqT-II, Fra-C, and Stn-II. The major observed deviation is that a substitution of leucine for tryptophan occurs at residue 113 of Hi2. The importance of the equivalent residue in EqT-II, Trp112, in membrane binding and SM recognition has been exemplified by studies in which this residue was mutated to phenylalanine or subject to ^{19}F NMR studies [52,53]. This substitution also appears to be prevalent in nature, as it exists in numerous anemone APs, as well as the ALP Dr1 from *D. rerio* [20,22]. Furthermore, a mutant of EqT-II containing this substitution exhibited similar SM specificity to the wild-type protein [6]. For these reasons, this substitution in Hi2 is not expected to significantly inhibit any possible membrane-binding and SM-recognition capabilities of the protein. While at a lower level of conservation the N-terminal region of Hi2 also appears to form an α-helix of similar length to EqT-II, Fra-C, and Stn-II. The notion that this α-helix is capable of inserting itself into a target membrane is supported by the amphipathic nature of the predicted helical wheel and its hydrophobic moment comparable to previously analyzed AP N-terminal α-helices [20]. Finally, residues associated with oligomerization were somewhat consistent between Hi2 and the model actinoporins. While Hi2 showed conservation at residues Lys76, Ile59, and Trp147, whose equivalents in anthozoan APs are associated with oligomerization, its RGD motif was another site of substitution [12,13,45,46]. Hi2 instead shows a substitution of Lys for Arg and Glu for Asp. That said, Lys and Glu are of a similar charge and hydrophobicity to Arg and Asp, respectively, and may allow for the retention of the oligomerization function [12]. Furthermore, this Lys substitution has been observed in natural actinoporins [20]. Hi2 and many other sponge ALPs also exhibited an acidic pI (Supplementary Table), which is uncharacteristic of the typically basic anthozoan APs [15,54]. This observation is not entirely unusual, however, as there is a precedent for acidic APs derived from anthozoans [20,54,55].

The identification of ALPs from glomeromycete fungi with high sequence similarity to known APs supports the previous notion of a pre-metazoan origin of these proteins [22]. The presence of these genes in fungi, sponges, and cnidarians, and their absence in choanoflagellates, ctenophores, and placozoans, could possibly be explained by a series of gene losses occurring throughout their history. However, additional assemblies from these organisms should be assessed prior to making this conclusion. Furthermore, the observation that many species of sea sponge were the source of numerous ALP isoforms may be an indication that duplication of this gene is a common event in this phylum, similar to the situation in cnidarians and molluscs [25,54]. This observed diversification paired with the fact that many sponge ALPs were derived from transcriptomes indicates that these are not simply genomic relics but do play some sort of functional role in sponges. In the phylogenetic analysis, the sea sponge ALPs were not found to cluster together in a clear monophyletic group, which appears to further exemplify the notion of a high degree of divergence of these proteins in the porifera. However, based on the strong signal that numerous sponge ALPs share with anthozoan APs, to the point of being grouped together, and considering the higher sequence similarity many sponge ALPs have with cnidarian APs, several of these proteins from sponges may instead be classifiable as APs.

4. Materials and Methods

4.1. Sample Collection

Haliclona indistincta (MIIG1388; Appendix A Figure A1a) was collected at Corranroo on 17 May 2019 and *H. viscosa* (MIIG1389 and MIIG1390; Figure A1b) was collected at Bridges of Ross on 1 August 2019. The following sample processing protocol was applied to both species. Visible epibionts were removed. The sponges were rinsed in sterile artificial seawater. The sponges were then dissected into ~1 cm^3 pieces and flash-frozen with liquid nitrogen. Samples were stored at −70 °C until further use. Voucher specimens were stored in ethanol.

4.2. RNA Extraction

A ~1 cm^3 piece of flash-frozen sponge tissue was submerged in 500 µL of Trizol in a 2 mL microcentrifuge tube. The tissue was semi-homogenized by hand with a plastic pestle. An additional 500 µL of Trizol was then added to the sample. The sample was mixed by gently inverting five times and allowed to incubate at room temperature for 5 min. The sample was then inverted and vortexed with a VWR Analogue mini vortex mixer at maximum speed for 2 min. A volume of 100 µL BCP was added to the sample. The sample was mixed by hand for 20 s and then vortexed for 10 s with a VWR Analogue mini vortex mixer at maximum speed. The sample was incubated at room temperature for 5–10 min. The sample was then centrifuged at 16,000 g for 15 min at 6 °C. The clear, aqueous layer at the top was transferred to a fresh microcentrifuge tube. RNA was purified by first adding 500 µL of 100% isopropanol to the aqueous phase sample. The sample was then inverted and vortexed with a VWR Analogue mini vortex mixer at maximum speed for 2 min. The sample was left to incubate for 10 min at room temperature. The sample was then centrifuged at 16,000 g for 15 min at 4 °C. The supernatant was discarded and 1 mL of 75% EtOH was added to the RNA pellet. The pellet was disrupted by vortexing. The RNA sample was then centrifuged for 5 min at 4 °C at 7500 g. The EtOH was carefully removed without disturbing the pellet. The washing with 75% EtOH was repeated once. The RNA sample was allowed to air dry until the edges of the pellet were visible. Finally, the pellet was resuspended in 100 µL molecular-grade water. The RNA sample was kept frozen at −70 °C until further use. A subsample of each RNA extraction was used for quality and quantity assessment on a 2100 Bioanalyzer RNA Eukaryotic Chip.

4.3. Transcriptome Sequencing

Samples were sent to Macrogen, Inc. for the preparation of Illumina TruSeq Stranded mRNA libraries from poly-A selection, with insert sizes of 150 bp. The libraries were sequenced on a Novaseq, with a targeted 40 million reads per sample.

4.4. Transcriptome Assembly

Previously sequenced raw cDNA Illumina reads of *H. cinerea* (culture held at Carna Marine Research Station), *H. oculata* (MIIG1250 and MIIG1251), *H. indistincta* (MIIG1093, MIIG1094, MIIG1095), and *H. simulans* (MIIG1248 and MIIG1249) were acquired from Prof. Grace P. McCormack and Dr. Jose Maria Aguilar-Camacho for use in this study (personal communication) [43].

Raw cDNA reads of *H. cinerea*, *H. indistincta*, *H. oculata*, *H. simulans*, and *H. viscosa* were processed with fastp version 0.2 using default settings to remove adapters and low-quality regions [56]. The processed reads were then assembled with Trinity version 2.8.5 using default settings [57]. Reads were pooled so that one transcriptome per species was assembled. Isoforms and low-expressed transcripts were retained in the final assembly. The longest translated open reading frames per transcript were extracted using TransDecoder version 5.5.0 [58]. Homology searches using these open reading frames as a query against the SwissProt database (accessed on 14 January 2020) with BLASTp version 2.9.0 [59,60], as well as the Pfam database (accessed on 14 January 2020) with HMMER version 3.2.1 [61,62], were performed. Significant homologous alignments were used to guide TransDecoder in identifying additional open reading frames. The completeness of the transcriptomes was then assessed using BUSCO version 5.1.2; specifically, the assemblies were queried against the latest version of the eukaryota_odb10 dataset (downloaded 13 April 2021) [63]. Transcriptome assembly, open reading frame extraction, BUSCO analysis, and the generation of the maximum likelihood trees were performed with an account at the Leibniz Supercomputing Centre.

4.5. Identification of Novel Actinoporin-like Proteins from Sea Sponges

While analyzing the output of the homology search against the Pfam database, it was noticed that numerous translated protein sequences possessed the Pfam domain

PF06369 representing sea anemone cytotoxic proteins, such as actinoporins. Due to the biotechnological potential of actinoporins, this prompted the screening of all publicly available sea sponge genomes [41,51,64–68] and transcriptomes [42,50,65,68–86] to see if other members of the phylum also encoded ALPs in their genome. If transcriptome assemblies were not provided in the original publication, the data were assembled in the same manner as the Irish *Haliclona*. The *Haliclona* ALPs were used as queries in a tblasn search against the other sponge assemblies with an e-value cutoff of 1e-4. The longest open reading frames were then extracted from the hits using TransDecoder version 5.5.0 [58]. These protein sequences were screened against the Pfam database (accessed on 14 January 2020) with HMMER version 3.2.1 [61,62] and only those indicated as possessing domain PF06369 were retained for further analysis. These identified sponge ALPs were also screened against the NCBI Conserved Domain Database to confirm that PF06369 was the primary conserved domain [87]. To explore the possible evolutionary origin of ALPs in animals, the genomes of the choanoflagellates *Monosiga brevicollis* and *Salpingoeca rosetta*, the ctenophores *Mnemiopsis leidyi* and *Pleurobrachia bachei*, and the placozoans *Trichoplax adhaerens* and *Hoilungia hongkongensis* were also screened using EqT-II as a query [88–93]. Furthermore, nineteen choanoflagellate transcriptomes were also screened in a similar manner [94].

4.6. Sequence Analysis and Structural Prediction

All identified sea sponge ALPs were used as a query against the NCBI non-redundant protein database to identify the closest homologous sequence [95]. SignalP 5.0 was used to identify the presence of signal peptides in sea sponge ALPs contained within a complete ORF [44]. The isoelectric point and molecular weight of complete, mature sponge ALPs were determined using the compute pI/Mw tool of ExPASy [96]. Protein structure prediction of all sponge ALPs was performed using Phyre2 Suite version 5.1 [97]. The quality of the predicted protein structure for Hi2 was assessed with ProQ3D and ModFOLD8 [98,99]. Structural alignment of sponge ALPs upon the crystal structure of chain A from EqT-II (1iaz) was performed with TM-Align [100]. Protein structures were visualized using UCSF Chimera version 1.15 [101]. A protein topology plot was created using Pro-origami [102]. A multiple sequence alignment of Hi2, EqT-II (P61914), Fra-C (B9W5G6), and Stn-II (P07845) was performed with MAFFT v7.490, with the L-INS-i alignment method using default settings [103]. The multiple sequence alignment was then visualized using Jalview version 2.11.1.4 [104]. Analysis of the N-terminal α-helix of Hi2 and generation of an Edmundson wheel projection were accomplished using HeliQuest and NetWheels [105,106]. RNA 3D structure prediction was performed with 3dRNA v2.0 [107].

4.7. Phylogenetic Analysis of Actinoporin-like Proteins from Sea Sponges

Sea sponge ALPs derived from complete ORFs were chosen for multiple sequence alignment and phylogenetic analysis. All alignments were performed using MAFFT 7.490, with the L-INS-i alignment method using default settings [103]. APs were represented by well-characterized actiniarian proteins, such as the aforementioned EqT-II, Fra-C, and Stn-II, as well as those from stony and soft corals. ALPs were represented by the series of HALT proteins from H. magnipapillata. It was observed that in general the sea sponge ALPs most consistently aligned with sequences from sea anemones, fungi of the class Glomeromycota, and teleost fish, in that order. To get more sequences for the phylogenetic tree, all complete sponge ALPs were queried against the NCBI nr database against these three taxonomic groups, as well as against molluscs and invertebrates which did not fall under these aforementioned phyla. The top hit from each category for each sponge sequence was then retrieved. This resulted in seven groups of sequences: glomeromycete fungi, sponges, anthozoan cnidarians, hydrozoan cnidarians, molluscs, miscellaneous invertebrates, and teleost fish (Table S5). All signal peptides were removed prior to alignments using SignalP version 5.0 [44]. Each of these groups were separately aligned to the mature sequences of EqT-II, Fra-C, and Stn-II. The individual alignments were then trimmed corresponding

to the boundaries of EqT-II, Fra-C, and Stn-II using using Jalview version 2.11.1.4 [104]. Several sponge ALPs, while complete, where excluded due to being excessively truncated compared to EqT-II or introducing significant gaps. All trimmed sequences were then pooled together and once more aligned. Maximum likelihood trees were constructed in IQ-TREE version 2.1.4 with 1000 bootstrap pseudoreplicates with the intention of visualizing the degree of similarity the sponge ALPs had to APs and ALPs from other phyla [108]. The resulting phylogenetic trees were modified using the Interactive Tree of Life (iTOL) v6 [109].

4.8. Generation of Figures

Several figures were further modified using the GNU Image Manipulation Program 2.10.28 [110] and Inkscape 0.92 [111].

5. Conclusions

Sea sponges, like some other invertebrates, are a source of ALPs. These proteins exhibit a high degree of predicted structural similarity as well as a phylogenetic signal to APs from cnidarians. One AP, Hi2, also possesses a majority of conserved residues associated with essential functions, including the recognition of SM, binding to membranes, oligomerization, and the formation of pores. While their ecological role in sponges remains to be determined, the aforementioned qualities encourage the exploration of these proteins for the biotechnological applications which have been proposed for anthozoan APs.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/md20010074/s1, Table S1. Presence; Table S2. Blastp nr; Table S3. pI-Mw; Table S4. TM-align; Table S5. Trees.

Author Contributions: K.S. conceived the experiments, aided in the collection of samples, carried out all other methodological tasks, prepared figures, and wrote the paper. G.P.M. supervised the research and assisted in interpreting the data and in writing the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 764840. The Department of Further and Higher Education, Research, Innovation and Science, through the Higher Education Authority (HEA), provided funding to higher education institutions to cover costed extensions for research activities that are at risk because of interruptions caused by the COVID-19 pandemic.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw RNA-seq reads are available at the NCBI BioProject PRJNA795170. Transcriptome assemblies, sea sponge ALP sequences, predicted protein structures, multiple sequence alignments, and maximum likelihood trees are available at Mendeley (https://data.mendeley.com/datasets/w9t6zsjjb7/1, accessed on 14 December 2021).

Acknowledgments: The authors thank the technicians of the NUIG Zoology Department for their help in collecting *H. indistincta*. Olivier P. Thomas and Daniel Rodrigues are thanked for the collection of *H. viscosa*. Jose Maria Aguillar-Camacho provided raw data of *H. cinerea*. Members of the Marie Skłodowska-Curie Actions innovative training network IGNITE—Comparative Genomics of Non-Model Invertebrates are thanked for their advice on methodology. The Irish marine biorepository is thanked for providing a high-quality image of *Haliclona viscosa* for the graphical abstract.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

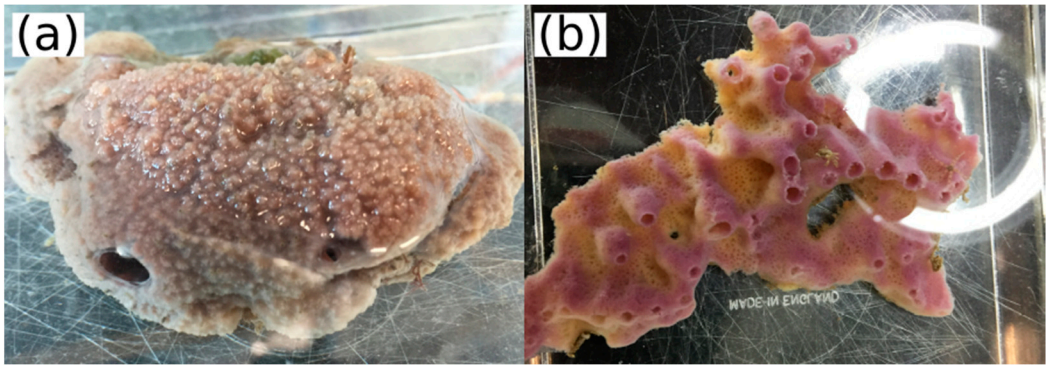


Figure A1. (**a**) *Haliclona indistincta* (MIIG1388). (**b**) *Haliclona viscosa* (MIIG1389).

References

1. Kem, W.R. Sea Anemone Toxins: Structure and Action. *Biol. Nematocysts* **1988**, 375–405.
2. Anderluh, G.; Maček, P. Cytolytic Peptide and Protein Toxins from Sea Anemones (Anthozoa: Actiniaria). *Toxicon* **2002**, *40*, 111–124. [CrossRef]
3. Athanasiadis, A.; Anderluh, G.; Maček, P.; Turk, D. Crystal Structure of the Soluble Form of Equinatoxin II, a Pore-Forming Toxin from the Sea Anemone Actinia Equina. *Structure* **2001**, *9*, 341–346. [CrossRef]
4. Mancheño, J.M.; Martín-Benito, J.; Martínez-Ripoll, M.; Gavilanes, J.G.; Hermoso, J.A. Crystal and Electron Microscopy Structures of Sticholysin II Actinoporin Reveal Insights into the Mechanism of Membrane Pore Formation. *Structure* **2003**, *11*, 1319–1328. [CrossRef]
5. Mechaly, A.E.; Bellomio, A.; Gil-Cartón, D.; Morante, K.; Valle, M.; González-Mañas, J.M.; Guérin, D.M.A. Structural Insights into the Oligomerization and Architecture of Eukaryotic Membrane Pore-Forming Toxins. *Structure* **2011**, *19*, 181–191. [CrossRef] [PubMed]
6. Bakrač, B.; Gutiérrez-Aguirre, I.; Podlesek, Z.; Sonnen, A.F.-P.; Gilbert, R.J.C.; Maček, P.; Lakey, J.H.; Anderluh, G. Molecular Determinants of Sphingomyelin Specificity of a Eukaryotic Pore-Forming Toxin. *J. Biol. Chem.* **2008**, *283*, 18665–18677. [CrossRef]
7. Bakrač, B.; Kladnik, A.; Maček, P.; McHaffie, G.; Werner, A.; Lakey, J.H.; Anderluh, G. A Toxin-Based Probe Reveals Cytoplasmic Exposure of Golgi Sphingomyelin. *J. Biol. Chem.* **2010**, *285*, 22186–22195. [CrossRef]
8. Malovrh, P.; Viero, G.; Serra, M.D.; Podlesek, Z.; Lakey, J.H.; Maček, P.; Menestrina, G.; Anderluh, G. A Novel Mechanism of Pore Formation: Membrane Penetration by the N-Terminal Amphipathic Region of Equinatoxin. *J. Biol. Chem.* **2003**, *278*, 22678–22685. [CrossRef] [PubMed]
9. Rojko, N.; Kristan, K.Č.; Viero, G.; Žerovnik, E.; Maček, P.; Serra, M.D.; Anderluh, G. Membrane Damage by an α-Helical Pore-Forming Protein, Equinatoxin II, Proceeds through a Succession of Ordered Step. *J. Biol. Chem.* **2013**, *288*, 23704–23715. [CrossRef]
10. Drechsler, A.; Potrich, C.; Sabo, J.K.; Frisanco, M.; Guella, G.; Dalla Serra, M.; Anderluh, G.; Separovic, F.; Norton, R.S. Structure and Activity of the N-Terminal Region of the Eukaryotic Cytolysin Equinatoxin II. *Biochemistry* **2006**, *45*, 1818–1828. [CrossRef] [PubMed]
11. Lam, Y.H.; Hung, A.; Norton, R.S.; Separovic, F.; Watts, A. Solid-State NMR and Simulation Studies of Equinatoxin II N-Terminus Interaction with Lipid Bilayers. *Proteins Struct. Funct. Bioinform.* **2010**, *78*, 858–872. [CrossRef]
12. García-Linares, S.; Richmond, R.; García-Mayoral, M.F.; Bustamante, N.; Bruix, M.; Gavilanes, J.G.; Martínez-del-Pozo, Á. The Sea Anemone Actinoporin (Arg-Gly-Asp) Conserved Motif Is Involved in Maintaining the Competent Oligomerization State of These Pore-Forming Toxins. *FEBS J.* **2014**, *281*, 1465–1478. [CrossRef] [PubMed]
13. Tanaka, K.; Caaveiro, J.M.M.; Morante, K.; González-Mañas, J.M.; Tsumoto, K. Structural Basis for Self-Assembly of a Cytolytic Pore Lined by Protein and Lipid. *Nat. Commun.* **2015**, *6*, 6337. [CrossRef]
14. Črnigoj Kristan, K.; Viero, G.; Dalla Serra, M.; Maček, P.; Anderluh, G. Molecular Mechanism of Pore Formation by Actinoporins. *Toxicon* **2009**, *54*, 1125–1134. [CrossRef] [PubMed]
15. Rojko, N.; Dalla Serra, M.; Maček, P.; Anderluh, G. Pore Formation by Actinoporins, Cytolysins from Sea Anemones. *Biochim. Biophys. Acta (BBA)-Biomembr.* **2016**, *1858*, 446–456. [CrossRef] [PubMed]
16. Mutter, N.L.; Soskine, M.; Huang, G.; Albuquerque, I.S.; Bernardes, G.J.L.; Maglia, G. Modular Pore-Forming Immunotoxins with Caged Cytotoxicity Tailored by Directed Evolution. *ACS Chem. Biol.* **2018**, *13*, 3153–3160. [CrossRef]
17. Wloka, C.; Mutter, N.L.; Soskine, M.; Maglia, G. Alpha-Helical Fragaceatoxin C Nanopore Engineered for Double-Stranded and Single-Stranded Nucleic Acid Analysis. *Angew. Chem. Int. Ed.* **2016**, *55*, 12494–12498. [CrossRef]

18. Makino, A.; Abe, M.; Murate, M.; Inaba, T.; Yilmaz, N.; Hullin-Matsuda, F.; Kishimoto, T.; Schieber, N.L.; Taguchi, T.; Arai, H.; et al. Visualization of the Heterogeneous Membrane Distribution of Sphingomyelin Associated with Cytokinesis, Cell Polarity, and Sphingolipidosis. *FASEB J.* **2015**, *29*, 477–493. [CrossRef]
19. Laborde, R.J.; Sanchez-Ferras, O.; Luzardo, M.C.; Cruz-Leal, Y.; Fernández, A.; Mesa, C.; Oliver, L.; Canet, L.; Abreu-Butin, L.; Nogueira, C.V.; et al. Novel Adjuvant Based on the Pore-Forming Protein Sticholysin II Encapsulated into Liposomes Effectively Enhances the Antigen-Specific CTL-Mediated Immune Response. *J. Immunol.* **2017**, *198*, 2772–2784. [CrossRef]
20. Macrander, J.; Daly, M. Evolution of the Cytolytic Pore-Forming Proteins (Actinoporins) in Sea Anemones. *Toxins* **2016**, *8*, 368. [CrossRef]
21. Ben-Ari, H.; Paz, M.; Sher, D. The Chemical Armament of Reef-Building Corals: Inter- and Intra-Specific Variation and the Identification of an Unusual Actinoporin in Stylophora Pistilata. *Sci. Rep.* **2018**, *8*, 251. [CrossRef] [PubMed]
22. Gutiérrez-Aguirre, I.; Trontelj, P.; Maček, P.; Lakey, J.H.; Anderluh, G. Membrane Binding of Zebrafish Actinoporin-like Protein: AF Domains, a Novel Superfamily of Cell Membrane Binding Domains. *Biochem. J.* **2006**, *398*, 381–392. [CrossRef]
23. von Reumont, B.M.; Campbell, L.I.; Richter, S.; Hering, L.; Sykes, D.; Hetmank, J.; Jenner, R.A.; Bleidorn, C. A Polychaete's Powerful Punch: Venom Gland Transcriptomics of Glycera Reveals a Complex Cocktail of Toxin Homologs. *Genome Biol. Evol.* **2014**, *6*, 2406–2423. [CrossRef] [PubMed]
24. Gerdol, M.; Luo, Y.-J.; Satoh, N.; Pallavicini, A. Genetic and Molecular Basis of the Immune System in the Brachiopod Lingula Anatina. *Dev. Comp. Immunol.* **2018**, *82*, 7–30. [CrossRef]
25. Gerdol, M.; Cervelli, M.; Oliverio, M.; Modica, M.V. Piercing Fishes: Porin Expansion and Adaptation to Hematophagy in the Vampire Snail Cumia Reticulata. *Mol. Biol. Evol.* **2018**, *35*, 2654–2668. [CrossRef]
26. Kawashima, Y.; Nagai, H.; Ishida, M.; Nagashima, Y.; Shiomi, K. Primary Structure of Echotoxin 2, an Actinoporin-like Hemolytic Toxin from the Salivary Gland of the Marine Gastropod Monoplex Echo. *Toxicon* **2003**, *42*, 491–497. [CrossRef]
27. Hoang, Q.T.; Cho, S.H.; McDaniel, S.F.; Ok, S.H.; Quatrano, R.S.; Shin, J.S. An Actinoporin Plays a Key Role in Water Stress in the Moss Physcomitrella Patens. *New Phytol.* **2009**, *184*, 502–510. [CrossRef]
28. Takara, T.; Nakagawa, T.; Isobe, M.; Okino, N.; Ichinose, S.; Omori, A.; Ito, M. Purification, Molecular Cloning, and Application of a Novel Sphingomyelin-Binding Protein (Clamlysin) from the Brackishwater Clam, Corbicula Japonica. *Biochim. Biophys. Acta (BBA)-Mol. Cell Biol. Lipids* **2011**, *1811*, 323–332. [CrossRef] [PubMed]
29. Glasser, E.; Rachamim, T.; Aharonovich, D.; Sher, D. Hydra Actinoporin-like Toxin-1, an Unusual Hemolysin from the Nematocyst Venom of Hydra Magnipapillata Which Belongs to an Extended Gene Family. *Toxicon* **2014**, *91*, 103–113. [CrossRef]
30. Luter, H.M.; Bannister, R.J.; Whalan, S.; Kutti, T.; Pineda, M.-C.; Webster, N.S. Microbiome Analysis of a Disease Affecting the Deep-Sea Sponge Geodia Barretti. *FEMS Microbiol. Ecol.* **2017**, *93*, fix074. [CrossRef] [PubMed]
31. Syue, S.-T.; Hsu, C.-H.; Soong, K. Testing of How and Why the Terpios Hoshinota Sponge Kills Stony Corals. *Sci. Rep.* **2021**, *11*, 7661. [CrossRef]
32. Padilla Verdín, C.J.; Carballo, J.L.; Camacho, M.L. Qualitative Assessment of Sponge-Feeding Organisms from the Mexican Pacific Coast. *Open Mar. Biol. J.* **2010**, *4*. [CrossRef]
33. Laport, M.S.; Santos, O.C.S.; Muricy, G. Marine Sponges: Potential Sources of New Antimicrobial Drugs. *Curr. Pharm. Biotechnol.* **2009**, *10*, 86–105. [CrossRef]
34. Qi, S.-H.; Ma, X. Antifouling Compounds from Marine Invertebrates. *Mar. Drugs* **2017**, *15*, 263. [CrossRef]
35. Calcabrini, C.; Catanzaro, E.; Bishayee, A.; Turrini, E.; Fimognari, C. Marine Sponge Natural Products with Anticancer Potential: An Updated Review. *Mar. Drugs* **2017**, *15*, 310. [CrossRef]
36. Müller, W.E.G.; Wang, X.; Binder, M.; von Lintig, J.; Wiens, M.; Schröder, H.C. Differential Expression of the Demosponge (Suberites Domuncula) Carotenoid Oxygenases in Response to Light: Protection Mechanism Against the Self-Produced Toxic Protein (Suberitine). *Mar. Drugs* **2012**, *10*, 177–199. [CrossRef]
37. do Nascimento-Neto, L.G.; Cabral, M.G.; Carneiro, R.F.; Silva, Z.; Arruda, F.V.S.; Nagano, C.S.; Fernandes, A.R.; Sampaio, A.H.; Teixeira, E.H.; Videira, P.A. Halilectin-3, a Lectin from the Marine Sponge Haliclona Caerulea, Induces Apoptosis and Autophagy in Human Breast Cancer MCF7 Cells Through Caspase-9 Pathway and LC3-II Protein Expression. *Anti-Cancer Agents Med. Chem. (Former. Curr. Med. Chem.—Anti-Cancer Agents)* **2018**, *18*, 521–528. [CrossRef] [PubMed]
38. Scarfì, S.; Pozzolini, M.; Oliveri, C.; Mirata, S.; Salis, A.; Damonte, G.; Fenoglio, D.; Altosole, T.; Ilan, M.; Bertolino, M.; et al. Identification, Purification and Molecular Characterization of Chondrosin, a New Protein with Anti-Tumoral Activity from the Marine Sponge *Chondrosia Reniformis* Nardo 1847. *Mar. Drugs* **2020**, *18*, 409. [CrossRef] [PubMed]
39. Mangel, A.; Leitão, J.M.; Batel, R.; Zimmermann, H.; Müller, W.E.G.; Schröder, H.C. Purification and Characterization of a Pore-Forming Protein from the Marine Sponge Tethya Lyncurium. *Eur. J. Biochem.* **1992**, *210*, 499–507. [CrossRef] [PubMed]
40. Wiens, M.; Korzhev, M.; Krasko, A.; Thakur, N.L.; Perović-Ottstadt, S.; Breter, H.J.; Ushijima, H.; Diehl-Seifert, B.; Müller, I.M.; Müller, W.E.G. Innate Immune Defense of the Sponge Suberites Domuncula against Bacteria Involves a MyD88-Dependent Signaling Pathway: Induction of a Perforin-like Molecule. *J. Biol. Chem.* **2005**, *280*, 27949–27959. [CrossRef] [PubMed]
41. Nichols, S.A.; Roberts, B.W.; Richter, D.J.; Fairclough, S.R.; King, N. Origin of Metazoan Cadherin Diversity and the Antiquity of the Classical Cadherin/β-Catenin Complex. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 13046–13051. [CrossRef] [PubMed]
42. Ereskovsky, A.V.; Richter, D.J.; Lavrov, D.V.; Schippers, K.J.; Nichols, S.A. Transcriptome Sequencing and Delimitation of Sympatric Oscarella Species (*O. carmela* and *O. pearsei* sp. Nov) from California, USA. *PLoS ONE* **2017**, *12*, e0183002. [CrossRef] [PubMed]

43. Aguilar-Camacho, J.M.; Doonan, L.; McCormack, G.P. Evolution of the Main Skeleton-Forming Genes in Sponges (Phylum Porifera) with Special Focus on the Marine Haplosclerida (Class Demospongiae). *Mol. Phylogenet. Evol.* **2019**, *131*, 245–253. [CrossRef]
44. Almagro Armenteros, J.J.; Tsirigos, K.D.; Sønderby, C.K.; Petersen, T.N.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 5.0 Improves Signal Peptide Predictions Using Deep Neural Networks. *Nat. Biotechnol.* **2019**, *37*, 420–423. [CrossRef]
45. Anderluh, G.; Barlič, A.; Potrich, C.; Maček, P.; Menestrina, G. Lysine 77 Is a Key Residue in Aggregation of Equinatoxin II, a Pore-Forming Toxin from Sea Anemone Actinia Equina. *J. Membr. Biol.* **2000**, *173*, 47–55. [CrossRef]
46. Morante, K.; Caaveiro, J.M.M.; Viguera, A.R.; Tsumoto, K.; González-Mañas, J.M. Functional Characterization of Val60, a Key Residue Involved in the Membrane-Oligomerization of Fragaceatoxin C, an Actinoporin from Actinia Fragacea. *FEBS Lett.* **2015**, *589*, 1840–1846. [CrossRef]
47. Peraro, M.D.; van der Goot, F.G. Pore-Forming Toxins: Ancient, but Never Really out of Fashion. *Nat. Rev. Microbiol.* **2016**, *14*, 77–92. [CrossRef] [PubMed]
48. Spicer, B.A.; Conroy, P.J.; Law, R.H.P.; Voskoboinik, I.; Whisstock, J.C. Perforin—A Key (Shaped) Weapon in the Immunological Arsenal. *Semin. Cell Dev. Biol.* **2017**, *72*, 117–123. [CrossRef] [PubMed]
49. Basulto, A.; Pérez, V.M.; Noa, Y.; Varela, C.; Otero, A.J.; Pico, M.C. Immunohistochemical Targeting of Sea Anemone Cytolysins on Tentacles, Mesenteric Filaments and Isolated Nematocysts of Stichodactyla Helianthus. *J. Exp. Zool. Part A Comp. Exp. Biol.* **2006**, *305A*, 253–258. [CrossRef]
50. Koutsouveli, V.; Cárdenas, P.; Santodomingo, N.; Marina, A.; Morato, E.; Rapp, H.T.; Riesgo, A. The Molecular Machinery of Gametogenesis in Geodia Demosponges (Porifera): Evolutionary Origins of a Conserved Toolkit across Animals. *Mol. Biol. Evol.* **2020**, *37*, 3485–3506. [CrossRef]
51. Srivastava, M.; Simakov, O.; Chapman, J.; Fahey, B.; Gauthier, M.E.A.; Mitros, T.; Richards, G.S.; Conaco, C.; Dacre, M.; Hellsten, U.; et al. The Amphimedon Queenslandica Genome and the Evolution of Animal Complexity. *Nature* **2010**, *466*, 720–726. [CrossRef]
52. Hong, Q.; Gutiérrez-Aguirre, I.; Barlič, A.; Malovrh, P.; Kristan, K.; Podlesek, Z.; Maček, P.; Turk, D.; González-Mañas, J.M.; Lakey, J.H.; et al. Two-Step Membrane Binding by Equinatoxin II, a Pore-Forming Toxin from the Sea Anemone, Involves an Exposed Aromatic Cluster and a Flexible Helix. *J. Biol. Chem.* **2002**, *277*, 41916–41924. [CrossRef]
53. Anderluh, G.; Razpotnik, A.; Podlesek, Z.; Maček, P.; Separovic, F.; Norton, R.S. Interaction of the Eukaryotic Pore-Forming Cytolysin Equinatoxin II with Model Membranes: 19F NMR Studies. *J. Mol. Biol.* **2005**, *347*, 27–39. [CrossRef] [PubMed]
54. Valle, A.; Alvarado-Mesén, J.; Lanio, M.E.; Álvarez, C.; Barbosa, J.A.R.G.; Pazos, I.F. The Multigene Families of Actinoporins (Part I): Isoforms and Genetic Structure. *Toxicon* **2015**, *103*, 176–187. [CrossRef]
55. Jiang, X.; Chen, H.; Yang, W.; Liu, Y.; Liu, W.; Wei, J.; Tu, H.; Xie, X.; Wang, L.; Xu, A. Functional Expression and Characterization of an Acidic Actinoporin from Sea Anemone Sagartia Rosea. *Biochem. Biophys. Res. Commun.* **2003**, *312*, 562–570. [CrossRef]
56. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [CrossRef]
57. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Trinity: Reconstructing a Full-Length Transcriptome without a Genome from RNA-Seq Data. *Nat. Biotechnol.* **2011**, *29*, 644–652. [CrossRef] [PubMed]
58. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M.; et al. De Novo Transcript Sequence Reconstruction from RNA-Seq Using the Trinity Platform for Reference Generation and Analysis. *Nat. Protoc.* **2013**, *8*, 1494–1512. [CrossRef]
59. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]
60. O'Donovan, C.; Martin, M.J.; Gattiker, A.; Gasteiger, E.; Bairoch, A.; Apweiler, R. High-Quality Protein Knowledge Resource: SWISS-PROT and TrEMBL. *Brief. Bioinform.* **2002**, *3*, 275–284. [CrossRef] [PubMed]
61. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER Web Server: Interactive Sequence Similarity Searching. *Nucleic Acids Res.* **2011**, *39*, W29–W37. [CrossRef]
62. Finn, R.D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; et al. Pfam: The Protein Families Database. *Nucleic Acids Res.* **2014**, *42*, D222–D230. [CrossRef]
63. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [CrossRef]
64. Kenny, N.J.; Francis, W.R.; Rivera-Vicéns, R.E.; Juravel, K.; de Mendoza, A.; Díez-Vives, C.; Lister, R.; Bezares-Calderón, L.A.; Grombacher, L.; Roller, M.; et al. Tracing Animal Genomic Evolution with the Chromosomal-Level Assembly of the Freshwater Sponge Ephydatia Muelleri. *Nat. Commun.* **2020**, *11*, 3676. [CrossRef]
65. Kenny, N.J.; Plese, B.; Riesgo, A.; Itskovich, V.B. Symbiosis, Selection, and Novelty: Freshwater Adaptation in the Unique Sponges of Lake Baikal. *Mol. Biol. Evol.* **2019**, *36*, 2462–2480. [CrossRef] [PubMed]
66. Francis, W.R.; Eitel, M.; Vargas, S.; Adamski, M.; Haddock, S.H.D.; Krebs, S.; Blum, H.; Erpenbeck, D.; Wörheide, G. The Genome of the Contractile Demosponge Tethya Wilhelma and the Evolution of Metazoan Neural Signalling Pathways. *bioRxiv* **2017**, 120998. [CrossRef]

67. Ryu, T.; Seridi, L.; Moitinho-Silva, L.; Oates, M.; Liew, Y.J.; Mavromatis, C.; Wang, X.; Haywood, A.; Lafi, F.F.; Kupresanin, M.; et al. Hologenome Analysis of Two Marine Sponges with Different Microbiomes. *BMC Genom.* **2016**, *17*, 158. [CrossRef]
68. Fortunato, S.A.V.; Adamski, M.; Ramos, O.M.; Leininger, S.; Liu, J.; Ferrier, D.E.K.; Adamska, M. Calcisponges Have a ParaHox Gene and Dynamic Expression of Dispersed NK Homeobox Genes. *Nature* **2014**, *514*, 620–623. [CrossRef] [PubMed]
69. Finoshin, A.D.; Adameyko, K.I.; Mikhailov, K.V.; Kravchuk, O.I.; Georgiev, A.A.; Gornostaev, N.G.; Kosevich, I.A.; Mikhailov, V.S.; Gazizova, G.R.; Shagimardanova, E.I.; et al. Iron Metabolic Pathways in the Processes of Sponge Plasticity. *PLoS ONE* **2020**, *15*, e0228722. [CrossRef]
70. Manousaki, T.; Koutsouveli, V.; Lagnel, J.; Kollias, S.; Tsigenopoulos, C.S.; Arvanitidis, C.; Magoulas, A.; Dounas, C.; Dailianis, T. A de Novo Transcriptome Assembly for the Bath Sponge Spongia Officinalis, Adjusting for Microsymbionts. *BMC Res. Notes* **2019**, *12*, 813. [CrossRef]
71. González-Aravena, M.; Kenny, N.J.; Osorio, M.; Font, A.; Riesgo, A.; Cárdenas, C.A. Warm Temperatures, Cool Sponges: The Effect of Increased Temperatures on the Antarctic Sponge *Isodictya* sp. *PeerJ* **2019**, *7*, e8088. [CrossRef]
72. Leiva, C.; Taboada, S.; Kenny, N.J.; Combosch, D.; Giribet, G.; Jombart, T.; Riesgo, A. Population Substructure and Signals of Divergent Adaptive Selection despite Admixture in the Sponge Dendrilla Antarctica from Shallow Waters Surrounding the Antarctic Peninsula. *Mol. Ecol.* **2019**, *28*, 3151–3170. [CrossRef]
73. Pita, L.; Hoeppner, M.P.; Ribes, M.; Hentschel, U. Differential Expression of Immune Receptors in Two Marine Sponges upon Exposure to Microbial-Associated Molecular Patterns. *Sci. Rep.* **2018**, *8*, 16081. [CrossRef] [PubMed]
74. Revilla-i-Domingo, R.; Schmidt, C.; Zifko, C.; Raible, F. Establishment of Transgenesis in the Demosponge Suberites Domuncula. *Genetics* **2018**, *210*, 435–443. [CrossRef]
75. Kenny, N.J.; de Goeij, J.M.; de Bakker, D.M.; Whalen, C.G.; Berezikov, E.; Riesgo, A. Towards the Identification of Ancestrally Shared Regenerative Mechanisms across the Metazoa: A Transcriptomic Case Study in the Demosponge Halisarca Caerulea. *Mar. Genom.* **2018**, *37*, 135–147. [CrossRef]
76. Leys, S. Ephydatia Muelleri Trinity Transcriptome. Available online: https://era.library.ualberta.ca/items/d643c77d-aefd-45f3-8ddb-985189eebcf2 (accessed on 2 April 2021).
77. Leys, S. Eunapius Fragilis Trinity Transcriptome. Available online: https://era.library.ualberta.ca/items/6139a88f-895d-44a7-bd0d-e22d455d2785 (accessed on 2 April 2021).
78. Simion, P.; Philippe, H.; Baurain, D.; Jager, M.; Richter, D.J.; Di Franco, A.; Roure, B.; Satoh, N.; Quéinnec, É.; Ereskovsky, A.; et al. A Large and Consistent Phylogenomic Dataset Supports Sponges as the Sister Group to All Other Animals. *Curr. Biol.* **2017**, *27*, 958–967. [CrossRef]
79. Díez-Vives, C.; Moitinho-Silva, L.; Nielsen, S.; Reynolds, D.; Thomas, T. Expression of Eukaryotic-like Protein in the Microbiome of Sponges. *Mol. Ecol.* **2017**, *26*, 1432–1451. [CrossRef]
80. Borisenko, I.; Adamski, M.; Ereskovsky, A.; Adamska, M. Surprisingly Rich Repertoire of Wnt Genes in the Demosponge Halisarca Dujardini. *BMC Evol. Biol.* **2016**, *16*, 123. [CrossRef]
81. Guzman, C.; Conaco, C. Comparative Transcriptome Analysis Reveals Insights into the Streamlined Genomes of Haplosclerid Demosponges. *Sci. Rep.* **2016**, *6*, 18774. [CrossRef] [PubMed]
82. Alié, A.; Hayashi, T.; Sugimura, I.; Manuel, M.; Sugano, W.; Mano, A.; Satoh, N.; Agata, K.; Funayama, N. The Ancestral Gene Repertoire of Animal Stem Cells. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E7093–E7100. [CrossRef] [PubMed]
83. Qiu, F.; Ding, S.; Ou, H.; Wang, D.; Chen, J.; Miyamoto, M.M. Transcriptome Changes during the Life Cycle of the Red Sponge, Mycale Phyllophila (Porifera, Demospongiae, Poecilosclerida). *Genes* **2015**, *6*, 1023–1052. [CrossRef]
84. Whelan, N.V.; Kocot, K.M.; Moroz, L.L.; Halanych, K.M. Error, Signal, and the Placement of Ctenophora Sister to All Other Animals. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 5773–5778. [CrossRef] [PubMed]
85. Riesgo, A.; Peterson, K.; Richardson, C.; Heist, T.; Strehlow, B.; McCauley, M.; Cotman, C.; Hill, M.; Hill, A. Transcriptomic Analysis of Differential Host Gene Expression upon Uptake of Symbionts: A Case Study with Symbiodinium and the Major Bioeroding Sponge Cliona Varians. *BMC Genom.* **2014**, *15*, 376. [CrossRef] [PubMed]
86. Riesgo, A.; Farrar, N.; Windsor, P.J.; Giribet, G.; Leys, S.P. The Analysis of Eight Transcriptomes from All Poriferan Classes Reveals Surprising Genetic Complexity in Sponges. *Mol. Biol. Evol.* **2014**, *31*, 1102–1120. [CrossRef]
87. Lu, S.; Wang, J.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; Gwadz, M.; Hurwitz, D.I.; Marchler, G.H.; Song, J.S.; et al. CDD/SPARCLE: The Conserved Domain Database in 2020. *Nucleic Acids Res.* **2020**, *48*, D265–D268. [CrossRef] [PubMed]
88. King, N.; Westbrook, M.J.; Young, S.L.; Kuo, A.; Abedin, M.; Chapman, J.; Fairclough, S.; Hellsten, U.; Isogai, Y.; Letunic, I.; et al. The Genome of the Choanoflagellate Monosiga Brevicollis and the Origin of Metazoans. *Nature* **2008**, *451*, 783–788. [CrossRef]
89. Fairclough, S.R.; Chen, Z.; Kramer, E.; Zeng, Q.; Young, S.; Robertson, H.M.; Begovic, E.; Richter, D.J.; Russ, C.; Westbrook, M.J.; et al. Premetazoan Genome Evolution and the Regulation of Cell Differentiation in the Choanoflagellate Salpingoeca Rosetta. *Genome Biol.* **2013**, *14*, R15. [CrossRef]
90. Ryan, J.F.; Pang, K.; Schnitzler, C.E.; Nguyen, A.-D.; Moreland, R.T.; Simmons, D.K.; Koch, B.J.; Francis, W.R.; Havlak, P.; Smith, S.A.; et al. The Genome of the Ctenophore Mnemiopsis Leidyi and Its Implications for Cell Type Evolution. *Science* **2013**, *342*, 1242592. [CrossRef]
91. Moroz, L.L.; Kocot, K.M.; Citarella, M.R.; Dosung, S.; Norekian, T.P.; Povolotskaya, I.S.; Grigorenko, A.P.; Dailey, C.; Berezikov, E.; Buckley, K.M.; et al. The Ctenophore Genome and the Evolutionary Origins of Neural Systems. *Nature* **2014**, *510*, 109–114. [CrossRef] [PubMed]

92. Srivastava, M.; Begovic, E.; Chapman, J.; Putnam, N.H.; Hellsten, U.; Kawashima, T.; Kuo, A.; Mitros, T.; Salamov, A.; Carpenter, M.L.; et al. The Trichoplax Genome and the Nature of Placozoans. *Nature* **2008**, *454*, 955–960. [CrossRef]
93. Eitel, M.; Francis, W.R.; Varoqueaux, F.; Daraspe, J.; Osigus, H.-J.; Krebs, S.; Vargas, S.; Blum, H.; Williams, G.A.; Schierwater, B.; et al. Comparative Genomics and the Nature of Placozoan Species. *PLoS Biol.* **2018**, *16*, e2005359. [CrossRef]
94. Richter, D.J.; Fozouni, P.; Eisen, M.B.; King, N. Gene Family Innovation, Conservation and Loss on the Animal Stem Lineage. *eLife* **2018**, *7*, e34226. [CrossRef]
95. Pruitt, K.D.; Tatusova, T.; Maglott, D.R. NCBI Reference Sequence (RefSeq): A Curated Non-Redundant Sequence Database of Genomes, Transcripts and Proteins. *Nucleic Acids Res.* **2005**, *33*, D501–D504. [CrossRef]
96. Artimo, P.; Jonnalagedda, M.; Arnold, K.; Baratin, D.; Csardi, G.; de Castro, E.; Duvaud, S.; Flegel, V.; Fortier, A.; Gasteiger, E.; et al. ExPASy: SIB Bioinformatics Resource Portal. *Nucleic Acids Res.* **2012**, *40*, W597–W603. [CrossRef]
97. Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J.E. The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. *Nat. Protoc.* **2015**, *10*, 845–858. [CrossRef] [PubMed]
98. Uziela, K.; Menéndez Hurtado, D.; Shu, N.; Wallner, B.; Elofsson, A. ProQ3D: Improved Model Quality Assessments Using Deep Learning. *Bioinformatics* **2017**, *33*, 1578–1580. [CrossRef]
99. McGuffin, L.J.; Aldowsari, F.M.F.; Alharbi, S.M.A.; Adiyaman, R. ModFOLD8: Accurate Global and Local Quality Estimates for 3D Protein Models. *Nucleic Acids Res.* **2021**, *49*, W425–W430. [CrossRef] [PubMed]
100. Zhang, Y.; Skolnick, J. TM-Align: A Protein Structure Alignment Algorithm Based on the TM-Score. *Nucleic Acids Res.* **2005**, *33*, 2302–2309. [CrossRef]
101. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [CrossRef]
102. Stivala, A.; Wybrow, M.; Wirth, A.; Whisstock, J.C.; Stuckey, P.J. Automatic Generation of Protein Structure Cartoons with Pro-Origami. *Bioinformatics* **2011**, *27*, 3315–3316. [CrossRef] [PubMed]
103. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef] [PubMed]
104. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.; Clamp, M.; Barton, G.J. Jalview Version 2—a Multiple Sequence Alignment Editor and Analysis Workbench. *Bioinformatics* **2009**, *25*, 1189–1191. [CrossRef] [PubMed]
105. Gautier, R.; Douguet, D.; Antonny, B.; Drin, G. HELIQUEST: A Web Server to Screen Sequences with Specific α-Helical Properties. *Bioinformatics* **2008**, *24*, 2101–2102. [CrossRef]
106. Mól, A.R.; Castro, M.S.; Fontes, W. NetWheels: A Web Application to Create High Quality Peptide Helical Wheel and Net Projections. *bioRxiv* **2018**, 416347. [CrossRef]
107. Wang, J.; Wang, J.; Huang, Y.; Xiao, Y. 3dRNA v2.0: An Updated Web Server for RNA 3D Structure Prediction. *Int. J. Mol. Sci.* **2019**, *20*, 4116. [CrossRef]
108. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [CrossRef]
109. Letunic, I.; Bork, P. Interactive Tree of Life (ITOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [CrossRef] [PubMed]
110. GNU Image Manipulation Program 2.10.28. Available online: https://www.gimp.org/ (accessed on 30 December 2020).
111. Inkscape 0.92. Available online: https://inkscape.org/ (accessed on 24 January 2020).