universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

„Ecophysiology of the symbiont microbiome in the sponge *Ianthella basta* – from 'omics' to single cell function"

verfasst von / submitted by

Florian Moeller

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien, 2023 / Vienna, 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 794 685 437
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Biologie
Betreut von / Supervisor:	Univ.-Prof. Mag. Dr. Michael Wagner

> Of the eons of geological periods recorded in the stratifications of the earth: of the myriad minute entomological organic existences concealed in cavities of the earth, beneath removable stones, in hives and mounds, of microbes, germs, bacteria, bacilli, spermatozoa: of the incalculable trillions of billions of millions of imperceptible molecules contained by cohesion of molecular affinity in a single pinhead: of the universe of human serum constellated with red and white bodies, themselves universes of void space constellated with other bodies, each, in continuity, its universe of divisible component bodies of which each was again divisible in divisions of redivisible component bodies, dividends and divisors ever diminishing without actual division till, if the progress were carried far enough, nought nowhere was never reached.
>
> ULYSSES
> JAMES JOYCE
> [17]

But if we resort to an indirect test, and ask Nature 'Who are the fittest: those who are continually at war with each other, or those who support one another?' we at once see that those animals which acquire habits of mutual aid are undoubtedly the fittest. They have more chances to survive, and they attain, in their respective classes, the highest development of intelligence and bodily organization. If the numberless facts which can be brought forward to support his view are taken into account, we may safely say that mutual aid is as much a law of animal life as mutual struggle, but that, as a factor of evolution, it most probably has a far greater importance, inasmuch as it favours the development of such habits and characters as insure the maintenance and further development of the species, together with the greatest amount of welfare and enjoyment of life for the individual, with the least waste of energy. [...]

The first thing which strikes us as soon as we begin studying the struggle for existence under both its aspects – direct and metaphorical – is the abundance of facts of mutual aid, not only for rearing progeny, as recognised by most evolutionists, but also for the safety of the individual and for providing it with the necessary food. With many large divisions of the animal kingdom mutual aid is the rule. Mutual aid is met with even amidst the lowest animals, and we must be prepared to learn some day, from the students of microscopical pond-life, most wonderful facts of mutual aid, even from the life of microorganisms.

MUTUAL AID AMONG ANIMALS.
KROPOTKIN, P.
The Nineteenth century: a monthly review, Mar. 1877-Dec. 1900; London
Vol. 28, Iss. 163, (Sep 1890): 337-354.

Acknowledgements

How did we get here if not for you.

Table of Contents

CHAPTER I

Introduction

Outline of thesis

Life emerging and developing in a fluid continuum of synergy, syntrophy, and symbiosis

Life on earth started at the edge of the thermodynamic limit. As organic matter organized itself within the prebiotic churn, through the self-assembly of proteins and nucleic acids in thermodynamically advantageous configurations, the eventual emergence of cellularity allowed for the efficient compartmentalization of energy, thereby allowing for a distinction to be made between different individual forms of life. From this primordial milieu, these first life forms emerged as microbes and organized amongst themselves through evolution, into various lineages dictated primarily by the methods of energy capture and the means of genetic propagation. Although the microbial last common ancestor (LCA) is believed to have initially eked out a living as a methanogen and/or acetogen by fixing CO_2 via the energy provided by H_2 [1–3], an abundant microbial community, sustained by these *ur*-producers, eventually developed, that would have recycled the newly generated organic material while likely gaining energy via methane oxidation, sulfate reduction and/or anoxygenic photosynthesis [4, 5]. The first microbial communities were thus possibly living in syntrophy, defined simply as 'obligately mutualistic metabolism' between microbial partners in an interdependent cycle of chemical exchange [6].

In fact, over 20 years ago, the Syntrophy hypothesis was advanced as a mechanism for the evolution of eukaryotes arising from a tripartite metabolic symbiosis wherein a fermenting deltaproteobacterial host transfers interspecies hydrogen to an endosymbiotic methanogenic archaeon and the resultant methane recycled by a versatile methanotrophic, facultative aerobic alphaproteobacterium (HM Syntrophy) [7, 8]. Beginning with phylogenomic trees where eukaryotes branched within the archaea [9], along with the discovery of Asgard archaea, a phylogenetically deep-branching lineage with more similar genes and shared genes in common with eukaryotes than other archaea [10, 11], the view that eukaryotes are but a secondary domain deriving from the evolutionary merger of specific archaeal and bacterial lineages, began to come into widespread acceptance [12–14]. Most eukaryogenetic models now commonly propose either a two-partner or three-partner symbiosis [15–18] highlighting the importance of this set of relations, with syntrophy representing an important facet, to the development of complex life.

Sponges: from unicellular to multicellular, from antiquity to modernity

The first eukaryotes are postulated to have emerged between the Great Oxidation Event (GOE), some 2.4 billion years ago (Ga), and the minimum age of the oldest unambiguous eukaryotic fossils [19], but it took until roughly 600 million years ago (Ma) before oxygen levels would approach modern conditions [20, 21]. It was during this oxygen ramp-up that unicellular eukaryotes began to stably associate into multicellular forms, first as colonies and simple multicellular aggregations, and finally as complex multicellular life forms exhibiting capabilities for cell-cell adhesion, intercellular communication, tissue differentiation mediated by networks of regulatory genes, programmed cell death, and a compartmentalized 3-D organization delineating regions of direct contact with the environment. The *Bangomorha* spp. red algae are the first incontrovertible multicellular eukaryotes to appear in the geologic record ~1.2 Ga [22] and are followed much later by macrophyte green algae, which arose 750 Ma [23, 24], and the metazoans, which arose ~650 Ma [25–27].

Sponges (phylum *Porifera*) are considered to be one of the most ancient [28, 29], if not the most ancient lineage of metazoans [30–32], with some recent discoveries even suggesting their emergence ~890 Ma [33]. Due to their striking physiological resemblance to the sponge cell type, the choanocyte, the unicellular protistan choanoflagellates have long been considered to be the closest relatives of the first animals [34, 35], a hypothesis more recently borne out by phylogenomic studies [36, 37]. Furthermore, since choanoflagellates can alternate between single-celled and multicellular 'rosette' states [38, 39] resembling a sponge choanosome, much research has been devoted to them to understand the origins of metazoan multicellularity. Recent research demonstrating that a newly isolated choanoflagellate, *Barroeca monosierra*, can harbor a stable and relatively simple microbiome within its rosette extracellular matrix [40], and that another choanoflagellate, *Salpingoeca rosetta* can undergo either multicellular development or mating in response to secreted bacterial cues [41–43], suggest a potentially undiscovered and additional vital role for microbial symbiosis in eukaryotic evolution. It is perhaps safe to state that sponge-microbe symbioses may thus

represent the oldest extant example of microbial-metazoan symbiosis and that microbes may have had a significant role in the development of animal multicellularity.

Over 8,600 formally described species of sponges currently inhabit freshwater to saltwater habitats, in shallow and deepwater areas, from the tropics to the poles. As sessile filter feeders feeding mainly on dissolved organic matter and particulate organic matter, which may include a variety of microorganisms, they can occupy up to 80% of the available substrate, support a variety of infauna species, connect the benthic and pelagic zones by filtering vast amounts of seawater, and mediate biogeochemical fluxes by respiring organic matter and facilitating the uptake and release of nutrients. Sponges are thus ecologically significant components of benthic environments. They lack true specialized tissues, organs, a nervous system, true muscle, a blood or circulatory system, a stomach or digestive system, and instead rely on a specialized aquiferous system for food acquisition and waste disposal (Figure 1a). Water is absorbed into the interior channels of the sponge through the numerous small pores (ostia) dotting the outer dermal layer and finally discharged through larger apertures known as oscula by the beating activity of internal flagella-bearing cells called choanocytes. After capture, food particles are carried into the sponge body's middle mesohyl region where they are phagocytosed by amoeboid sponge cells known as archaeocytes. The mesohyl is made up of non-cellular collagen fibers, spicules, and several motile cell types that are responsible for reproduction, digestion, and spicule synthesis. The sponge body is made up of two cell layers: an outer pinacoderm and an interior choanoderm.

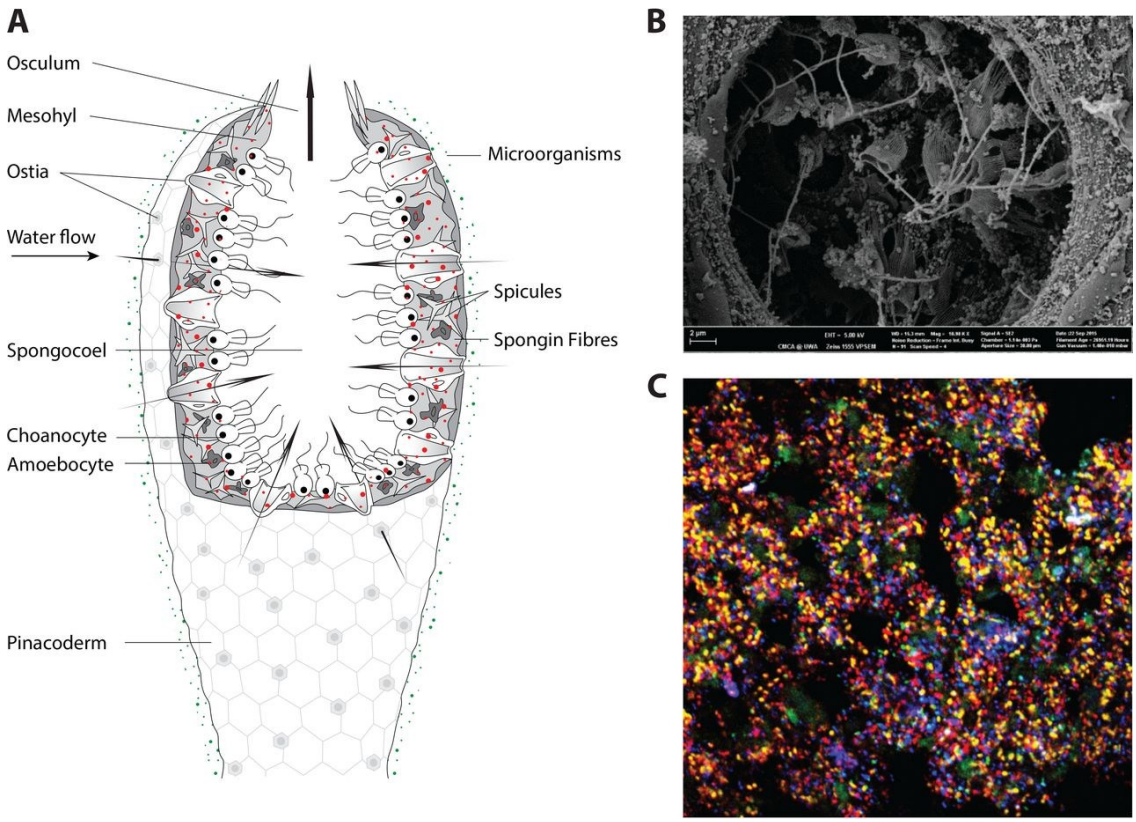


FIG 1 Adapted from [44]. (A) Overview of a marine sponge body plan showing a schematic representation of an asconoid sponge. Seawater enters the sponge through tiny pores (ostia) in the surface layer (pinacoderm). Flagellated choanocyte cells are arranged in rings lining the sponge's aquiferous system and are responsible for moving water through the sponge until it is discharged via the exhalant opening (osculum). Environmental microorganisms are shown as green cells, whereas the dense community of symbiotic microorganisms that exists within the mesohyl is portrayed as red cells. Siliceous spicules provide additional structural support to many sponge species. (B) Multicolor double labeling of oligonucleotide probes for fluorescence *in situ* hybridization (DOPE-FISH) analysis of the sponge mesohyl using rRNA-targeted probes for specific sponge-associated microorganisms reveals the density of microbial cells. *Poribacteria*, yellow; *Nitrospira*, pink; *Chloroflexi*, cyan; *Deltaproteobacteria*, light green; *Gammaproteobacteria*, red; *Archaea*, blue. Larger dark-green areas are sponge autofluorescence. Photo courtesy of Faris Behnam and Michael Wagner, Department of Microbial Ecology, University of Vienna. (C) Scanning electron micrograph showing the density of microbial cells surrounding the sponge choanocyte chamber, where individual flagellated choanocyte cells are responsible for drawing water through the sponge and removing bacteria and other particles for subsequent digestion. Photo courtesy of Peta Clode, Centre for Microscopy, Characterisation and Analysis, University of Western Australia.

Naturalists have long been fascinated by sponges, from antiquity to classical modernity, whether it be Aristotle who noted their animal-like characteristics and symbiotic relationships to small crustacean and fish species [45], to Linnaeus and Lamarck who grouped them midway between the plant and animal kingdoms alongside corals, sea anemones and medusae [46]. By the late 18th century, sponges began to be categorically classified as animals, although the debate would linger for another century [47]. Ernst Haeckel's Gastrea

theory [48], which posits that a hollow ball of clonal flagellated cells, similar to sponge choanocytes, was the initial step in the evolution of animal multicellularity, remains, with some modern adaptations, one of the oldest and most persistent evolutionary theories on the origin of metazoans [49–51].

The 18th century also saw significant advancements in the study of parasites, and their associated plant and animal diseases, but it was not until the 19th century that the germ theory of disease and thus the appreciation for the impact of microbial actors began to crystallize [52]. By 1879, the term 'symbiosis' was introduced into the scientific discourse by de Bary who highlighted the 'living together of unlike organisms' with descriptions of the duckweed fern (genus *Azolla*) living in mutualism with cyanobacteria, and lichens, the dependent association of certain fungi with algae or cyanobacteria that can survive independently [53]. Descriptions of mycorrhizae, symbiotic fungal associations with vascular plants in the rhizosphere, as well as mutualistic relations between nitrogen-fixing bacteria and other vascular plants, especially legumes, followed soon thereafter [52]

The catalog of microbial symbiotic associations in various configurations across the tree of life continued to grow in the 20th century. Within sponges, the observation of putative symbiotic bacteria (Figure 1b and 1c) became more and more frequent (see references in [54]) until the pioneering work from the 1970s/1980s [55–59] firmly established the field of sponge symbiosis with the introduction of still relevant concepts such as the existence of sponge-specific microbes [60] and their purported ability to escape from sponge digestion in the mesohyl via the masking of capsular sheaths [61]. Other findings established in this era include the fact that sponge symbionts can make up to 38% of the sponge tissue volume [54] (Figure 1b and 1c), and can mediate important biogeochemical and functional roles such as carbon fixation [62], along with the provision of photosynthates, mostly in the form of glycerol by cyanobacterial symbionts [63], and N_2 fixation [58]. Research into the biogeochemical functions of sponges continued to progress, such that nutrient flux analyses and stable isotope studies coupled sometimes with emerging molecular biology methods, began to demonstrate a variety of symbiont-mediated process such as methanotrophy (along with the ingestion of the methanotrophs by the host sponge [64]), nitrification [65–67], denitrification [67, 68], anammox [68], and sulfate reduction [69].

The high modern sponge – the holosponge

As the molecular revolution continued unabated into the 21st century with the rapid advancement of sequencing technologies, bioinformatic processing power, and sophisticated state-of-the-art chemical instrumentation, a variety of conclusions regarding the marine sponge microbiome started to coalesce. Microorganisms spanning 60 different phyla are now known to associate with marine sponges [70–72], although the abundance of symbionts [73], as well as the levels of diversity and richness of these symbiont communities can vary between sponge species [70, 72, 74]. These symbionts frequently belong to evolutionary clades that only contain sequences originating from sponges (and occasionally corals), indicating extremely specific connections [75, 76]. Furthermore, comparative analyses of the metagenomes of sponge-associated and seawater communities revealed that features such as transposable elements, defense mechanisms (such as CRISPR-Cas, toxin-antitoxin, and restriction-modification systems), and eukaryote-like proteins that may be involved in symbiont recognition and interaction were more prevalent in symbiont consortia [77–80], and that this pattern of enrichment reflects a functional convergence among marine sponge microbiomes.

A significant factor in the evolutionary and ecological survival of sponges is the large variety of secondary metabolites that sponges and/or the accompanying microorganisms make that are known to offer protection against predators and epibionts [81]. In fact, sponges are considered to be the most important source for marine product discovery [82]. In some cases, the genes required for heterologous biotechnological production of the compounds have also been identified [83, 84] and have provided important new insights into the origin (i.e., sponge or symbiont derivation) of sponge natural product chemistry, with numerous cases demonstrating that symbiotic bacteria can be key contributors to the production of marine natural therapeutics [84–87].

While much of previous research on symbiont contributions to sponge metabolism was largely focused on autotrophic processes, the majority of sponges are heterotrophic filter feeders, (where dissolved organic matter — DOM — in fact, can comprise 81–95% of the daily diet of coral reef sponges [88–90]) and exhibit rapid cell turnover, thereby releas-

ing large amounts of particulate organic matter [91, 92]. Although the sponge microbiome has primarily been assumed to have a large role in the assimilation of DOM [93, 94], several compound-specific stable isotope probing (SIP) studies have indicated that both host sponge cells and bacterial symbionts may be active in DOM assimilation [95–97]. Not only have the previous metagenome-based comparative analyses shed light on the adaptive and defensive traits that directly enable symbiotic relationships with sponges, while also confirming the presence of the key genes involved in nitrification, denitrification, and anammox, but they further hint at additional syntrophic interactions mediated by the marine sponge microbiome. A diverse and abundant range of membrane transporters (e.g., ABC-type transporters) are now known to be encoded by the sponge microbiome [78] that may facilitate DOM import and the succeeding breakdown of complex carbohydrates, a seemingly defining characteristic of sponge symbioses in regard to carbon metabolism [80, 98], and there is growing evidence that the symbionts may consume the extracellular matrix and the biomass of sponge cells [80, 98, 99]. The fundamental functions of the sponge microbiome may also confer possible advantages to host metabolism by supplying it with micronutrients like vitamin B_1 and vitamin B_{12}, pathways for which, have been found to be more abundant in sponge symbionts [77, 78, 100, 101], indicating that they may be able to meet the needs of the host, and be facilitated by the aforementioned membrane transporters.

The prevalence and importance of the numerous documented cases of microbial symbioses with multicellular eukaryotes has brought forth the concept of the 'holobiont' which encompasses the host animal and its associated microbiome [102], and which can be regarded as complex ecosystems with interdependent inhabitants sustaining a variety of processes, that can only function as a whole [103–105]. These individual ecosystems then engage in impactful interactions with nearby holobionts, thereby integrating themselves into bigger communities and ecosystems that engage in interactions at ever larger scales, such that even one member of the holobiont can have an impact that is much greater than the whole metaorganism [102]. Notable examples of these 'scaling' ecological impacts exhibited by sponge holobionts include their contribution to ecosystem gross primary productivity either via photoautotrophy [106] or chemoautotrophy (particularly in deep-sea environments) [107, 108], the provision of putative anti-predatory bioactive compounds [101, 109] to sponge communities that are demonstrably structured by predation [110, 111]

(thereby influencing benthic community structure), and the assimilation and recycling of DOM by sponge holobionts within reef food webs [95, 112]. With the exception of cases involving sponge holobiont contributions to gross primary productivity via chemoautotrophy, a persistent challenge in ecosystem- and holobiont-scale studies continues to be to link diversity with function, and even once function has been attributed, to properly apportion the contributions of each holobiont member.

The list of proposed symbiotic functions from this era have undoubtedly grown quite expansive, and unsurprisingly the examples of clear-cut functional assignments to specific symbionts has been rather rare. Dominant sponge holobiont members are notoriously hard to cultivate and experimentally viable sponge holobiont cell models or axenic cell cultures are also currently lacking. Although metatranscriptomic and metaproteomic-based studies of sponge holobionts have been undertaken [100, 113–117], which demonstrably corroborate the expression of previously detected genes and pathways involved in many key proposed functions, the lack of targeted physiological experiments in these studies precludes unequivocal functional assignments to specific sponge symbionts. Several recent exciting studies in sponge holobiont systems have employed pulse-chase isotopic tracer techniques along with nanoscale secondary ion mass spectrometry (NanoSIMS), and have elucidated, at the cellular level, the directionality and type of cells (e.g., choanocyte vs. microbial) involved in DOM or POM uptake and processing [118–120]. Although groundbreaking these studies do not account for phylogenetic identity, or specific metabolic pathways, and use non-targeted bulk DOM preparations (e.g. glucose, algal-derived DOM, bulk amino acids). Nevertheless, the breadth and scope of the functions that have been hypothesized, along with the use of state-of-the-art techniques, provide numerous opportunities for the application of focused, hypothesis-based studies in order to continue to make tighter and more detailed links between diversity and function.

Research objectives

Marine sponges often contain highly diverse microbial communities that contribute symbiotic functions essential for host fitness and survival. Whilst this makes them attractive models for symbiosis research, the high microbial complexity can also make it difficult to unravel

the nature of the symbiotic interactions. The sponge *Ianthella basta* is an abundant and widely distributed species found throughout the Indo-West Pacific [121] that maintains highly stable symbioses with three dominant microorganisms. With such low microbial diversity and high symbiont specificity [70, 122], *I. basta* represents an ideal model for detailed analysis of sponge symbiotic function. The three primary symbionts of *I. basta* comprise an α-proteobacterium, a γ-proteobacterium and a thaumarchaeote and these associations are maintained over broad latitudinal gradients [123], in the face of acute local and global environmental pressures [124] and across different host health states [123]. To explore the entire symbiome of *I. basta*, extensive metagenomic and metaproteomic analyses of these microbial symbionts were undertaken and combined with targeted experimental and *in silico* investigations, to ascertain the nature of the multi-species interactions.

Firstly, we wanted to explore the genome and the proteome as well as the ecophysiology of the dominant thaumarchaeal symbiont of *I. basta*. Thaumarchaea are frequently found in sponge hosts and can make up significant portions of the community, in some cases even dominating the communities [125], with a particularly high diversity of Thaumarchaeota falling into sponge specific clusters [75]. In fact, aside from associations with the mammalian skin (humans included) [126–128] and ascidian [129–131] microbiomes, sponges are one of the few metazoans that Thaumarchaea consistently associate with in symbiotic relationships. Thaumarchaeota have been discovered in sponge larvae [132] and within adult sponges that acquire their symbionts horizontally, demonstrating the same variety in transmission routes as reported in bacterial sponge symbionts [133]. Thaumarchaeota play a significant role in sponge nitrogen metabolism and the elimination of nitrogenous waste since they are well-known ammonia-oxidizers [134, 135]. While many studies simply assume that the presence of ammonia oxidation in sponges is indicated by the molecular detection of microbes associated with known nitrifiers [78, 136], this assumption may be misrepresentative given the functional versatility of Thaumarchaeota [137–140]. Chapter III thus details the first detailed study of a sponge thaumarchaeal symbiont that combines comparative genomics, phylogenomics, and proteomics in conjunction with focused physiological experiments using stable-isotope based functional assays along with the aid of an ammonia-oxidizing archaeal-specific inhibitor.

Next, as reported in Chapter IV, the ecophysiology of the dominant γ-proteobacterial symbiont was similarly interrogated within a metaproteogenomic context and with ecophysiological experiments, but this time with a primary focus on the environmentally abundant organosulfonate taurine as an intermediate for sponge-symbiont interactions and with the additional application of NanoSIMS to track stable isotope-labeled taurine within the sponge holobiont. Taurine (2-aminoethanesulfonate) is widespread in animals and algae but is rare in the bacterial and plant kingdoms. Among the marine invertebrates, arthropods and molluscs tend to have the highest concentrations of taurine (reviewed in [141]), reaching concentrations as high as 182 µmol/g wet wt in the bivalve *Solemya reidi* [142] while reaching up to 74% of the free amino acid (FAA) pool in some bivalve species and comprising up to 39% of the FAA pool in sponges [143]. Concentrations of taurine reported from marine sponges range between 1.05-12.5 µmol/g wet wt sponge [144] although it is important to note that the concentrations can vary considerably, even amongst closely related species from the same genus [143]. Reduced sulfur compounds are well known drivers of marine symbiosis in chemoautotrophic systems (reviewed in [145]), and reduced organic sulfur compounds have also received attention in marine holobiont systems [146–148]. Recently, metagenomic observations suggest that marine sponge symbionts have the capacity for sulfur oxidation alongside the utilization of taurine [149–153]. Although metabolic exchange between marine bacteria and eukaryotic phytoplankton involving taurine has been convincingly suggested [154–157], to date we have no empirical evidence for the utilisation of host derived taurine by marine symbionts and of its role in mediating holobiont function, despite the abundance and widespread distribution of this compound in marine animals.

Finally, in Chapter V, additional metagenomic sequencing and metaproteomics of more *I. basta* individuals attempt to further shed light on the metabolic potential of *I. basta's* microbiome and virome, specifically by focusing on the dominant α-proteobacterium and other low abundant microbial denizens, while also including the genomic contributions of viruses. The data and analyses provided in this chapter, thus provide the opportunity to more firmly establish the full genomic and metabolic potential of the *I. basta* holobiont and offer a firm theoretical framework from which further targeted ecophysiological studies can be undertaken.

References

1. Martin W, Russell MJ. On the origin of biochemistry at an alkaline hydrothermal vent. *Phil Trans R Soc B* 2007; **362**: 1887–1926.
2. Lane N, Martin WF. The Origin of Membrane Bioenergetics. *Cell* 2012; **151**: 1406–1416.
3. Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-Sathi S, et al. The physiology and habitat of the last universal common ancestor. *Nat Microbiol* 2016; **1**: 16116.
4. Hayes JM. Early life on Earth. *Nobel symposium*. 1994. pp 220–36.
5. Arndt NT, Nisbet EG. Processes on the Young Earth and the Habitats of Early Life. *Annu Rev Earth Planet Sci* 2012; **40**: 521–549.
6. Morris BEL, Henneberger R, Huber H, Moissl-Eichinger C. Microbial syntrophy: interaction for the common good. *FEMS Microbiol Rev* 2013; **37**: 384–406.
7. Moreira D, López-García P. Symbiosis Between Methanogenic Archaea and δ-Proteobacteria as the Origin of Eukaryotes: The Syntrophic Hypothesis. *J Mol Evol* 1998; **47**: 517–530.
8. López-García P, Moreira D. Metabolic symbiosis at the origin of eukaryotes. *Trends in Biochemical Sciences* 1999; **24**: 88–93.
9. Cox CJ, Foster PG, Hirt RP, Harris SR, Embley TM. The archaebacterial origin of eukaryotes. *Proc Natl Acad Sci USA* 2008; **105**: 20356–20361.
10. Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 2015; **521**: 173–179.
11. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström D, Juzokaite L, Vancaester E, et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 2017; **541**: 353–358.
12. Williams TA, Embley TM. Changing ideas about eukaryotic origins. *Phil Trans R Soc B* 2015; **370**: 20140318.
13. Williams TA, Cox CJ, Foster PG, Szöllősi GJ, Embley TM. Phylogenomics provides robust support for a two-domains tree of life. *Nat Ecol Evol* 2020; **4**: 138–147.
14. Koonin EV. Archaeal ancestors of eukaryotes: not so elusive any more. *BMC Biol* 2015; **13**: 84.
15. Sousa FL, Neukirchen S, Allen JF, Lane N, Martin WF. Lokiarchaeon is hydrogen dependent. *Nat Microbiol* 2016; **1**: 16034.
16. Spang A, Stairs CW, Dombrowski N, Eme L, Lombard J, Caceres EF, et al. Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. *Nat Microbiol* 2019; **4**: 1138–1148.
17. Imachi H, Nobu MK, Nakahara N, Morono Y, Ogawara M, Takaki Y, et al. Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature* 2020; **577**: 519–525.

18. López-García P, Moreira D. The Syntrophy hypothesis for the origin of eukaryotes revisited. *Nat Microbiol* 2020; **5**: 655–667.

19. Javaux EJ. Challenges in evidencing the earliest traces of life. *Nature* 2019; **572**: 451–460.

20. Lyons TW, Reinhard CT, Planavsky NJ. The rise of oxygen in Earth's early ocean and atmosphere. *Nature* 2014; **506**: 307–315.

21. Knoll AH, Bergmann KD, Strauss JV. Life: the first two billion years. *Phil Trans R Soc B* 2016; **371**: 20150493.

22. Butterfield NJ. Modes of pre-Ediacaran multicellularity. *Precambrian Research* 2009; **173**: 201–211.

23. Becker B. Snow ball earth and the split of Streptophyta and Chlorophyta. *Trends in Plant Science* 2013; **18**: 180–183.

24. Laurin-Lemay S, Brinkmann H, Philippe H. Origin of land plants revisited in the light of sequence contamination and missing data. *Current Biology* 2012; **22**: R593–R594.

25. Love GD, Grosjean E, Stalvies C, Fike DA, Grotzinger JP, Bradley AS, et al. Fossil steroids record the appearance of Demospongiae during the Cryogenian period. *Nature* 2009; **457**: 718–721.

26. Maloof AC, Rose CV, Beach R, Samuels BM, Calmet CC, Erwin DH, et al. Possible animal-body fossils in pre-Marinoan limestones from South Australia. *Nature Geosci* 2010; **3**: 653–659.

27. Erwin DH, Laflamme M, Tweedt SM, Sperling EA, Pisani D, Peterson KJ. The Cambrian Conundrum: Early Divergence and Later Ecological Success in the Early History of Animals. *Science* 2011; **334**: 1091–1097.

28. Whelan NV, Kocot KM, Moroz LL, Halanych KM. Error, signal, and the placement of Ctenophora sister to all other animals. *Proc Natl Acad Sci USA* 2015; **112**: 5773–5778.

29. Ryan JF, Pang K, Schnitzler CE, Nguyen A-D, Moreland RT, Simmons DK, et al. The Genome of the Ctenophore Mnemiopsis leidyi and Its Implications for Cell Type Evolution. *Science* 2013; **342**: 1242592–1242592.

30. Simion P, Philippe H, Baurain D, Jager M, Richter DJ, Di Franco A, et al. A Large and Consistent Phylogenomic Dataset Supports Sponges as the Sister Group to All Other Animals. *Current Biology* 2017; **27**: 958–967.

31. Kapli P, Telford MJ. Topology-dependent asymmetry in systematic errors affects phylogenetic placement of Ctenophora and Xenacoelomorpha. *Sci Adv* 2020; **6**: eabc5162.

32. Costa-Paiva E, Mello B, Bezerra B, Coates CJ, Halanych KM, Brown F, et al. Molecular dating of the blood pigment hemocyanin provides new insight into the origin of animals. *Geobiology* 2022; **20**: 333–345.

33. Turner EC. Possible poriferan body fossils in early Neoproterozoic microbial reefs. *Nature* 2021; **596**: 87–91.

34. King N. The Unicellular Ancestry of Animal Development. *Developmental Cell* 2004; **7**: 313–325.

35. Mah JL, Christensen-Dalsgaard KK, Leys SP. Choanoflagellate and choanocyte collar-flagellar systems and the assumption of homology: Collar-flagellar systems: a comparison. *Evolution & Development* 2014; **16**: 25–37.

36. Shalchian-Tabrizi K, Minge MA, Espelund M, Orr R, Ruden T, Jakobsen KS, et al. Multigene Phylogeny of Choanozoa and the Origin of Animals. *PLoS ONE* 2008; **3**: e2098.

37. Torruella G, de Mendoza A, Grau-Bové X, Antó M, Chaplin MA, del Campo J, et al. Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi. *Current Biology* 2015; **25**: 2404–2410.

38. Dayel MJ, Alegado RA, Fairclough SR, Levin TC, Nichols SA, McDonald K, et al. Cell differentiation and morphogenesis in the colony-forming choanoflagellate Salpingoeca rosetta. *Developmental Biology* 2011; **357**: 73–82.

39. Sebé-Pedrós A, Degnan BM, Ruiz-Trillo I. The origin of Metazoa: a unicellular perspective. *Nat Rev Genet* 2017; **18**: 498–512.

40. Hake KH, West PT, McDonald K, Laundon D, Garcia De Las Bayonas A, Feng C, et al. Colonial choanoflagellate isolated from Mono Lake harbors a microbiome. 2021. Microbiology.

41. Woznica A, Gerdt JP, Hulett RE, Clardy J, King N. Mating in the Closest Living Relatives of Animals Is Induced by a Bacterial Chondroitinase. *Cell* 2017; **170**: 1175-1183.e11.

42. Woznica A, Cantley AM, Beemelmanns C, Freinkman E, Clardy J, King N. Bacterial lipids activate, synergize, and inhibit a developmental switch in choanoflagellates. *Proc Natl Acad Sci USA* 2016; **113**: 7894–7899.

43. Alegado RA, Brown LW, Cao S, Dermenjian RK, Zuzow R, Fairclough SR, et al. A bacterial sulfonolipid triggers multicellular development in the closest living relatives of animals. *eLife* 2012; **1**: e00013.

44. Webster NS, Thomas T. The Sponge Hologenome. *mBio* 2016; **7**.

45. Aristotle. Historia animalium, History of animals. 1965. Harvard University Press, Cambridge, Massachusetts, USA.

46. Brusca RC, Brusca GJ, Haver NJ. Invertebrates. 1990. Sinauer Associates Sunderland, MA.

47. Hooper JNA, Van Soest RWM, Debrenne F. Phylum Porifera Grant, 1836. In: Hooper JNA, Van Soest RWM, Willenz P (eds). *Systema Porifera*. 2002. Springer US, Boston, MA, pp 9–13.

48. Haeckel E. Die Gastraea-theorie, die phylogenetische classification des thierreichs und die homologie der keimblatter. *Jenaische Z f Naturwiss* 1874; **8**: 1–55.

49. Hashimshony T, Feder M, Levin M, Hall BK, Yanai I. Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer. *Nature* 2015; **519**: 219–222.

50. Arendt D, Benito-Gutierrez E, Brunet T, Marlow H. Gastric pouches and the mucociliary sole: setting the stage for nervous system evolution. *Phil Trans R Soc B* 2015; **370**: 20150286.

51. Richter DJ, King N. The Genomic and Cellular Foundations of Animal Origins. *Annu Rev Genet* 2013; **47**: 509–537.

52. Egerton FN. History of Ecological Sciences, Part 52: Symbiosis Studies. *Bulletin of the Ecological Society of America* 2015; **96**: 80–139.

53. Bary A de. Die Erscheinung der Symbiose: Vortrag. 1879. De Gruyter.

54. Vacelet J, Donadey C. Electron microscope study of the association between some sponges and bacteria. *Journal of Experimental Marine Biology and Ecology* 1977; **30**: 301–314.

55. Reiswig HM. Particle feeding in natural populations of three marine demosponges. *The Biological Bulletin* 1971; **141**: 568–591.

56. Reiswig HM. Bacteria as food for temperate-water marine sponges. *Can J Zool* 1975; **53**: 582–589.

57. Wilkinson CR. Microbial associations in sponges. I. Ecology, physiology and microbial populations of coral reef sponges. *Mar Biol* 1978; **49**: 161–167.

58. Wilkinson CR, Fay P. Nitrogen fixation in coral reef sponges with symbiotic cyanobacteria. *Nature* 1979; **279**: 527–529.

59. Wilkinson CR, Vacelet J. Transplantation of marine sponges to different conditions of light and current. *Journal of Experimental Marine Biology and Ecology* 1979; **37**: 91–104.

60. Wilkinson CR, Nowak M, Austin B, Colwell RR. Specificity of bacterial symbionts in Mediterranean and Great Barrier Reef sponges. *Microb Ecol* 1981; **7**: 13–21.

61. Wilkinson CR, Garrone R, Vacelet J. Marine sponges discriminate between food bacteria and bacterial symbionts: electron microscope radioautography and in situ evidence. *Proceedings of the Royal society of London Series B Biological sciences* 1984; **220**: 519–528.

62. Wilkinson CR. Net Primary Productivity in Coral Reef Sponges. *Science* 1983; **219**: 410–412.

63. Wilkinson CR. Nutrient translocation from symbiontic cyanobacteria to coral reef sponges. In: Levi C, Boury-Esnault N (eds). *Biologie des spongiaires*. 1979. Colloques Internationaux du Centre National de la Recherché Scientifique, Paris, pp 373–380.

64. Vacelet J, Boury-Esnault N, Fiala-Medioni A, Fisher CR. A methanotrophic carnivorous sponge. *Nature* 1995; **377**: 296–296.

65. Southwell MW, Popp BN, Martens CS. Nitrification controls on fluxes and isotopic composition of nitrate from Florida Keys sponges. *Marine Chemistry* 2008; **108**: 96–108.

66. Bayer K, Schmitt S, Hentschel U. Physiology, phylogeny and *in situ* evidence for bacterial and archaeal nitrifiers in the marine sponge *Aplysina aerophoba*. *Environmental Microbiology* 2008; **10**: 2942–2955.

67. Schläppy M-L, Schöttner SI, Lavik G, Kuypers MMM, de Beer D, Hoffmann F. Evidence of nitrification and denitrification in high and low microbial abundance sponges. *Marine Biology* 2010; **157**: 593–602.

68. Hoffmann F, Radax R, Woebken D, Holtappels M, Lavik G, Rapp HT, et al. Complex nitrogen cycling in the sponge *Geodia barretti*. *Environmental Microbiology* 2009; **11**: 2228–2243.

69. Hoffmann F, Larsen O, Thiel V, Rapp HT, Pape T, Michaelis W, et al. An Anaerobic World in Sponges. *Geomicrobiology Journal* 2005; **22**: 1–10.

70. Webster NS, Taylor MW, Behnam F, Lücker S, Rattei T, Whalan S, et al. Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environmental Microbiology* 2009.

71. Schmitt S, Tsai P, Bell J, Fromont J, Ilan M, Lindquist N, et al. Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. *The ISME Journal* 2012; **6**: 564–576.

72. Reveillaud J, Maignien L, Eren AM, Huber JA, Apprill A, Sogin ML, et al. Host-specificity among abundant and rare taxa in the sponge microbiome. *The ISME Journal* 2014; **8**: 1198–1209.

73. Hentschel U, Usher KM, Taylor MW. Marine sponges as microbial fermenters: Marine sponges as microbial fermenters. *FEMS Microbiology Ecology* 2006; **55**: 167–177.

74. Lee OO, Wang Y, Yang J, Lafi FF, Al-Suwailem A, Qian P-Y. Pyrosequencing reveals highly diverse and species-specific microbial communities in sponges from the Red Sea. *ISME J* 2011; **5**: 650–664.

75. Simister RL, Deines P, Botté ES, Webster NS, Taylor MW. Sponge-specific clusters revisited: a comprehensive phylogeny of sponge-associated microorganisms: Sponge-associated microorganisms. *Environmental Microbiology* 2012; **14**: 517–524.

76. Taylor MW, Tsai P, Simister RL, Deines P, Botte E, Ericson G, et al. 'Sponge-specific' bacteria are widespread (but rare) in diverse marine environments. *ISME J* 2013; **7**: 438–443.

77. Thomas T, Rusch D, DeMaere MZ, Yung PY, Lewis M, Halpern A, et al. Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. *The ISME Journal* 2010; **4**: 1557–1567.

78. Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, Webster NS, et al. Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proceedings of the National Academy of Sciences* 2012; **109**: E1878–E1887.

79. Horn H, Slaby BM, Jahn MT, Bayer K, Moitinho-Silva L, Förster F, et al. An Enrichment of CRISPR and Other Defense-Related Features in Marine Sponge-Associated Microbial Metagenomes. *Frontiers in Microbiology* 2016; **7**.

80. Slaby BM, Hackl T, Horn H, Bayer K, Hentschel U. Metagenomic binning of a marine sponge microbiome reveals unity in defense but metabolic specialization. *ISME J* 2017; **11**: 2465–2478.

81. Hentschel U, Piel J, Degnan SM, Taylor MW. Genomic insights into the marine sponge microbiome. *Nature Reviews Microbiology* 2012; **10**: 641–654.

82. Blunt JW, Copp BR, Hu W-P, Munro MHG, Northcote PT, Prinsep MR. Marine natural products. *Nat Prod Rep* 2008; **25**: 35.

83. Fisch KM, Gurgui C, Heycke N, van der Sar SA, Anderson SA, Webb VL, et al. Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. *Nat Chem Biol* 2009; **5**: 494–501.

84. Piel J, Hui D, Wen G, Butzke D, Platzer M, Fusetani N, et al. Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proc Natl Acad Sci USA* 2004; **101**: 16222–16227.

85. Bewley CA, Faulkner DJ. Lithistid Sponges: Star Performers or Hosts to the Stars. *Angewandte Chemie International Edition* 1998; **37**: 2162–2178.

86. Sudek S, Lopanik NB, Waggoner LE, Hildebrand M, Anderson C, Liu H, et al. Identification of the Putative Bryostatin Polyketide Synthase Gene Cluster from " *Candidatus* Endobugula sertula", the Uncultivated Microbial Symbiont of the Marine Bryozoan *Bugula neritina*. *J Nat Prod* 2007; **70**: 67–74.

87. Partida-Martinez LP, Hertweck C. Pathogenic fungus harbours endosymbiotic bacteria for toxin production. *Nature* 2005; **437**: 884–888.

88. Goeij JM de, Berg H van den, Oostveen MM van, Epping EHG, Duyl FC van. Major bulk dissolved organic carbon (DOC) removal by encrusting coral reef cavity sponges. *Marine Ecology Progress Series* 2008; **357**: 139–151.

89. Yahel G, Sharp JH, Marie D, Häse C, Genin A. In situ feeding and element removal in the symbiont-bearing sponge *Theonella swinhoei* : Bulk DOC is the major source for carbon. *Limnol Oceanogr* 2003; **48**: 141–149.

90. Mueller B, de Goeij JM, Vermeij MJA, Mulders Y, van der Ent E, Ribes M, et al. Natural Diet of Coral-Excavating Sponges Consists Mainly of Dissolved Organic Carbon (DOC). *PLoS ONE* 2014; **9**: e90152.

91. Alexander BE, Liebrand K, Osinga R, van der Geest HG, Admiraal W, Cleutjens JPM, et al. Cell Turnover and Detritus Production in Marine Sponges from Tropical and Temperate Benthic Ecosystems. *PLoS ONE* 2014; **9**: e109486.

92. De Goeij JM, De Kluijver A, Van Duyl FC, Vacelet J, Wijffels RH, De Goeij AFPM, et al. Cell kinetics of the marine sponge Halisarca caerulea reveal rapid cell turnover and shedding. *Journal of Experimental Biology* 2009; **212**: 3892–3900.

93. Reiswig HM. Water transport, respiration and energetics of three tropical marine sponges. *Journal of Experimental Marine Biology and Ecology* 1974; **14**: 231–249.

94. Weisz JB, Lindquist N, Martens CS. Do associated microbial abundances impact marine demosponge pumping rates and tissue densities? *Oecologia* 2008; **155**: 367–376.

95. Rix L, de Goeij JM, Mueller CE, Struck U, Middelburg JJ, van Duyl FC, et al. Coral mucus fuels the sponge loop in warm- and cold-water coral reef ecosystems. *Scientific Reports* 2016; **6**: 18715.

96. Rix L, de Goeij JM, van Oevelen D, Struck U, Al-Horani FA, Wild C, et al. Differential recycling of coral and algal dissolved organic matter via the sponge loop. *Functional Ecology* 2017; **31**: 778–789.

97. Goeij JM de, Moodley L, Houtekamer M, Carballeira NM, Duyl FC van. Tracing 13C-enriched dissolved and particulate organic carbon in the bacteria-containing coral reef

sponge Halisarca caerulea: Evidence for DOM-feeding. *Limnology and Oceanography* 2008; **53**: 1376–1386.

98. Kamke J, Sczyrba A, Ivanova N, Schwientek P, Rinke C, Mavromatis K, et al. Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. *ISME J* 2013; **7**: 2287–2300.
99. Astudillo-García C, Slaby BM, Waite DW, Bayer K, Hentschel U, Taylor MW. Phylogeny and genomics of SAUL, an enigmatic bacterial lineage frequently associated with marine sponges: Phylogeny and genomics of SAUL. *Environ Microbiol* 2018; **20**: 561–576.
100. Fiore CL, Labrie M, Jarett JK, Lesser MP. Transcriptional activity of the giant barrel sponge, Xestospongia muta Holobiont: molecular evidence for metabolic interchange. *Frontiers in Microbiology* 2015; **6**.
101. Lackner G, Peters EE, Helfrich EJN, Piel J. Insights into the lifestyle of uncultured bacterial natural product factories associated with marine sponges. *PNAS* 2017; **114**: E347–E356.
102. McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, et al. Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci USA* 2013; **110**: 3229–3236.
103. Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman DA. The Application of Ecological Theory Toward an Understanding of the Human Microbiome. *Science* 2012; **336**: 1255–1262.
104. Koskella B, Hall LJ, Metcalf CJE. The microbiome beyond the horizon of ecological and evolutionary theory. *Nat Ecol Evol* 2017; **1**: 1606–1615.
105. Coyte KZ, Schluter J, Foster KR. The ecology of the microbiome: Networks, competition, and stability. *Science* 2015; **350**: 663–666.
106. Cardini U, Bednarz VN, van Hoytema N, Rovere A, Naumann MS, Al-Rshaidat MMD, et al. Budget of Primary Production and Dinitrogen Fixation in a Highly Seasonal Red Sea Coral Reef. *Ecosystems* 2016; **19**: 771–785.
107. Vacelet J, Fiala-Médioni A, Fisher C, Boury-Esnault N. Symbiosis between methane-oxidizing bacteria and a deep-sea carnivorous cladorhizid sponge. *Mar Ecol Prog Ser* 1996; **145**: 77–85.
108. van Duyl F, Hegeman J, Hoogstraten A, Maier C. Dissolved carbon fixation by sponge–microbe consortia of deep water coral mounds in the northeastern Atlantic Ocean. *Mar Ecol Prog Ser* 2008; **358**: 137–150.
109. Flórez LV, Biedermann PHW, Engl T, Kaltenpoth M. Defensive symbioses of animals with prokaryotic and eukaryotic microorganisms. *Nat Prod Rep* 2015; **32**: 904–936.
110. Pawlik JR. Coral reef sponges: Do predatory fishes affect their distribution? *Limnol Oceanogr* 1998; **43**: 1396–1399.
111. Wulff J. Bottom-up and top-down controls on coral reef sponges: disentangling within-habitat and between-habitat processes. *Ecology* 2017; **98**: 1130–1139.

112. de Goeij JM, van Oevelen D, Vermeij MJA, Osinga R, Middelburg JJ, de Goeij AFPM, et al. Surviving in a Marine Desert: The Sponge Loop Retains Resources Within Coral Reefs. *Science* 2013; **342**: 108–110.

113. Liu M, Fan L, Zhong L, Kjelleberg S, Thomas T. Metaproteogenomic analysis of a community of sponge symbionts. *The ISME Journal* 2012; **6**: 1515–1525.

114. Moitinho-Silva L, Díez-Vives C, Batani G, Esteves AI, Jahn MT, Thomas T. Integrated metabolism in sponge–microbe symbiosis revealed by genome-centered metatranscriptomics. *The ISME Journal* 2017; **11**: 1651–1666.

115. Radax R, Rattei T, Lanzen A, Bayer C, Rapp HT, Urich T, et al. Metatranscriptomics of the marine sponge Geodia barretti: tackling phylogeny and function of its microbial community: Metatranscriptomics in Geodia barretti. *Environmental Microbiology* 2012; **14**: 1308–1324.

116. Jahn MT, Markert SM, Ryu T, Ravasi T, Stigloher C, Hentschel U, et al. Shedding light on cell compartmentation in the candidate phylum Poribacteria by high resolution visualisation and transcriptional profiling. *Sci Rep* 2016; **6**: 35860.

117. Campana S, Riesgo A, Jongepier E, Fuss J, Muyzer G, de Goeij JM. Meta-transcriptomic comparison of two sponge holobionts feeding on coral- and macroalgal-dissolved organic matter. *BMC Genomics* 2022; **23**: 674.

118. Achlatis M, Pernice M, Green K, de Goeij JM, Guagliardo P, Kilburn MR, et al. Single-cell visualization indicates direct role of sponge host in uptake of dissolved organic matter. *Proceedings of the Royal Society B: Biological Sciences* 2019; **286**: 20192153.

119. Hudspith M, Rix L, Achlatis M, Bougoure J, Guagliardo P, Clode PL, et al. Subcellular view of host–microbiome nutrient exchange in sponges: insights into the ecological success of an early metazoan–microbe symbiosis. *Microbiome* 2021; **9**: 44.

120. Rix L, Ribes M, Coma R, Jahn MT, de Goeij JM, van Oevelen D, et al. Heterotrophy in the earliest gut: a single-cell view of heterotrophic carbon and nitrogen assimilation in sponge-microbe symbioses. *ISME J* 2020; **14**: 2554–2567.

121. Bergquist PR, Kelly-Borges M. Systematics and Biogeography of the Genus Ianthella (Demospongiae: Verongida: Ianthellidae) in the south-west Pacific. *Beagle: Records of the Museums and Art Galleries of the Northern Territory, The* 1995; **12**: 151.

122. Freckelton ML, Luter HM, Andreakis N, Webster NS, Motti CA. Qualitative variation in colour morphotypes of Ianthella basta (Porifera: Verongida). *Hydrobiologia* 2012; **687**: 191–203.

123. Luter HM, Whalan S, Webster NS. Exploring the Role of Microorganisms in the Disease-Like Syndrome Affecting the Sponge Ianthella basta. *Applied and Environmental Microbiology* 2010; **76**: 5736–5744.

124. Luter HM, Whalan S, Webster NS. Thermal and Sedimentation Stress Are Unlikely Causes of Brown Spot Syndrome in the Coral Reef Sponge, Ianthella basta. *PLoS ONE* 2012; **7**: e39779.

125. Preston CM, Wu KY, Molinski TF, DeLong EF. A psychrophilic crenarchaeon inhabits a marine sponge: Cenarchaeum symbiosum gen. nov., sp. nov. *Proc Natl Acad Sci USA* 1996; **93**: 6241–6246.

126. Koskinen K, Pausan MR, Perras AK, Beck M, Bang C, Mora M, et al. First Insights into the Diverse Human Archaeome: Specific Detection of Archaea in the Gastrointestinal Tract, Lung, and Nose and on Skin. *mBio* 2017; **8**: e00824-17.

127. Moissl-Eichinger C, Probst AJ, Birarda G, Auerbach A, Koskinen K, Wolf P, et al. Human age and skin physiology shape diversity and abundance of Archaea on skin. *Sci Rep* 2017; **7**: 4039.

128. Umbach AK, Stegelmeier AA, Neufeld JD. Archaea Are Rare and Uncommon Members of the Mammalian Skin Microbiome. *mSystems* 2021; **6**: e00642-21.

129. Dror H, Novak L, Evans JS, López-Legentil S, Shenkar N. Core and Dynamic Microbial Communities of Two Invasive Ascidians: Can Host–Symbiont Dynamics Plasticity Affect Invasion Capacity? *Microb Ecol* 2019; **78**: 170–184.

130. Erwin PM, Pineda MC, Webster N, Turon X, López-Legentil S. Down under the tunic: bacterial biodiversity hotspots and widespread ammonia-oxidizing archaea in coral reef ascidians. *ISME J* 2014; **8**: 575–588.

131. Martínez-García M, Stief P, Díaz-Valdés M, Wanner G, Ramos-Esplá A, Dubilier N, et al. Ammonia-oxidizing *Crenarchaeota* and nitrification inside the tissue of a colonial ascidian. *Environmental Microbiology* 2008; **10**: 2991–3001.

132. Steger D, Ettinger-Epstein P, Whalan S, Hentschel U, de Nys R, Wagner M, et al. Diversity and mode of transmission of ammonia-oxidizing archaea in marine sponges. *Environmental Microbiology* 2008; **10**: 1087–1094.

133. Sipkema D, de Caralt S, Morillo JA, Al-Soud WA, Sørensen SJ, Smidt H, et al. Similar sponge-associated bacteria can be acquired via both vertical and horizontal transmission: Microbial transmission in *Petrosia ficiformis*. *Environ Microbiol* 2015; **17**: 3807–3821.

134. Pester M, Schleper C, Wagner M. The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. *Current Opinion in Microbiology* 2011; **14**: 300–306.

135. Baker BJ, De Anda V, Seitz KW, Dombrowski N, Santoro AE, Lloyd KG. Diversity, ecology and evolution of Archaea. *Nat Microbiol* 2020; **5**: 887–900.

136. Mohamed NM, Saito K, Tal Y, Hill RT. Diversity of aerobic and anaerobic ammonia-oxidizing bacteria in marine sponges. *The ISME Journal* 2010; **4**: 38–48.

137. Mussmann M, Brito I, Pitcher A, Sinninghe Damste JS, Hatzenpichler R, Richter A, et al. Thaumarchaeotes abundant in refinery nitrifying sludges express amoA but are not obligate autotrophic ammonia oxidizers. *Proceedings of the National Academy of Sciences* 2011; **108**: 16771–16776.

138. Palatinszky M, Herbold C, Jehmlich N, Pogoda M, Han P, von Bergen M, et al. Cyanate as an energy source for nitrifiers. *Nature* 2015; **524**: 105–108.

139. Kitzinger K, Padilla CC, Marchant HK, Hach PF, Herbold CW, Kidane AT, et al. Cyanate and urea are substrates for nitrification by Thaumarchaeota in the marine environment. *Nature Microbiology* 2019; **4**: 234–243.

140. Dekas AE, Parada AE, Mayali X, Fuhrman JA, Wollard J, Weber PK, et al. Characterizing Chemoautotrophy and Heterotrophy in Marine Archaea and Bacteria With Single-Cell Multi-isotope NanoSIP. *Front Microbiol* 2019; **10**: 2682.

141. Huxtable RJ. Physiological actions of taurine. *Physiological Reviews* 1992; **72**: 101–163.

142. Lee R, Childress J, Desaulniers N. The effects of exposure to ammonia on ammonia and taurine pools of the symbiotic clam. *Journal of Experimental Biology* 1997; **200**: 2797–2805.

143. Hooper JNA, Capon RJ, Keenan CP, Parry DL, Smit N. Chemotaxonomy of marine sponges: Families Microcionidae, Raspailiidae and Axinellidae, and their relationships with other families in the orders Poecilosclerida and Axinellida (Porifera : Demospongiae). *Invert Systematics* 1992; **6**: 261–301.

144. Shinagawa A, Suzuki T, Konosu S. The Role of Free Amino Acids and Betaines in Intracellular Osmoregulation of Marine Sponges. *Nippon Suisan Gakkaishi* 1992; **58**: 1717–1722.

145. Dubilier N, Bergin C, Lott C. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nat Rev Microbiol* 2008; **6**: 725–740.

146. Raina J-B, Tapiolas D, Willis BL, Bourne DG. Coral-Associated Bacteria and Their Role in the Biogeochemical Cycling of Sulfur. *Appl Environ Microbiol* 2009; **75**: 3492–3501.

147. Raina J-B, Tapiolas DM, Forêt S, Lutz A, Abrego D, Ceh J, et al. DMSP biosynthesis by an animal and its role in coral thermal stress response. *Nature* 2013; **502**: 677–680.

148. Kleiner M, Wentrup C, Lott C, Teeling H, Wetzel S, Young J, et al. Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *PNAS* 2012; **109**: E1173–E1182.

149. Tian R-M, Wang Y, Bougouffa S, Gao Z-M, Cai L, Bajic V, et al. Genomic analysis reveals versatile heterotrophic capacity of a potentially symbiotic sulfur-oxidizing bacterium in sponge: A versatile SOB as a potential sponge symbiont. *Environmental Microbiology* 2014; **16**: 3548–3561.

150. Gauthier M-EA, Watson JR, Degnan SM. Draft Genomes Shed Light on the Dual Bacterial Symbiosis that Dominates the Microbiome of the Coral Reef Sponge Amphimedon queenslandica. *Frontiers in Marine Science* 2016; **3**.

151. Karimi E, Keller-Costa T, Slaby BM, Cox CJ, da Rocha UN, Hentschel U, et al. Genomic blueprints of sponge-prokaryote symbiosis are shared by low abundant and cultivatable Alphaproteobacteria. *Scientific Reports* 2019; **9**: 1999.

152. Engelberts JP, Robbins SJ, de Goeij JM, Aranda M, Bell SC, Webster NS. Characterization of a sponge microbiome using an integrative genome-centric approach. *The ISME Journal* 2020; **14**: 1100–1110.

153. Botté ES, Nielsen S, Abdul Wahab MA, Webster J, Robbins S, Thomas T, et al. Changes in the metabolic potential of the sponge microbiome under ocean acidification. *Nature Communications* 2019; **10**: 4134.

154. Amin SA, Hmelo LR, van Tol HM, Durham BP, Carlson LT, Heal KR, et al. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* 2015; **522**: 98–101.

155. Durham BP, Sharma S, Luo H, Smith CB, Amin SA, Bender SJ, et al. Cryptic carbon and sulfur cycling between surface ocean plankton. *Proc Natl Acad Sci USA* 2015; **112**: 453–457.

156. Durham BP, Boysen AK, Carlson LT, Groussman RD, Heal KR, Cain KR, et al. Sulfonate-based networks between eukaryotic phytoplankton and heterotrophic bacteria in the surface ocean. *Nat Microbiol* 2019; **4**: 1706–1715.

157. Landa M, Burns AS, Roth SJ, Moran MA. Bacterial transcriptome remodeling during sequential co-culture with a marine dinoflagellate and diatom. *ISME J* 2017; **11**: 2677–2690.

CHAPTER II

Overview of

publications and manuscripts

Chapter III

Mansucript title:

Characterization of a thaumarchaeal symbiont that drives incomplete nitrification in the tropical sponge *Ianthella basta*

Author names:

Florian U. Moeller, Nicole S. Webster, Craig W. Herbold, Faris Behnam, Daryl Domman, Mads Albertsen, Maria Mooshammer, Stephanie Markert, Dmitrij Turaev, Dörte Becher, Thomas Rattei, Thomas Schweder, Andreas Richter, Margarete Watzka, Per Halkjaer Nielsen, Michael Wagner

Reference:

Moeller FU, Webster NS, Herbold CW, Behnam F, Domman D, Albertsen M, et al. Characterization of a thaumarchaeal symbiont that drives incomplete nitrification in the tropical sponge Ianthella basta. *Environmental Microbiology* 2019; **21**: 3831–3854.

Author contributions:

Conceptualization — F.U.M., N.S.W., M.A., T.R., T.S., P.H.N., M.W.
Sample collection — F.U.M., F.B., N.S.W.
Data collection and analysis — F.U.M., C.W.H., F.B., D.D., M.A., M.M., S.M., D.T., D.B., T.R., A.R., M.W.
Writing — F.U.M., N.S.W., M.W.
Supervision — M.W., N.S.W.
Funding — M.W.

Chapter IV

Mansucript title:

Taurine as a key intermediate for host-symbiont interaction in the tropical sponge *Ianthella basta*

Author names:

Florian U. Moeller, Craig W. Herbold, Arno Schintlmeister, Maria Mooshammer, Cherie Motti, Faris Behnam, Margarete Watzka, Thomas Schweder, Mads Albertsen, Andreas Richter, Nicole S. Webster, and Michael Wagner

Reference:

Currently under review after resubmission and minor revisions since 12 February 2023. Submitted as a preprint to bioRxiv.

Moeller FU, Herbold CW, Schintlmeister A, Mooshammer M, Motti C, Behnam F, et al. Taurine as a key intermediate for host-symbiont interaction in the tropical sponge Ianthella basta. *bioRxiv* (2022): 2022-09.

Author contributions:

Conceptualization — F.U.M., N.S.W., M.W.
Sample collection — F.U.M., F.B., N.S.W.
Data collection and analysis — F.U.M., C.W.H., A.S., C.M., F.B., M.W., T.S., M.A., A.R., N.S.W.
Writing — F.U.M., N.S.W., M.W.

Supervision — M.W., N.S.W.
Funding — M.W.

Chapter V

Mansucript title:

Metabolic reconstruction of the near complete microbiome of the model sponge *Ianthella basta*

Author names:

Joan Pamela Engelberts, Steven J. Robbins, Craig W. Herbold, **Florian U. Moeller**, Nico Jehmlich, Patrick W. Laffy, Michael Wagner, Nicole S. Webster

Reference:

Engelberts JP, Robbins SJ, Herbold CW, **Moeller FU**, Jehmlich N, Laffy PW, et al. Metabolic reconstruction of the near complete microbiome of the model sponge *Ianthella basta*. *Environmental Microbiology* 2022; 1462-2920.16302.

Author contributions:

Conceptualization — J.P.E., F.U.M., N.S.W., M.W.
Sample collection — F.U.M., F.B., N.S.W., J.P.E
Data collection and analysis — F.U.M., J.P.E., S.J.R., C.W.H., N.J., P.L.
Writing — J.P.E., F.U.M., N.S.W., M.W.
Supervision — M.W., N.S.W.
Funding — N.W., M.W.

CHAPTER III

Characterization of a thaumarchaeal symbiont that drives incomplete nitrification in the tropical sponge *Ianthella basta*

Authors:

Florian U. Moeller, Nicole S. Webster, Craig W. Herbold, Faris Behnam, Daryl Domman, Mads Albertsen, Maria Mooshammer, Stephanie Markert, Dmitrij Turaev, Dörte Becher, Thomas Rattei, Thomas Schweder, Andreas Richter, Margarete Watzka, Per Halkjaer Nielsen, Michael Wagner

Published as an Open Access publication in *Environmental Microbiology* (2019; https://doi.org/10.1111/1462-2920.14732).

environmental microbiology

Environmental Microbiology (2019) **21**(10), 3831–3854 doi:10.1111/1462-2920.14732

Characterization of a thaumarchaeal symbiont that drives incomplete nitrification in the tropical sponge *Ianthella basta*

Florian U. Moeller,[1] **Nicole S. Webster**,[2,3] **Craig W. Herbold**,[1] **Faris Behnam**,[1] **Daryl Domman**,[1] **Mads Albertsen**,[4] **Maria Mooshammer**,[1] **Stephanie Markert**,[5,8] **Dmitrij Turaev**,[6] **Dörte Becher**,[7] **Thomas Rattei**,[6] **Thomas Schweder**,[5,8] **Andreas Richter**,[9] **Margarete Watzka**,[9] **Per Halkjaer Nielsen**[4] **and Michael Wagner**[1,4*]

[1]*Centre for Microbiology and Environmental Systems Science, Division of Microbial Ecology, University of Vienna, Austria.*

[2]*Australian Institute of Marine Science, Townsville, Queensland, Australia.*

[3]*Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, University of Queensland, St Lucia, Queensland, Australia.*

[4]*Center for Microbial Communities, Department of Chemistry and Bioscience, Aalborg University, 9220, Aalborg, Denmark.*

[5]*Institute of Marine Biotechnology e.V, Greifswald, Germany.*

[6]*Centre for Microbiology and Environmental Systems Science, Division of Computational Systems Biology, University of Vienna, Austria.*

[7]*Institute of Microbiology, Microbial Proteomics, University of Greifswald, Greifswald, Germany.*

[8]*Institute of Pharmacy, Pharmaceutical Biotechnology, University of Greifswald, Greifswald, Germany.*

[9]*Centre for Microbiology and Environmental Systems Science, Division of Terrestrial Ecosystem Research, University of Vienna, Austria.*

Summary

Marine sponges represent one of the few eukaryotic groups that frequently harbour symbiotic members of the *Thaumarchaeota*, which are important chemoautotrophic ammonia-oxidizers in many environments. However, in most studies, direct demonstration of ammonia-oxidation by these archaea within sponges is lacking, and little is known about sponge-specific adaptations of ammonia-oxidizing archaea (AOA). Here, we characterized the thaumarchaeal symbiont of the marine sponge *Ianthella basta* using metaproteogenomics, fluorescence *in situ* hybridization, qPCR and isotope-based functional assays. '*Candidatus* Nitrosospongia ianthellae' is only distantly related to cultured AOA. It is an abundant symbiont that is solely responsible for nitrite formation from ammonia in *I. basta* that surprisingly does not harbour nitrite-oxidizing microbes. Furthermore, this AOA is equipped with an expanded set of extracellular subtilisin-like proteases, a metalloprotease unique among archaea, as well as a putative branched-chain amino acid ABC transporter. This repertoire is strongly indicative of a mixotrophic lifestyle and is (with slight variations) also found in other sponge-associated, but not in free-living AOA. We predict that this feature as well as an expanded and unique set of secreted serpins (protease inhibitors), a unique array of eukaryotic-like proteins, and a DNA-phosporothioation system, represent important adaptations of AOA to life within these ancient filter-feeding animals.

Introduction

Marine sponges (phylum *Porifera*) are among the most basal metazoan lineages (Simion *et al*., 2017), with fossil records suggesting their evolutionary emergence more than 600 million years ago (Love *et al*., 2009). Sponges form a major part of the marine benthic fauna across the world's oceans (Bell, 2008) where they mediate critical biogeochemical processes through their filtration of immense volumes of seawater (Maldonado *et al*., 2012; de Goeij *et al*., 2013). Many sponges harbour dense, diverse and species-specific communities of microbes (Hentschel *et al*., 2012; Thomas *et al*., 2016), and these associations are often temporally and geographically stable (Luter *et al*., 2010; Schmitt *et al*., 2012; Astudillo-García *et al*., 2017). Functional roles that have been assigned to specific sponge-associated microorganisms include the provision of photosynthates (Wilkinson, 1983) and fixed N_2 (Wilkinson and Fay, 1979)

Received 21 January, 2019; revised 14 June, 2019; accepted 17 June, 2019. *For correspondence. E-mail wagner@microbial-ecology.net; Tel. (+43) 14277 91200; Fax (+43) 14277 825701.

by cyanobacterial symbionts and the production of bioactive secondary metabolites (Wilson *et al*., 2014; Freeman *et al*., 2016; Agarwal *et al*., 2017). Furthermore, sponge symbionts can serve as an endogenous food source as exemplified by biomass transfer from a sponge-associated sulfate-reducing bacterial community to host cells (Hoffmann *et al*., 2005), and the ingestion of symbiotic methanotrophs in a deep-sea carnivorous sponge (Vacelet *et al*., 1995). Whilst a vast array of additional putative symbiotic functions have been hypothesized from taxonomic or metagenomic data (Webster and Thomas, 2016), unequivocal evidence for specific sponge symbiont physiologies is comparably rare.

Nitrogen cycling in sponge holobionts has received considerable attention and sponge-symbiont-driven nitrification, denitrification and anaerobic ammonium oxidation has been described (Bayer *et al*., 2008; Southwell *et al*., 2008; Hoffmann *et al*., 2009; Schläppy *et al*., 2010; Radax *et al*., 2012a). Sponges are marine filter feeders and during decay of organic matter (from food and sponge cells) ammonia is released. Since ammonia is more toxic than nitrite and nitrate, for aquatic animals (Camargo and Alonso, 2006), nitrification that is performed by sponge-associated microbes might be beneficial for the host. In addition, growth of nitrifier symbionts on waste products of the host (NH_4^+ and CO_2) could also be a way of efficiently recycling those into new microbial food biomass. Consistently, many marine sponges harbour symbionts phylogenetically related to nitrifying microbes (Steger *et al*., 2008; Hoffman *et al*., 2009; Off *et al*., 2010), indicating that ammonia oxidation via nitrite to nitrate is a widely distributed process in these animals. More specifically, molecular signatures of proteobacterial as well as thaumarchaeal ammonia oxidizers are frequently detected in sponges, although these microbes rarely co-occur in the same host (Bayer *et al*., 2008; Radax *et al*., 2012b). Most sponges that contain symbionts related to ammonia-oxidizers also host bacteria affiliated with known nitrite-oxidizers, particularly members of the genus *Nitrospira* (Hoffmann *et al*., 2009; Off *et al*., 2010; Reveillaud *et al*., 2014; Moitinho-Silva *et al*., 2017a). Whilst many studies simply equate the molecular detection of microbes related to recognized nitrifiers with the occurrence of canonical nitrification in sponges (Mohamed *et al*., 2010; Fan *et al*., 2012), this assumption could be misleading due to the functional versatility of both Thaumarchaeota and *Nitrospira* (Mußmann *et al*., 2011; Koch *et al*., 2014; Daims *et al*., 2015; Palatinszky *et al*., 2015).

Thaumarchaeotes are important ammonia oxidizers in many environments (Pester *et al*., 2011). Marine sponges host a particularly high diversity of thaumarchaeotes, with many of the 16S rRNA gene sequences falling into phylogenetic clusters that exclusively contain sponge-derived sequences (Simister *et al*., 2012). The detection of thaumarchaeotes in sponge larvae also suggests vertical transmission or early environmental acquisition of these symbionts (Sharp *et al*., 2007; Steger *et al*., 2008; Schmitt *et al*., 2012). The first genomic information of a member of this phylum was derived from *Ca*. Cenarchaeum symbiosum in the marine sponge *Axinella mexicana* (Schleper *et al*., 1998; Hallam *et al*., 2006), and subsequently thaumarchaeal genomes were recovered from the deep sea glass sponge *Lophophysema eversa* (Tian *et al*., 2016; referred to as DSGS-AOA in this manuscript) and the temperate sponge *Cymbastella concentrica*, with the latter also shown to transcribe genes for ammonia oxidation (referred to as CCThau in Moitinho Silva *et al*., 2017a and in this manuscript). However, little information is available on the genomic plasticity and mechanisms of host adaptation in sponge thaumarchaeal symbionts. Furthermore, with the exception of a thaumarchaeal symbiont that inhabits the cold-water sponge *Geodia barretti* (Radax *et al*., 2012a, b), direct evidence for the catalysis of ammonia oxidation by sponge-associated thaumarchaeotes is lacking. Whilst transcription and translation of key functional genes like the *amoA* gene encoding a subunit of the ammonia monooxygenase of thaumarchaeotes has been detected across multiple sponge species (Liu *et al*., 2012; Fiore *et al*., 2015; Moitinho-Silva *et al*., 2017a), experimental validation of their involvement in nitrification is required to confirm this activity *in situ* (Mußmann *et al*., 2011).

To better understand, the physiological capability of Thaumarchaeota in sponges, we used a metaproteogenomic approach to characterize the Thaumarchaeota symbiont of the marine sponge *Ianthella basta*. *Ianthella basta* is an abundant and ecologically important reef sponge found throughout the Indo-Pacific (Bergquist and Kelly-Borges, 1995). In contrast to many microbially diverse sponge species, 16S rRNA gene surveys have shown that *I. basta* harbours only three dominant microbial phylotypes which belong to a novel α-proteobacterial lineage, a γ-proteobacterial clade within the UBA10353 order and LS-SOB family (Parks *et al*., 2018), and the Thaumarchaeota (Luter *et al*., 2010). This community structure is stable among different host colour morphotypes (Freckelton *et al*., 2012), between individuals sampled from different geographic regions (Luter *et al*., 2010), across different host health states (Luter *et al*., 2010) and in sponges exposed to different environmental stressors (Luter *et al*., 2012). Here we (i) quantify the abundance of the thaumarchaeal symbionts in *I. basta*, (ii) confirm their activity as ammonia oxidizers, (iii) document the expression of almost 100 thaumarchaeal genes *in situ* and (iv) reveal sponge-specific adaptations of these archaea including a putative mixotrophic lifestyle. Furthermore, we demonstrate that nitrite oxidation surprisingly does not occur in *I. basta*.

Results and discussion

Quantification of the thaumarchaeal I. basta *symbiont using FISH and qPCR*

Fluorescence *in situ* hybridization (FISH) was performed on fixed cryosections from one sponge individual using the general archaeal probe Arch915 that is fully complementary to the 16S rRNA of the single archaeal phylotype known to inhabit *I. basta* (Fig. 1). Quantitative FISH across 10 images revealed that the thaumarchaeal symbiont comprised 24% ± 1.6% [standard error (SE)] of the total bacterial and archaeal cells detected with probes Arch915 and the probe set EUB338-I-III targeting most bacteria. Absolute quantification of the thaumarchaeal symbiont in five sponge individuals (including those used for metagenome sequencing) using specifically designed qPCR primers, revealed an average absolute abundance of 2.41 ± 0.7 (SE) × 10^{10} 16S rRNA gene copies per gram wet weight of sponge tissue. Consistent with the relative abundances derived from FISH, the ratio of thaumarchaeal 16S rRNA sequences to the total 16S rRNA genes derived from qPCR assays targeting the additional α- and γ-proteobacterial symbionts of *I. basta* (data not shown) was 22% ± 2.3% (SE). The qPCR experiments were performed with five sponge individuals, and sponge samples within an individual, were randomly selected, rendering it highly unlikely that the abundance pattern of the thaumarchaeal symbiont throughout a sponge varies significantly. Collectively, these data demonstrate that the *I. basta* thaumarchaeote is a dominant member of the *I. basta* microbiome, occurring at densities that exceed the total microbial biomass of many tropical sponge species (Taylor *et al*., 2007).

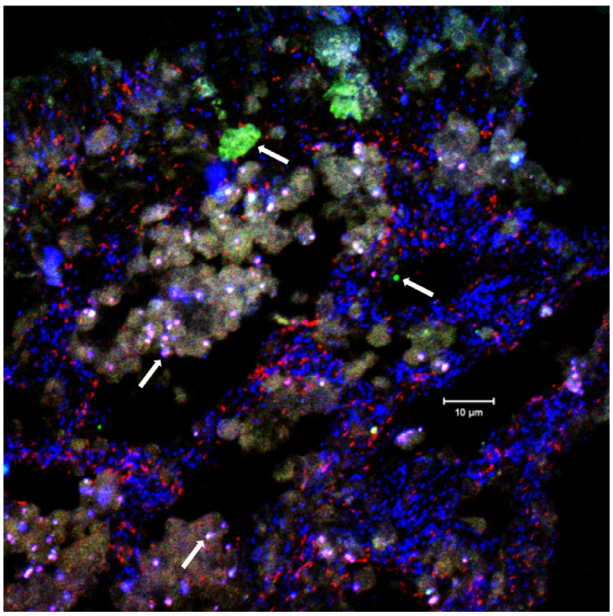


Fig. 1. Fluorescence *in situ* hybridization of a 5 µm cryosection of *I. basta* using double-labelled (Stoecker *et al*., 2010) probe Arch915 in red and the double-labelled probe EUB338-I-III set in blue. Green and white/lila structures represent autofluorescence (see white arrows). As *I. basta* harbours '*Ca*. Nitrosospongia ianthellae' as the only archaeon, all red signals represent this AOA. Blue signals represent bacterial symbionts.

Metaproteogenomic analyses of the thaumarchaeal symbiont of I. basta

An extensive metagenomic data set consisting of 25.8 Gbp of sequence information was obtained for an *I. basta* individual by Illumina sequencing. To facilitate the assembly of symbiont genomes, cell fractions were specifically concentrated for sponge symbionts and dissociated from sponge nuclei. As a result, we recovered a 1.99 Mbp metagenome-assembled genome (MAG) consisting of 113 contigs, representing a nearly complete thaumarchaeal genome (99%) with very low contamination (Supporting Information Table S1). Of all genome-sequenced thaumarchaeotes, the *I. basta* symbiont has the highest average genomic amino acid identity (AAI) of 58.1% with *Ca*. Nitrosopumilus piranensis. As this represents a new species within a new genus (60%–80% AAI is typical for organisms grouped at the genus level; Luo *et al*., 2014; Konstantinidis *et al*., 2017), we propose the name *Ca*. Nitrosospongia ianthellae. The name *Nitrosospongia* describes this organism's ability to oxidize ammonia to nitrite (*nitrosus* is the Latin adjective meaning full of natron; here intended to mean nitrous) whilst residing within a sponge (*spongia* is the feminine Latin noun for sponge). The species name *ianthellae* refers to its discovery and description as a symbiont of the marine sponge, *I. basta*. Furthermore, 16S rRNA gene, *amoA* and concatenated marker gene phylogenies suggest that *Ca*. N. ianthellae is a member of the family *Ca. Nitrosopumilaceae* (Qin *et al*., 2016) (Fig. 2 and Supporting Information Fig. S1). Based on the 16S rRNA gene sequence, *Ca*. N. ianthellae is a member of the sponge-specific sequence cluster 174 (Simister *et al*., 2012) which does not include the three other sponge thaumarchaeal symbionts (*C. symbiosum*, CCThau and DSGS-AOA, a putative deep-sea glass sponge thaumarchaeal symbiont with the assembly accession number GCA_001541925.1) for which genome sequences are available (Hallam *et al*., 2006; Tian *et al*., 2016; Moitinho-Silva *et al*., 2017a). The lack of a close relationship between *Ca*. N. ianthellae and the other sponge-associated symbionts is consistent with the topology of the concatenated single copy conserved marker gene tree (Fig. 2B), whilst unexpectedly *amoA* phylogeny supported a clustering of *Ca*. N. ianthellae with *C. symbiosum* (Supporting Information Fig. S1). Whilst the *amoA* genes of *C. symbiosum* and *Ca*. N. ianthellae showed compositional bias (Li *et al*., 2014), it was not possible to determine whether this caused their monophyletic grouping in the

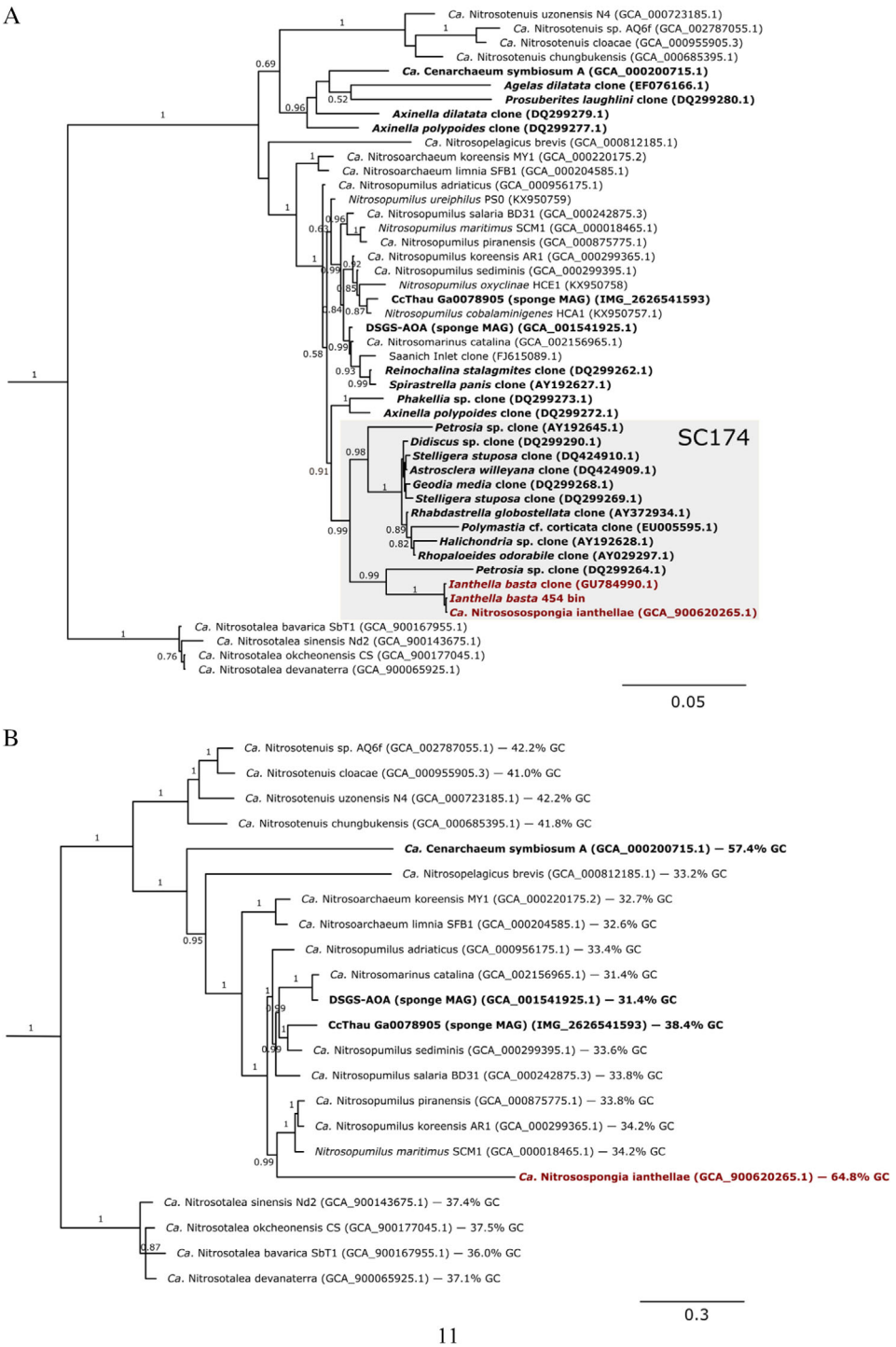


Fig. 2. Phylogeny of *Ca*. Nitrososospongia ianthellae.
A. Bayesian 16S rRNA gene tree.
B. Bayesian phylogenomic tree based on 34 concatenated universal, single-copy marker genes identified with CheckM (Parks *et al*., 2015). Bayesian posterior support values >0.5 are indicated for each branch. Outgroups for both trees consisted of all three genome-sequenced members of the *Nitrososphaera* cluster, both members of the *Nitrosocosmicus* clade, and *Ca*. Nitrosocaldus icelandicus. In all trees, sequences obtained from sponges are depicted in bold.

amoA gene tree. Upon individual addition of *C. symbiosum* and *Ca*. N. ianthellae *amoA* gene sequences to an *amoA* data set, their phylogenetic position in the tree varied dependent on taxa selection (data not shown). Querying the *Ca*. N. ianthellae 16S rRNA gene against the sponge microbiome project (SMP) database (Moitinho-Silva *et al*., 2017b) and against most other publicly available 16S rRNA gene amplicon data sets demonstrated that the habitat of *Ca*. N. ianthellae is restricted to a few sponge species (*I. basta*, *Ancorina alata*, *Stellata maori*, *Stellata aremaria* and *Xestospongia exigua*; see Supporting Information for more details).

An additional unpublished metagenomic data set (250 Mbp generated from pyrosequencing) derived from

a different *I. basta* individual was also screened to confirm the presence of genes of interest (Supporting Information Table S2) and support the phylogenetic inferences displayed in Fig. 2. Whilst the shallow metagenome was excluded from detailed analyses, it confirmed that closely related thaumarchaeal symbionts inhabited both sponge individuals. The average nucleotide identity (ANI) between the two thaumarchaeote MAGs was 98.2%, with 82% coverage of the smaller (Illumina data set) MAG, confirming that members of the same thaumarchaeal species (Konstantinidis *et al*., 2017) reside in both sponge individuals. Both thaumarchaeal symbiont MAGs from *I. basta* possess the highest GC content (64.8%) of any genome-sequenced thaumarchaeote (Fig. 2B) (see Supporting Information for a more detailed discussion of the high GC content).

Metaproteomic analysis was performed on the *I. basta* individual used for Illumina metagenomic sequencing. A total of 513 proteins was detected, 96 of which were specifically assigned to the *I. basta* thaumarchaeote (representing 5.4% of the genes in the MAG). When combining the normalized spectral abundance factor (NSAF; Florens *et al*., 2006) values for all samples and analyses, the 96 thaumarchaeal proteins comprised 10.5% of the 513 identified proteins (Supporting Information Table S3). Of the thaumarchaeal proteins, 60.2% (67.4% NSAF) were encoded by gene families shared by all thaumarchaeotes, and 7.1% (3.1% NSAF) were encoded by genes unique to the *I. basta* thaumarchaeote. Of the 96 expressed proteins, 88 were also encoded at a predicted average amino acid identity of 99.4% $\pm$ 1.4% (SD) in the second MAG recovered by pyrosequencing. The other eight protein homologues were also found in the second MAG but at much lower predicted amino acid identities (from 31.4% to 89%).

Core metabolism of Ca*. N. ianthellae*

AOA oxidize ammonia to conserve energy. Like their bacterial counterparts they activate ammonia with the help of an ammonia monooygenase (AMO). In bacterial ammonia oxidizers AMO forms hydroxylamine that is oxidized by hydroxylamine dehydrogenase to NO (Caranto and Lancaster, 2017) that is further oxidized to nitrite by a yet unknown enzyme. *Ca*. N. ianthellae encodes the recognized repertoire of AOA for ammonia oxidation. All genes encoding the subunits of the ammonia monooxygenase enzyme (*amoA*, *amoB* and *amoC*), including the hypothetical gene *amoX* were identified. *AmoB* and *C* were also detected as proteins (Supporting Information Table S3). The *amo* gene arrangement [*amoA-amoX-amoC-amoB*] is syntenic to most analysed members of *Ca*. Nitrosopumilaceae (Lehtovirta-Morley *et al*., 2011; Park *et al*., 2014). Consistent with other AOAs, no canonical hydroxylamine dehydrogenase was found but *Ca*. N. ianthellae encodes lineage one multicopper oxidases (MCO), which have been suggested as candidates for archaeal hydroxylamine dehydrogenases (Kerou *et al*., 2016). Like most other AOA (except for *Ca*. C. symbiosum), *Ca*. N. ianthellae encodes the putative NO-forming nitrite reductase (*nirK*; found to be highly expressed as protein; Supporting Information - Table S3) that has been suggested to play an important role for archaeal ammonia oxidation (Kozlowski *et al*., 2016; Carini *et al*., 2018) a hypothesis that nicely explains the inhibitory effect of the NO scavenger PTIO used in our incubation experiments described below. Interestingly however, the purple cupredoxin Nmar_1307 from *Nitrosopumilus maritimus* that is capable of oxidizing NO to NO_2^- (Hosseinzadeh *et al*., 2016) is absent in *Ca*. N. ianthellae and other sponge-associated AOA (and also many other AOA), but there are several mononuclear cupredoxins encoded by *Ca*. N. ianthellae that could have the same function. Consistent with all other AOA, no canonical NO- or N_2O-reductases were found suggesting that *Ca*. N. ianthellae cannot perform denitrification, although – like in many other thaumarchaeotes (Liu *et al*., 2012, Zhalnina *et al*., 2014; Santoro *et al*., 2015) – the putative nitric oxide reductase accessory proteins, NorQ and NorD are encoded in the genome (with the NorD subunit being expressed).

Urea and cyanate can be used by some ammonia oxidizers as ammonia source for energy conservation and assimilation. Consistent with marine and non-marine thaumarchaeota (Hallam *et al*., 2006; Tourna *et al*., 2011; Spang *et al*., 2012; Park *et al*., 2014; Bayer *et al*., 2016), *Ca*. N. ianthellae possesses a complete urease gene cluster in addition to a gene encoding a urea active transporter (*DUR3*; 86% homologous to *Ca*. Nitrosopumilus piranensis) and is thus capable of taking up (or using internally produced) urea and converting it to ammonia and CO_2. The sponge symbionts *C. symbiosum* and CCThau also encode urea transporters and urease gene clusters. Furthermore, urease subunit gene transcripts from CCThau were found in the metatranscriptomes from *C. concentrica* (Moitinho-Silva *et al*., 2017a). Thus, it seems likely that urea is not only used by some free-living marine AOA (Alonso-Sáez *et al*., 2012; Kitzinger *et al*. 2019) and a few cultured AOA (Fig. 3), but also appears to be a common substrate for group I.1a Thaumarchaeota residing in marine sponges. In this context it is also interesting to note that some sponges excrete urea (Morley *et al*., 2016). In contrast to *Nitrososphaera gargensis* (Palatinszky *et al*., 2015), *Ca*. N. ianthellae does not contain a cyanase for ammonia generation from cyanate, but like all other AOA, it encodes a protein with modest homology to a creatinine-amidohydrolase indicating that it could utilize sponge-derived creatinine and convert it to creatine.

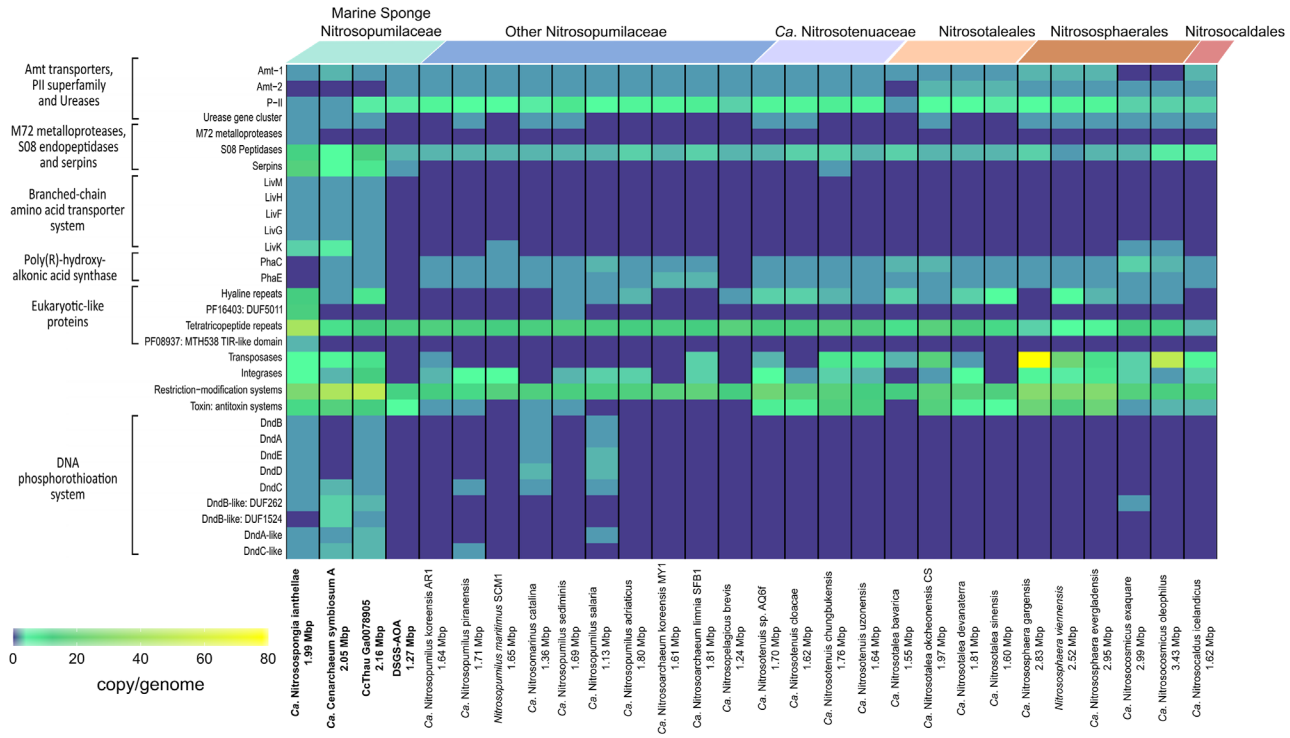


Fig. 3. Heat map showing the distribution and gene copy number per genome of selected genes and gene classes among genome-sequenced AOA. The colour scale indicates copies per genome and MAG respectively. Sponge-derived MAGs start on the left and are depicted in bold, followed by members of *Ca*. Nitrosopumilaceae, *Ca*. Nitrosotenuaceae, *Ca*. Nitrosotaleales, the *Nitrososphaerales*, and *Ca*. Nitrosocaldales. Genome sizes are listed next to each species name. An extended version of this Figure is available as Supporting Information Fig. S6.

Whilst all genome-sequenced AOA symbionts lack a canonical creatinase that would form urea from creatine, AOA including *Ca*. N. ianthellae do possess a Xaa-Pro aminopeptidase that has been hypothesized as a functional analog (Moitinho-Silva *et al*., 2017a).

As expected, *Ca*. N. ianthellae encodes the thaumarchaeal 3-hydroxypropionate/4-hydroxybutyrate pathway for autotrophic CO_2 fixation (see Supporting Information for more details). Regarding nitrogen assimilation, *Ca*. N. ianthellae encodes and expresses an *amt* transporter of the *amt-2* lineage which has been hypothesized to be a high affinity ammonia transporter based on indirect evidence from transcriptomic data (Nakagawa and Stahl, 2013; Offre *et al*., 2014). The *amt-2* gene of *Ca*. N. ianthellae is nested within the complex V ATP synthase operon (of which four of the encoded subunits were found to be expressed). Interestingly, in contrast to most other AOA, no putative low affinity *amt* transporter gene was found in *Ca*. N. ianthellae, a feature shared with the sponge symbionts *Ca*. C. symbiosum (Hallam *et al*., 2006) and CCThau (Fig. 3; Moitinho-Silva *et al*., 2017a). Under the assumption that *amt* affinity groups can reliably be inferred from annotation, this finding might indicate that the dense thaumarchaeal populations in some marine sponges including *I. basta* (see also Radax *et al*., 2012a) may be perpetually ammonia limited and thus might rely on a high affinity ammonia transporter in combination with the uptake of organic nitrogen sources like urea or amino acids. Regarding nitrogen assimilation, it is also noteworthy, that the almost complete bin of *Ca*. N. ianthellae contains only a single gene encoding a member of the nitrogen regulatory protein PII superfamily, whereas, with the exception of *Ca. C. symbiosum*, higher copy numbers of these genes are generally found in other AOA (Fig. 3; Kerou *et al*., 2016). This observation might indicate that *Ca*. N. ianthellae lives in a rather stable environment in regard to nitrogen availability compared to other AOA. Furthermore, in contrast to most other AOA, the two key genes coding for subunits of the polyhydroxyalkanoate (PHA) synthase *phaC* and *phaE* are apparently lacking in *Ca*. N. ianthellae. We also found no indications for use of alternative carbon storage compounds such as starch or glycogen in *Ca*. N. ianthellae. The apparent lack of carbon storage would be consistent with balanced growth in a chemically stable environment.

The negatively charged surface of the Thaumarchaeota S-layer was recently proposed to help concentrate the charged solute ammonium into the pseudo-periplasmic space (Li *et al*., 2018). Intriguingly, *Ca*. N. ianthellae encodes many putative surface-layer (S-layer) proteins

(Supporting Information Fig. S2), with six of them being expressed at combined NSAF values of 6.78% (Supporting Information Table S3). The theoretical isoelectric points of all *Ca*. N. ianthellae S-layer proteins (3.62–4.92) were in a similar range to those calculated for *N. maritimus* and *Nitrosoarchaeum limnia* SFB1 S-layer proteins (3.4–4.08), consistent with the proposed mechanism for charged solute acquisition. In addition to its ammonium concentration effect, the expanded group of S-layer proteins in *Ca*. N. ianthellae (and *Ca. C. symbiosum*) may be involved in adhesion, surface recognition or other types of interactions within the extracellular matrix of the sponge environment.

Shared gene families of sponge AOA

During genome annotation of *Ca*. N. ianthellae, we particularly focused on gene families unique to this organism or to *Ca*. N. ianthellae and other thaumarchaeal sponge symbionts (among all genome-sequenced thaumarchaeotes) as these gene families likely represent adaptations to a sponge-associated lifestyle. Whilst 40% (represented by 681 gene families) of the *Ca*. N. ianthellae genes were members of core gene families represented in every sequenced thaumarchaeote, 35% (represented by 616 gene families) had no close orthologues among other thaumarchaeotes. Additional pairwise comparisons between *Ca*. N. ianthellae and all other thaumarchaeote genomes were performed in order to identify shared gene families of species pairs that are absent in all other thaumarchaeotes. Interestingly, the highest number occurred with CCThau (n = 18) and *Ca*. C. symbiosum (n = 11) (Supporting Information Fig. S3). Furthermore, 14 gene families were found to be exclusively shared among all three thaumarchaeal sponge symbionts, whilst only one gene family was found to be shared with DSGS-AOA and *C. symbiosum*. In total, 44 gene families were exclusively shared with at least one other thaumarchaeal sponge symbiont representing 2.4% of the *Ca*. N. ianthellae genes. These genes are likely candidates for adaptations of AOA to life within sponges.

Ca*. N. ianthellae and other sponge AOA possess proteases and their inhibitors*

Among the genes that were unique to *Ca*. N. ianthellae or exclusively shared with other sponge AOA were several genes predicted to be involved in degradation of extracellular protein and inactivation of extracellular proteases, which may also be host-derived. *Ca*. N. ianthellae possesses a putatively exported metalloprotease of the M72 family (Drapeau, 1980; Passmore *et al*., 2015) that was also detected in the metaproteome. This large protease (2027 AA) contains hyaline repeat domains suggestive of an additional adhesive property. The M72 family of metalloendopeptidases have so far not been found in any *Archaea* (Trame *et al*., 2014) and most characterized members of this enzyme family are peptidyl-Asp metalloendopeptidases that hydrolyze bonds on the NH_2-terminal side of aspartic acid and cysteic acid residues (Drapeau, 1980). Furthermore, the *I. basta* thaumarchaeal symbiont contained 11 genes affiliated with four gene families that encode subtilisin-like serine protease domains (S08A family endopeptidases) (Supporting Information Fig. 4A). Although serine endopeptidases of the S08A family are also found in other thaumarchaeotes, except for *C. symbiosum* (four copies), and CCThau (13 copies), most contain only two copies. Furthermore, it is noteworthy that two of the four S08A family endopeptidase gene families were either exclusively found in *Ca*. N. ianthellae or shared with CCThau (Fig. 4, Supporting Information Fig. 4A). Interestingly, several of the *Ca*. N. ianthellae S08A endopeptidases are predicted to be exported without a membrane anchor, whilst most other thaumarchaeal S08A serine endopeptidases seem to be membrane anchored. This suggests that the *Ca*. N. ianthellae S08A endopeptidases act at larger distances outside of the pseudoperiplasmic space where host- or food-derived proteins that cannot pass through the S-layer will be available.

Ca. N. ianthellae also encodes 15 genes for the I4 family of serpins, which represent serine protease inhibitors. Some of these cluster with serpins from *C. symbiosum* (Hallam *et al*., 2006) and CCThau, although among the non-host associated AOA, only *Nitrosotenuis chungkubensis* encodes a serpin (Supporting Information Fig. 4B). Nine of the *Ca*. N. ianthellae serpins are predicted to be extracellular and are frequently found adjacent to S08A family endopeptidases in the genome (Fig. 4B). Serpins belong to a large family of irreversible inhibitory substrates of proteases, often but not exclusively of the serine class, that has been widely characterized in mammals, insects, plants as well as some viruses, and which can mediate host–microbe interactions (Ventura *et al*., 2012). These findings suggest that AOA living as sponge symbionts might use serpins to regulate their own secreted endopeptidases and/or sponge serine proteases found to be highly expressed in other sponges (Riesgo *et al*., 2014), including within the host of *C. symbiosum* (Zaikova, E., PhD thesis, 2007).

Peptide and amino acid uptake by Ca*. N. ianthellae and other sponge AOA*

Ca. N. ianthellae, like many other AOA, is well-equipped for uptake of oligo- and dipeptides as well as amino acids by encoding a set of transporters widely distributed in this clade. After uptake, *Ca*. N. ianthellae, like other AOA, has the genomic repertoire to degrade these peptides

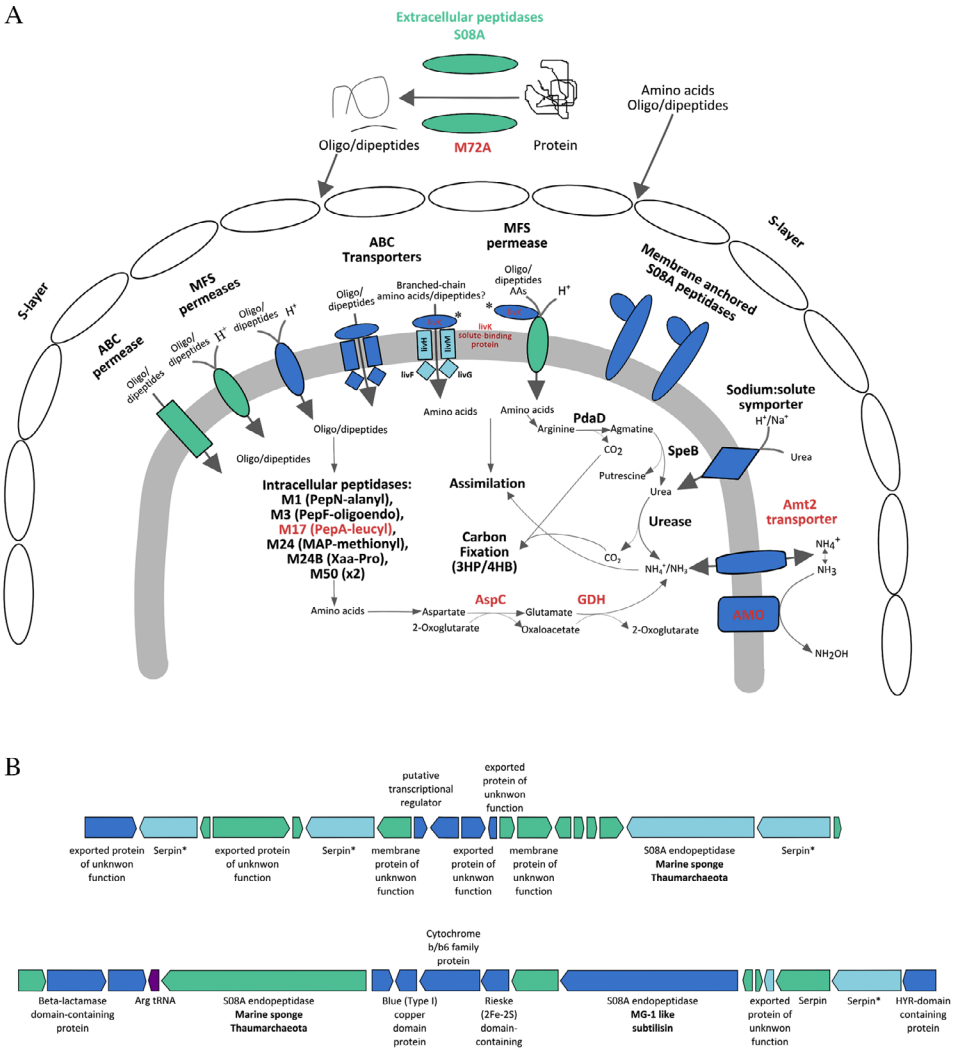


Fig. 4. A. Reconstructed metabolic pathways of genes and their expression (red) detected in *Ca*. N. ianthellae proposed to be involved in extracellular (and intracellular) protein degradation as well as amino acid transport and assimilation. Predicted proteins (and their respective subunits when relevant) are colour coded to denote the degree of homology among all sequenced Thaumarchaeota: green – *Ca*. N. ianthellae unique gene families; light blue – shared exclusively among thaumarchaeal sponge symbionts; dark blue – ubiquitously found in Thaumarchaeota. Extracellular proteins derived from the marine environment as well as the sponge mesohyl may be degraded extracellularly or in the thaumarchaeal pseudo-periplasmic space. These resultant oligo/dipeptides and amino acids which can also be derived from the environment can then be transported by a suite of ABC transporters and major facilitator superfamily (MFS; Newstead, 2015 and references therein) permeases into the cytoplasm to be further degraded by intracellular peptidases or assimilated. Amino acids such as arginine and aspartate can be further degraded to form NH_3/NH_4^+ for assimilation or export to the pseudo-periplasmic space for ammonia oxidation. Arginine can be decarboxylated by arginine decarboxylase (PdaD) to agmatine which can then be degraded to urea by agmatinase (SpeB). Both proteins are ubiquitously distributed among the Thaumarchaeota. All but one of the branched-chain amino acid transporter subunits (LivFGHMK operon) are exclusively found among thaumarchaeal sponge symbionts, whilst the periplasmic solute binding subunit (LivK – found to be expressed and denoted by an asterisk) can be found not only among sponge symbionts but also in *N. maritimus* and members of the genus *Nitrosocosmicus* (see also Fig. 3).
B. Genetic context map depicting examples of colocalized S08A endopeptidases and serine protease inhibitors (serpins). Colour coding as in panel (A), except purple denotes the presence of a tRNA, sites at which gene insertions are common. The asterisk next to serpins colour coded as sponge-specific (light-blue) denotes that *Ca*. *N. chungbukensis* is the sole non-sponge symbiont encoding a serpin and belonging to this orthologous group (see Supporting Information Fig. S4B). AspC, aspartate aminotransferase; GDH, glutamate dehydrogenase.

and amino acids and release ammonia (Fig. 4). In addition to features common to many AOA, *Ca*. N. ianthellae also contains the *livFGHMK* operon which in proteobacteria and cyanobacteria encodes a high-affinity branched-chain amino acid transport system (which can also transport other amino acids) belonging to the ATP binding cassette (ABC) superfamily of transporters (Hoshino, 1979; Adams *et al*., 1990; Hosie *et al*., 2002; Picossi *et al*., 2005) (Fig. 4). Among all other thaumarchaeote genomes, only *C. symbiosum*, CCThau, and a thaumarchaeote MAG from the Caspian Sea (Mehrshad *et al*., 2016) encode this transporter (Supporting Information Fig. S5). The periplasmic

subunit LivK from *Ca*. N. ianthellae was expressed as protein (Supporting Information Table S3) and this subunit was also reported to be expressed in CCThau (Moitinho-Silva *et al*., 2017a). Adjacent to the expressed *livK* in *Ca*. N. ianthellae is a gene encoding three consecutive *livK* domains (along with the requisite ligand binding sites) which is congruent with *C. symbiosum* containing 4 copies of the *livK* solute binding component (Supporting Information Fig. S5A). In *C. symbiosum*, the transporter encoding the *livFGHMK* operon is located next to several extracellular trypsin-like serine protease encoding genes, suggesting that the transporter is involved in amino acid uptake in this organism. However, given the substrate promiscuity of this transporter family (Adams *et al*., 1990; Valladares *et al*., 2002; Beckers *et al*., 2004; Picossi *et al*., 2005) (Supporting Information Fig. S5), experimental validation will be required before more specific predictions on its function in sponge AOA can be made. Such investigations would be particularly important, as multiple *livFGHMK* operons are also present in two γ-proteobacterial sponge symbionts (Gauthier *et al*., 2016), suggesting that amino-acid transport and utilization mediated by this transporter-type could be an important feature of sponge symbionts and may even contribute to sponge-mediated dissolved organic matter transfer to higher trophic levels (de Goeij *et al*., 2013). Proteases are also important for sponge metabolism and a body of research has focused on sponge proteases and their inhibitors from a biodiscovery context (Arreguín *et al*., 1993; Wilkesman and Schröder, 2002; Wilkesman and Schröder, 2007; Tabares *et al*., 2011). In these analyses generally, sponge holobiont samples are used for enzyme or inhibitor purification and characterization. Thus, our observation that abundant archaeal sponge symbionts encode and express putatively exported proteases and serpins, demonstrate that in such assays, it will remain unclear whether it is the microbial symbionts, or the host animals that are producing the analysed biomolecules.

Mixotrophy of Thaumarchaeota

A mixotrophic lifestyle for Thaumarchaeota has been inferred since the earliest reports on planktonic thaumarchaeotes described incorporation of organic carbon into signature lipids (Ingalls *et al*., 2006) and incorporation of labelled amino acids leucine (Ouverney and Fuhrman, 2000; Herndl *et al*., 2005) and aspartate (Teira *et al*., 2006) into single planktonic cells. Whilst initial findings – that pure cultures of Thaumarchaeota in group I.1a (Qin *et al*., 2014) and group I.1b (Tourna *et al*., 2011) could have increased growth rates in the presence of α-keto acids and ammonia – were subsequently proven to be a result of the H_2O_2 scavenging ability of α-keto acids (Kim *et al*., 2016; Qin *et al*., 2017), abundant genes and transcripts of S08A family serine proteases and other metallopeptidases in MG-1 deep-sea thaumarchaea have also been attributed to mixotrophic activity (Li *et al*., 2015). A deep sea Thaumarchaeota has also been found to encode the *livFGHMK* operon (Mehrshad *et al*., 2016), which clusters with the two sponge thaumarchaeotes (Supporting Information Fig. S5), highlighting both environments as potential ecological reservoirs of mixotrophic thaumarchaea. Furthermore, growth of thaumarchaeotes in an industrial wastewater treatment plant that was uncoupled to ammonia-oxidation has been described (Mußmann *et al*., 2011). Interestingly, many members of the Aigarchaeota, a sister group to the Thaumarchaeota (Guy and Ettema, 2011), also encode the full *livFGHMK* operon, and a single cell aigarchaeal-like genome from cold marine sediments encodes additional extracellular proteases, di- or tripeptide transporters and aminotransferases (Lloyd *et al*., 2013). It is therefore tempting to speculate that the thaumarchaeal ancestor was a mixotroph or even strict heterotroph (as also indicated by genomic and experimental data that deep branching clade I1c and d members in the Thaumarchaeota lack genes required for ammonia oxidation; Beam *et al*., 2014; Lin *et al*., 2015; Weber *et al*., 2015), and that sponge Thaumarchaeota along with a few other members of this clade retained the capability to use amino acids due to specialized environmental conditions. Future experiments would need to demonstrate uptake of amino acids by *Ca*. N. ianthellae to reveal whether it fuels assimilation or heterotrophic growth and/or is used for sequential intracellular generation and oxidation of ammonia (de Boer and Laanbroek, 1989; Burton and Prosser, 2001).

Eukaryotic-like proteins (ELPs) in Ca*. N. ianthellae*

Ca. N. ianthellae encodes a number of genes containing domains postulated to have an evolutionary origin within the eukaryotes and which are thought to be important for modulating interactions between bacteria and eukaryotic hosts (Callebaut *et al*., 2000; Lurie-Weinberger *et al*., 2010; Patterson *et al*., 2014). Recent metagenomic analyses revealed an abundance of genes encoding such eukaryotic-like proteins (ELPs) in the bacterial symbionts of sponges (Fan *et al*., 2012; Reynolds and Thomas, 2016; Díez-Vives *et al*., 2017), with many of these being expressed (Díez-Vives *et al*., 2017). Four types of ELPs are found in *Ca*. N. ianthellae: Proteins with tetratricopeptide repeats (TPR), the Toll-interleukin-1 receptor (TIR)–like domain PF08937 (DUF1863; Cort *et al*., 2000; Essuman *et al*., 2018), immunoglobulin-like (Ig-like) domains (DUF5011; Shigeno-Nakazawa *et al*., 2016), and hyaline repeats (HYR; Callebaut *et al*., 2000) and an extensive discussion of these ELPs is provided in the Supporting Information.

Mobile and selfish genetic elements in Ca. *N. ianthellae*

In contrast to other members of *Ca*. Nitrosopumilaceae, *Ca*. N. ianthellae along with the other two sponge AOA, *C. symbiosum* and CcThau (but not DSGS-AOA) are enriched in transposases, restriction-modification (RM) systems (including a Type II restriction endonuclease, PF13156, which is exclusively found in sponge AOA – Supporting Information Fig. S6), toxin-antitoxin (T-A) systems, as well as genes putatively involved in DNA phosphorothioation (Fig. 3, Supporting Information Fig. S6). However, no differential enrichment of integrases was detected within sponge-associated thaumarchaotes. The abundance of transposases and other mobile/selfish genetic elements (MGEs/SGEs) in sponge AOA is consistent with what has been reported for other sponge-associated microbes (Fan *et al*., 2012, Horn *et al*., 2016), suggesting that evolution of AOA sponge symbionts compared to free-living marine AOA (see also Fig. 3) is more heavily shaped by horizontal gene transfer in the concentrated milieu of environmental bacteria and viruses resulting from sponge feeding and pumping activity.

Among the above-mentioned genetic elements, the complete genetic repertoire for DNA phosporothioation (PT) (Wang *et al*., 2007; You *et al*., 2007) in *Ca*. N. ianthellae and CcThau (*dndA, B, C, D, E*) is particularly noteworthy as among cultured AOA, only *Ca*. Nitrosomarinus catalina (Ahlgren *et al*., 2017) and *Ca*. Nitrosopumilus salaria also encode this DNA modification system. These genes possibly encode a primitive immune system by enabling discrimination between self and non-self DNA (but can also epigenetically affect the transcriptome of bacteria; Tong *et al*. 2018) and might thus – in concert with the better understood canonical RM systems (Fig. S6) – be particularly useful for sponge symbionts living in a system where they are prone to genetic exchanges. In addition, *Ca*. N. ianthellae, CcThau and *C. symbiosum* have genes with low similarity to *dndA, B* and *C* respectively. The *dndB*-like genes contain domains (DUF262, DUF1524) previously identified as components of R-M systems (Miller, 2012; Machnika *et al*., 2015) which are known to be enriched in sponge microbiomes (Fan *et al*., 2012, Horn *et al*., 2016).

Ca. N. ianthellae also encodes most enzymes necessary for the production of archaeosine, a highly modified tRNA nucleoside (Phillips *et al*., 2012) which putatively confers structural stability that is in some archaea important for growth at low temperature (Blaby *et al*. 2010). Interestingly, among all analysed Thaumarchaeota, the critical enzyme (aTGT) of this pathway is missing the RNA binding site (PUA domain) whilst maintaining the conserved substrate binding pocket found in *Crenarchaeota* (Phillips *et al*., 2012). Although the PUA domain is dispensable for archaeosine formation (Sabina and Söll, 2006) it is as yet unknown whether *Ca*. N. ianthellae can use this for DNA modification, as recently described for another restriction-modification system variant in a *Salmonella* species (Thiaville *et al*., 2016).

Ammonia-oxidation is exclusively mediated by Ca. *N. bastadiens in* I. basta *and is coupled to carbon fixation*

The presence of *amoA*-encoding and expressing thaumarchaea does not prove that these microbes actually perform ammonia-oxidation in a system (Mußmann *et al*., 2011). Consequently, nitrification rates of *I. basta* harbouring the thaumarchaeal symbiont were experimentally determined to verify ammonia-oxidizing activity and to obtain insights into the mean nitrifying activity per symbiont cell. Interpretation of these data were facilitated by the fact that *I. basta* according to previous 16S rRNA gene based surveys (Webster *et al*., 2009; Luter *et al*., 2010; Freckelton *et al*., 2012; Luter *et al*., 2012) and our metagenomic and metaproteomic data contains a single AOA symbiont species and does not harbour bacterial ammonia-oxidizers or comammox organisms (Daims *et al*., 2015). Incubation experiments with freshly collected sponge clones were performed in the presence of different ammonium concentrations ranging from ambient seawater (0.29 $\pm$ 0.1 μM) via 25 to 100 μM. Gross and net nitrification rates as well as net fluxes of ammonium, nitrite and nitrate were determined in 24 h laboratory incubation experiments that were repeated several times over 7 days (Supporting Information Fig. S7). Net nitrification rates were inferred from the measurable increase of the nitrification products nitrite and nitrate, but can underestimate the actual nitrification rate if these products are concurrently consumed in the system. Gross nitrification rates were determined by the isotope pool dilution technique that directly measures the total gross nitrite and nitrate formation and is not affected by concurrent consumption of these compounds (for details see experimental procedures). Across all treatments, gross and net nitrification rates were similar in magnitude and highly correlated (gross rates = 1.13 $\times$ net rates +0.6; R^2 = 0.94; $p < 0.01$), suggesting that nitrate removal (e.g. via assimilation or denitrification) did not occur at significant rates, with the exception of sponges in the 100 μM NH_4^+ treatment (Fig. 5A). Gross nitrification rates of sponges incubated with 25 or 100 μM NH_4^+ were significantly greater than in ambient seawater ($p < 0.05$; ANOVA followed by Tukey HSD Test) (Fig. 5A). The stimulation of nitrification by increased ammonium availability, which has also been observed for other sponge species (Corredor *et al*., 1988; Bayer *et al*., 2007; Bayer *et al*., 2008; Schläppy *et al*., 2010), indicates ammonium limitation of the symbiotic nitrifiers under ambient

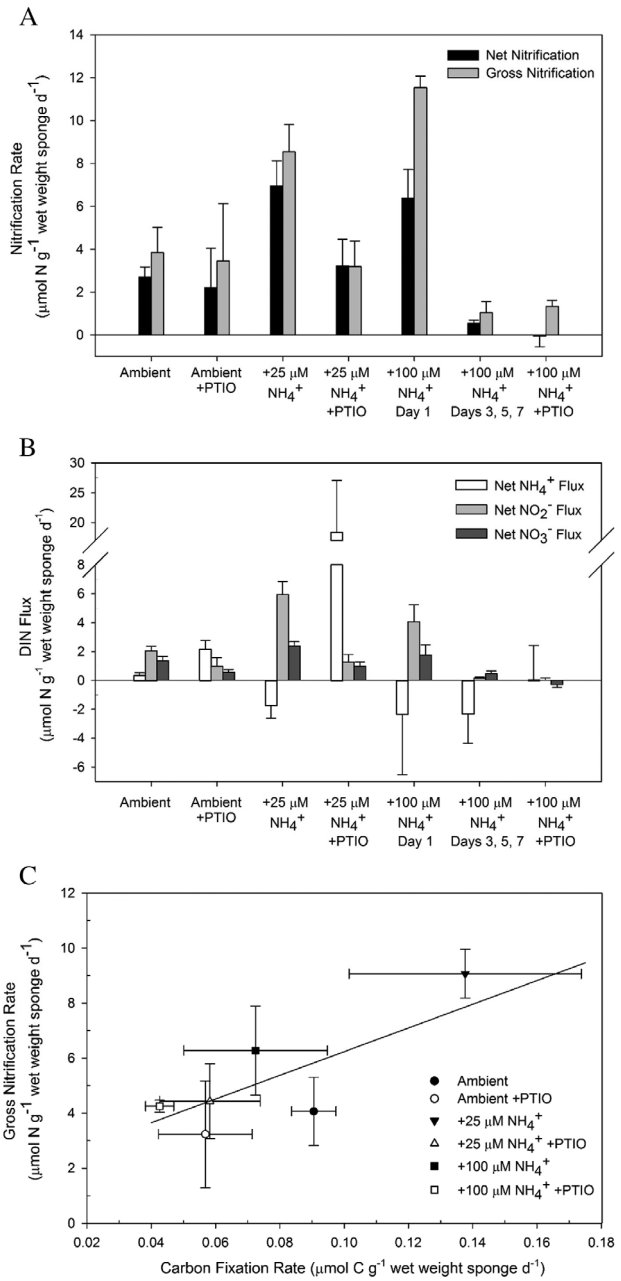


Fig. 5. Nitrification activity of the *I. basta* holobiont during 7-day incubation experiments at ambient conditions and with added ammonium (25 or 100 μM) and with/without the AOA inhibitor PTIO. A schematic overview of the corresponding experimental setup is given in Supporting Information Fig. S7.

A. Depicts net nitrification (as calculated from the addition of net NO_2^- and NO_3^- flux) and gross nitrification rates (estimated using the $^{15}NO_3^-$ isotope pool dilution method).

B. Net fluxes of all DIN species for the ambient, +25 μM NH_4^+ and +100 μM NH_4^+ treatments. Rates of the different days in the ambient and +25 μM $NH4^+$ incubations were averaged as they were not statistically different.

C. The relationship between gross nitrification rates of the *I. basta* holobiont and the carbon fixation rate of sponge clones sampled on day 7. The carbon fixation rates derived from the *I. basta* holobiont nitrification experiments displayed a positive and significant correlation with gross nitrification rates (R = 0.665, $p < 0.005$). Error bars in panel C reflect the standard error of the sample mean, where $n = 3$ for the carbon fixation rates and for all weighted gross nitrification rates in all treatment conditions. PTIO additions started 2 days after the experiment was commenced; hence, fixation of ^{13}C-labelled bicarbonate during the first 2 days was not influenced by inhibition.

conditions. Almost no net ammonium release was observed from the *I. basta* holobiont in unamended seawater, indicating similar *in situ* rates of ammonium production and consumption (Fig. 5B). Net nitrification rates of the *I. basta* holobiont under ambient experimental conditions were slightly higher but generally comparable to those reported for other sponge species (Supporting Information Table S4; to the best of our knowledge no gross nitrifications reports for sponges have been published). These data likely reflect nitrification rates under natural field conditions, as NH_4^+ concentrations in the unamended treatment in the aquaria experiments were consistent with those reported for Orpheus Island, where *I. basta* was collected (Jompa and McCook, 2002).

Net nitrification rates in sponges subjected to 100 μM NH_4^+ for 24 h were significantly lower than gross nitrification rates ($p < 0.05$; two-tailed t-test, Fig. 5A), indicating increased nitrate consumption most likely by denitrification by other sponge holobiont members (in this context it should be kept in mind that oxygen concentration will decrease if nitrification is stimulated by ammonia addition). Gross and net nitrification rates also became significantly depressed after prolonged exposure (> 3 days) to 100 μM NH_4^+, being significantly lower than the 24 h incubations at 100 μM NH_4^+ and the ambient seawater controls ($p < 0.01$, Mann–Whitney U-test) (Fig. 5A). From AOA pure culture studies, there are no indications that 100 μM NH_4^+ can be inhibitory for members of this clade, with concentrations of 2 to 20 mM NH_4^+ needed for inhibition (Hatzenpichler *et al*., 2008; Martens-Habbena *et al*., 2009; Tourna *et al*., 2011; Li *et al*., 2016; Sauder *et al*., 2017; Sauder *et al*., 2018). Whilst we cannot exclude that *Ca*. N. ianthellae is particularly sensitive to ammonium, it is also possible that the inhibition was caused by indirect effects. For example, if the sponge host was adversely affected at 100 μM NH_4^+, decay of sponge cells could cause stimulation of heterotrophic microbes and thus oxygen limitation. Typical acute toxicities in freshwater fish and invertebrates (96-h LC_{50}) are reached at unionized ammonia (NH_3) concentrations between 4.7 and 38 μM (with salinity typically having a slight ameliorating effect and tests on marine invertebrates very rare; Boardman *et al*., 2004) and susceptibility to ammonia toxicity can be increased by low O_2 concentrations (Camargo and Alonso, 2006). At ambient incubation conditions about 4.5% of total ammonia was present as NH_3, hence *I. basta* may have experienced physiological stress and

intermittent hypoxic conditions in the 100 μM NH_4^+ treatments.

PTIO, a scavenger of free radical nitric oxide (NO) (Amano and Noda, 1995; Ellis *et al*., 2001), has been described as a specific inhibitor of AOA (Yan *et al*., 2012; Martens-Habbena *et al*., 2015) and was therefore used to demonstrate that *Ca*. N. ianthellae is responsible for ammonia-oxidation in *I. basta*. As expected, the addition of PTIO to sponge incubations in the ambient and 25 μM NH_4^+ amended treatments, resulted in significantly depressed gross and net nitrification rates and increased ammonium accumulation when compared to the non-PTIO treated sponge incubations ($p < 0.05$; except for gross nitrification rates under ambient conditions, $p = 0.885$; Mann–Whitney U-test; Fig. 5A,B). In the case of the 100 μM NH_4^+ amendment, gross and net nitrification was already strongly reduced in the absence of PTIO after 3 days of incubation, so further inhibition by PTIO could not be demonstrated under these conditions (Fig. 5A). Residual ammonia oxidizing activity in some of the experiments in the presence of 75 μM PTIO was unexpected as pure AOA cultures are known to be fully or almost completely inhibited by this concentration of PTIO (Shen *et al*., 2013; Martens-Habbena *et al*., 2015). NO production by *I. basta* (many sponges express NO synthases, Riesgo *et al*., 2014**)** may have contributed to PTIO inactivation thereby lowering its inhibitory effect on the AOA. To our knowledge, PTIO has not previously been used in sponge microbiome research (but is commonly applied to eukaryotic tissues including sponges; Müller *et al*., 2006; Ueda *et al*., 2016) and further optimization of the concentration is recommended for future applications in these animals. Study design of such experiments should also take into consideration that PTIO has been shown to be cytotoxic for some microbes (Kits *et al*., 2019).

In 29 of the 33 sponge incubations (incubations with added PTIO or inhibited by addition of 100 μM NH_4^+ were excluded) (Fig. 5B), accumulation rates of NO_2^- exceeded those of NO_3^- by an average factor of 2.8 and this difference was found to be significant in the ambient and 25 μM NH_4^+ treatments (both $p < 0.01$, Mann–Whitney U-test). In contrast, no significant difference between NO_2^- and NO_3^- production was observed in the PTIO-amended incubations nor in those experiments where prolonged exposure to 100 μM NH_4^+ negatively affected net nitrification rates (all: $p > 0.05$, Mann–Whitney U-test).

In addition, several control experiments were performed in order to better understand the influence of the experimental setup and the seawater microbial community on our results. All nitrification experiments discussed above were performed in intermittently closed aquaria (Supporting Information Fig. S7) to prevent loss of labelled CO_2 but to provide sufficient oxygen exchange to maintain sponge health. Control experiments were performed at ambient and 100 μM NH_4^+, to compare DIN fluxes and inferred net nitrification rates using this setup, to incubations performed in constantly open containers. Interestingly, slightly but significantly higher NO_2^- fluxes and net nitrification rates were observed with ambient seawater in the intermittently closed aquaria (both $p < 0.01$, two-tailed t-test; Supporting Information Fig. S8), indicating that ammonia-oxidizers in the closed system either benefited from the minimized loss of the added bicarbonate (via CO_2 off gassing) or from potentially increased ammonia production from stressed sponge clones. Furthermore, we determined the contribution of the seawater microbial community to net nitrification by using aquaria seawater without sponges. These experiments revealed minimal nitrification with net rates of only 0.06 ± 0.3 μM N d^{-1} ($n = 12$).

To calculate cell specific ammonia oxidation rates for *Ca*. N. ianthellae, its 16S rRNA genes were quantified at the end of the 7 day sponge incubation under ambient and +25 μM NH_4^+ treatments. Sponges subjected to ambient and +25 μM NH_4^+ treatments contained thaumarchaeal symbiont gene copy numbers of $8.58 \pm 4.9 \times 10^{10}$ and $1.64 \pm 0.12 \times 10^{10}$ per g wet weight (SE, $n = 3$, for both) respectively. By dividing the net nitrification rate on day 7 by the 16S rRNA gene copy number of *Ca*. N. ianthellae on this day, average cell specific rates were estimated to be 0.11 ± 0.08 and 0.66 ± 0.09 fmol NH_4^+ oxidized per cell per day (SE, $n = 3$, for both) for the ambient and +25 μM NH_4^+ incubations respectively. Inferred cell specific ammonia oxidation rates for *Ca*. N. ianthellae were ~20–120 times lower than what has been reported for *N. maritimus* (Martens-Habbena *et al*., 2009) and ~6–65 times lower than coastal marine seawater (Wuchter *et al*., 2006), whilst the highest ammonia oxidation rate of *Ca*. N. ianthellae was comparable to the lowest rates detected in the sponge *Phakellia ventilabrum* (Radax *et al*., 2012a). The relatively low cell specific ammonia oxidation rate might reflect that not all *I. basta* AOA symbiont cells are physiologically active (or even in the process of digestion by the host) or that the sponge host exerts some control over the ammonia oxidation of the symbiont.

Furthermore, as ^{13}C-bicarbonate was added during all experiments (Supporting Information Fig. S7), we were able to measure the $\delta^{13}C$ values of sponge tissues and calculate inorganic carbon fixation rates to compare with the gross nitrification rates. PTIO was found to reduce the $\delta^{13}C$ value in both the ambient and + 25 μM NH_4^+ treatments (Supporting Information Fig. S9; both: $p < 0.05$, one-tailed t-test), whilst no significant differences were observed with the +100 μM NH_4^+ treatments. In addition, gross nitrification rates and carbon fixation were significantly positively correlated (Fig. 5C). These data suggest that *Ca*. N. ianthellae contributes significantly to carbon fixation by the sponge holobiont. This finding is consistent

with (i) the detection of key genes for CO_2 fixation in the *I. basta* thaumarchaeote metagenome bin, (ii) the detection of some of the respective proteins in the metaproteome, and (iii) the absence of autotrophic CO_2 fixation pathways in the metagenomic bins of the alpha- and gamma-symbionts (data not shown). In addition, the slope of the positive linear relationship between the gross nitrification rate and the carbon fixation rate (Fig. 5C; 43.0), reflects the gross nitrification: carbon fixation ratio (N:C ratio), which is most likely over-estimated since the experimental setup of the intermittently closed aquaria still allowed for the loss of added ^{13}C-bicarbonate. In addition, CO_2 production derived from respiratory activities within the *I. basta* holobiont were not accounted for, which may also lead to an underestimation of carbon fixation rates. Despite these possible biases the N:C ratio determined for *I. basta* is within the range typically found for aquatic environmental samples (2–60) (Andersson *et al*., 2006), but higher than the values we have inferred for *N. maritimus* (~7–29; Könneke *et al*., 2005; Martens-Habbena *et al*., 2009). The relatively high N:C ratio along with the low cell-specific ammonia-oxidation rates within the *I. basta* holobiont are consistent with the proteogenomic-derived hypothesis that *Ca*. N. ianthellae is not sustaining its population only by chemolithoautotrophic growth on ammonia, but more likely grows as a mixotroph.

Surprisingly, nitrite-oxidizing bacteria (NOB) could not be detected in *I. basta* by amplicon sequencing, metagenomic sequencing, or with FISH. Consistent with the absence of NOB, NO_2^- accumulated to high concentrations ranging from 8 to 21 μM in ambient treatments containing *I. basta* (Fig. 5B, Supporting Information Fig. S8), making this sponge one of the few natural systems in which greater NO_2^- than NO_3^- concentrations occur (Brezonik and Lee, 1968; Lam *et al*., 2011; Schaefer and Hollibaugh, 2017 and refs. therein). Whilst nitrite accumulation might have been more pronounced in the closed aquaria system than in the open ocean, we have shown before that strong nitrite gradients do occur in nitrifying open biofilm systems and possibly affect microbial community structure (Maixner *et al*. 2006). It will be interesting to explore in future studies how high the nitrite concentration is within *I. basta* and whether nitrite production contributes (via nitrite toxicity; Camargo and Alonso, 2006 and references therein) to protection from predators or to the unusually low microbial diversity in *I. basta*.

Given the apparent absence of NOB in *I. basta*, the observed production of nitrate (Fig. 5) is difficult to explain. One possibility is that some NO released by either the host or a host-associated microorganism is detoxified to nitrate by another member of the sponge holobiont. However, we did not detect genes with homology to those encoding the NO detoxifying enzyme, flavohemoglobin-NO-dioxygenase (Hmp), which converts NO together with O_2 to NO_3^- (Gardner, 2012 and references therein), in our microbial metagenomic datasets. Still, NO dioxygenases (NODs), could be encoded in the sponge genome since neuroglobin-like sequences are encoded in other sponges (Lechauve *et al*., 2013) and members of this enzyme family have been shown to have NOD activity *in vitro* (Brunori *et al*., 2005). Furthermore, a NADH-cytochrome *b5* reductase, which belongs to the same family as Hmp, was found to be highly expressed at all life stages of the sponge *Amphimedon queenslandica* (Conaco *et al*., 2012). Alternatively, partial nitrite oxidation may be catalysed by free-living NOB in the seawater, although nitrification rates in the control were negligible.

Conclusions and outlook

In this study, we combined proteogenomic and experimental analyses to show that *Ca*. N. ianthellae, the first characterized but yet uncultured representative of a new genus within the thaumarchaeotes, is responsible for ammonia oxidation in the widespread marine sponge *I. basta*. Whilst *Ca*. N. ianthellae is equipped with the typical genetic repertoire of free-living AOA, it also exhibits a number of putative adaptations to a host-associated lifestyle. Several of these adaptive features were unique to *Ca*. N. ianthellae, whereas others were shared exclusively with the previously described sponge symbionts *Ca*. C. symbiosum (Hallam *et al*., 2006) and the AOA symbiont of the sponge *C. concentrica* (Moitinho-Silva *et al*., 2017a). Many of these putative adaptations to a sponge-associated lifestyle were not encoded in the recently sequenced genome of an AOA from a deep-sea glass sponge (DSGS-AOA; Tian *et al*., 2016), indicating that this archaeon is not an obligate symbiont or that thriving in a deep-sea glass sponge requires very different traits. Our results confirm an emerging view that marine sponge microbiomes tend to converge on a few shared functional traits, a process shaped by the environmental niche provided by the sponge host and governed by specific ecological factors such as high dissolved nutrient loads and frequent contact with resident and transient microorganisms (Liu *et al*., 2012). Whereas other studies have focused on this phenomenon by comparing functional convergence across entire sponge microbiomes (Fan *et al*., 2012; Liu *et al*., 2012; Horn *et al*., 2016), we demonstrate that in symbiotic marine sponge Thaumarchaeota, a similar evolutionary convergence is achieved that stands in contrast to their strictly chemoautotrophic non-symbiotic free-living relatives and that this process is likely not solely achieved by gene acquisition but rather by selective gene retention and gene family expansion.

With *I. basta* emerging as a model species for sponge symbiosis research, additional work should be undertaken to address the conspicuous absence of nitrite-oxidizing bacteria and ascertain why, in contrast to most tropical sponge species, *I. basta* hosts such a low diversity of microbial symbionts. Furthermore, hypotheses about the interaction of *Ca*. N. ianthellae with other members of the sponge holobiont should be experimentally confirmed. For example, assimilation of amino acids and peptides should be tested using stable isotope probing and the function of unique serine proteases, serpins and ELPs should be analysed via heterologous gene expression to reveal mechanistic insights into the interaction of this archaeon with its host. Finally, the mechanism for symbiont acquisition should be assessed via screening for *Ca*. N. ianthellae in gametes and larvae.

Experimental procedures

Sponge collection

Large adult specimens (n = 4) of the sponge *Ianthella basta* were collected from Orpheus Island (18°36.878′S, 146°29.990′E), Queensland, Australia, cut into 10 cm x 10 cm explants and transferred to racks on the reef. After a 12 week healing period in the field, sponge clones were collected in two separate sampling trips by scuba diving between September and October 2011 and transported to the indoor temperature-controlled aquarium at the Australian Institute of Marine Science (AIMS), Townsville, where they were acclimated at ambient temperature (25°C) for 48 h and then randomly assigned to experimental treatments. The two adult *I. basta* specimens used for metaproteogenomic analyses were collected from Orpheus Island in October 2010 and 2011 at Orpheus Island. Upon sample collection, specimens were cut into small strips, immediately snap-frozen in liquid N_2 and subsequently stored at −80°C until DNA extraction.

Sponge incubations and nitrification rate measurements

To infer nitrification activity of the *I. basta* holobiont, a 7-day incubation experiment was performed using triplicate sponge clones under different ammonium concentrations: ambient (0.29 ± 0.1 μM), 25 and 100 μM NH_4^+. A schematic representation of the experimental design is given in Supporting Information Fig. S7. Specifically, sponge clones (1.5–13.0 g post-experimental wet weight; mean = 5.1; standard deviation = 2.2) were incubated at 25°C in the dark in 1.5 l acid-washed glass containers completely filled with 5 μm filtered seawater or with seawater amended with 25 or 100 μM NH_4Cl. Additionally, abiotic control experiments were performed with ambient seawater without sponge clones that were incubated for 24 h under identical conditions. Sponge clones were transferred to a new container with fresh seawater every 24 h during the 7-day incubation, in order to reduce the effects of O_2 depletion and NH_4^+ accumulation. To assess carbon fixation in concert with nitrification, $NaH^{13}CO_3^-$ (100 μM; 99% ^{13}C) was added every 12 h to all experimental treatments. To prevent loss of ^{13}C-labelled CO_2, containers were closed for 6 h after the $NaH^{13}CO_3^-$ addition and subsequently opened until the next $NaH^{13}CO_3^-$ addition to avoid oxygen depletion. The effect of such intermittently closed containers on nitrification activity was assessed in a control experiment by comparing ambient seawater and 100 μM NH_4Cl amended seawater in constantly open containers (Supporting Information Fig. S8). Nitrification activity was determined as gross nitrification rates using a ^{15}N isotope pool dilution technique (Inselsbacher *et al*., 2007), and net fluxes of individual dissolved inorganic N species (DIN), namely NH_4^+, NO_2^-, NO_3^-. Gross nitrification rates were also measured on day 1, 3 and 5, where at the beginning of each day $K^{15}NO_3$ was added to a final $^{15}NO_3^-$ concentration of 1% – 10% of the nitrate pool (Murphy *et al*., 2003) in the filtered ambient seawater. Gross nitrification rates were calculated from seawater samples collected after a short equilibration following ^{15}N-label addition (0 h) and after 20 h. Net fluxes of NH_4^+, NO_2^- and NO_3^- were measured on day 1, 3, 5 and 7 and were calculated as change in concentration over time (i.e., between 6 and 18 h after ^{15}N-label addition to avoid biased net NO_3^- fluxes through a stimulation of consumptive processes from the added $^{15}NO_3^-$). Net nitrification refers to the change in concentration of NO_2^- + NO_3^-. Furthermore, we used the AOA-specific inhibitor PTIO (2-phenyl-4,4,5,5-tetramethylimidazoline-1-oxyl 3-oxide; Tokyo Chemical Industry) (Martens-Habbena *et al*., 2015) in additional parallel incubations, which was added at a concentration of 75 μM at the beginning of day 3 and 7.

Seawater samples (10 ml) were taken from each aquarium and filtered using 0.45 μm Sartorius Minisart cellulose acetate filters (Göttingen, Germany). Duplicate samples for dissolved inorganic nitrogen (NH_4^+, NO_2^-, NO_3^-) were measured on a Seal AA3 segmented flow analyser and referenced against OSIL standards and in-house reference samples. For ^{15}N-analysis of NO_2^- + NO_3^-, sample water was filtered through pre-combusted GF/Fs (Whatman International; treated for 4 h at 450°C), and subsequently through 0.2 μm filters (Sartorius). All samples were immediately frozen at −20°C for later analysis. Prior to shipment to the University of Vienna, samples were thawed at room temperature and the microbial inhibitor pheynylmercuric acetate was added (to a final concentration of 10 μM). Upon arrival in Vienna the samples were promptly stored at −80°C. Nitrite and nitrate were isolated together from seawater

by sequential microdiffusion (Sørensen and Jensen, 1991). To remove ammonium from sample water, 100 mg MgO, and an acid trap (acidified cellulose filter disc enclosed in a semi-permeable Teflon membrane) was added to 9 ml of sample and 1.5 ml of 3 M KCl. After 5 days shaking at 35°C, the acid traps were removed, and 50 mg of Devarda's alloy was added along with a new acid trap, and shaken at 35°C for 7 days. Devarda's alloy is a reducing catalyst converting both NO_2^- and NO_3^- to NH_4^+ and the subsequently formed NH_3 was collected in the acid trap. Acid traps were dried over concentrated sulfuric acid and analysed for ^{15}N by an elemental analyser (EA 1110, CE Instruments, Milan, Italy) coupled to an isotope ratio mass spectrometer (IRMS; Finnigan MAT Delta^{Plus} IRMS with a Finnigan MAT ConFlo III interface). Gross nitrification rates were calculated based on Wanek *et al*. (2010). Gross and net rates are expressed as μmol nitrogen species per gram wet weight sponge per day (μmol N g^{-1} d^{-1}).

For the determination of ^{13}C enrichment in whole sponge tissue at the end of the incubation, sponge tissue was freeze-dried, ground to a fine powder, and stored at dry conditions prior to analysis. The δ ^{13}C values of sponge tissue were determined using an EA-IRMS system as described above. Carbon fixation rates were calculated based on the δ ^{13}C values treatments amended with 100 μM $NaH^{13}CO_3^-$. For the calculations, we used a background (natural abundance) bicarbonate concentration of 1975 μM.

DNA extraction from whole sponge tissue and qPCR for symbiont quantification

Between 80 and 150 mg of *I. basta* tissue was thawed, rinsed successively (3x) in 1X calcium- and magnesium-free artificial seawater (CMF-ASW) and immediately ground into a paste with a mortar and pestle in liquid N_2. After resuspension in TE buffer (10 mM Tris–HCL, pH 8.0, 1 mMEDTA), DNA was extracted from the suspension using an adapted SDS-based isolation method (Zhou *et al*., 1996) and using 1% polyvinylpyrrolidone. DNA was extracted from the two individuals used for metaproteogenomics (see below) as well as three healthy individuals from a previous study (Luter *et al*., 2010). Additionally, DNA was extracted from a subset of sponge clones that were subjected to nitrification incubations. Quantitative PCR (qPCR) was used to estimate the number of specific thaumarchaeal, as well as α- and γ-proteobacterial symbionts by quantifying the 16S rRNA gene using specific primers designed for each symbiont phylotype. The following primer sets were thus used for the dominant *I. basta* thaumarchaeal, and α-, and γ-proteobacterial symbionts respectively; IBthaum16S_523F, 5′-CCG TAG CCT GCC CTG TTA G-3′, IBthaum16S_727R, 5′-GCT TTC ATC CCT CAC CGT-3′; IBalpha16S_1010F, 5′-CGG AGA CGC TTC CTT CG –3′, IBalpha16S_1206R, 5'-GCC CAG CCC ATA AAT GC-3′; IBgamma16S_466F, 5′-TAC CCY TGY GTT TTG ACG-3′, IBgamma16S_655R, 5′- CCR CTT CTC TCT RCC ATA C-3′. An iCycler real-time PCR system (Bio-Rad, Hercules, CA) was used to measure all samples, in duplicate wells per reaction and reactions were performed in a 25 μl volume with 1 μl of DNA template. All symbiont 16S rRNA gene qPCR assays used SYBR®Green reaction mixtures containing 12.5 μl iQ SYBR®Green Supermix (Bio-Rad), and optimized concentrations of 400 nM primer as well as 0.25 mg ml^{-1} BSA. Cycling conditions were 95°C for 5 min followed by 40 cycles of 95°C for 40 s, 58°C for 30 s, and 72°C for 40 s. Fluorescence intensities were measured after each cycle, and a final elongation at 72°C was followed by a melting curve analysis from 55–95°C in 10 s increments of 0.5°C.

Standard curves were generated for each primer set using serial dilutions of a standard containing a known number of the target sequences. Standards used the M13 primer set to amplify 16S rRNA gene clones derived from the sponge symbionts. PCR products were visualized on an agarose gel, purified separately using the QIAquick PCR Purification Kit (Qiagen), followed by fluorometric quantification of DNA concentrations using PicoGreen (Molecular Probes, Eugene, OR) and a NanoDrop ND-3300 Fluorospectrometer (NanoDrop). Gene abundance was calculated based on DNA concentration and product size. Dilution series ranging from 10^6 to 10^0 copies μl^{-1} were used to generate standard curves. Final 16S rRNA gene abundances for each microbial symbiont were then normalized by the wet weight of the sponge tissue used for DNA extraction.

Cryosectioning and FISH for symbiont quantification

The *I. basta* individual collected for metagenomic sequencing in October 2010 was also assessed using FISH. Briefly, after sample collection, the *I. basta* specimen was cut into tissue strips of ~2 mm^3, fixed in 4% PFA for 1 h at room temperature and stored in ethyl alcohol (EtOH)-phosphate-buffered saline (PBS) at –20°C. For FISH, PFA-fixed samples of *I. basta* were embedded in Neg-50 (Richard-Allan Scientific), and cut to 5-μm sections (Leica CM3050 S). A double-labelled Arch915 probe in Cy3 (Thermo Fisher Scientific, Waltham, MA, USA) was used for the microscopic visualization and quantification of the thaumarchaeal symbiont of *I. basta*. To calculate the relative abundance of the thaumarchaeal symbiont, equimolar amounts of the double-labelled probes EUB338-I, EUB338-II, and EUB338-III (Fluos and Cy5) (Stoecker *et al*., 2010) were used for quantification of most bacteria. Hybridizations were prepared with an equimolar mixture of both probes and using 25% formamide in the hybridization buffer, with the stringency of the washing buffer adjusted

accordingly (Daims *et al*., 2001). As a negative control, the non-EUB338-I (reverse complementary probe to EUB338-I) was applied on one section per well per slide hybridized (Wallner *et al*., 1993). All hybridized samples were analysed with a confocal laser scanning microscope (CLSM) (LSM 510 Meta; Zeiss, Oberkochen, Germany). Archaeal and bacterial cells were counted by eye on 10 randomly selected images derived from multiple tissue sections obtained from a single *I. basta* individual, and the proportion of archaeal cells to total prokaryotic cells was calculated.

Microbial cell enrichment for metaproteogenomics

To separate symbiont cells from host tissue prior to DNA extraction, ~15.7 g wet weight of *I. basta* was rinsed successively (3x) in 1X CMF-ASW. Sponge tissue was cut into small pieces (<1 cm^3), ground on ice in 1X CMF-ASW with a mortar and pestle, transferred into a glass douncer on ice, and sponge tissue was dissociated through shear force and vortexing. The supernatant was transferred into multiple Eppendorf microcentrifuge tubes, centrifuged at 39 x g for 15 min at 4°C to pellet larger sponge particles, filtered through a 5 μm Sartorius filter and centrifuged at 10,844 × g for 15 min at 4°C to pellet microbial cells. Microscopic examination of the pellet revealed an enrichment of microbial cells and an absence of sponge nuclei. For the sponge individual sampled in October 2010, DNA was extracted from this pellet using the procedure referenced above. For the individual sampled in October 2011, the pellets were resuspended in 1X CMF-ASW and the suspension was layered in 1.8 ml amounts onto 6.5 ml cushions of 30% Gastrografin dissolved in 1X CMF-ASW + 0.2 M EDTA, and centrifuged at 40,008 x g for 1 h at 4°C in a Beckman L-100 XP ultracentrifuge with the SW 41 Ti swing rotor (Beckman Coulter). Following centrifugation, a cell-rich layer above the 30% Gastrografin cushion ('Fraction A') and the cell-rich pellet ('Fraction B') were carefully removed, resuspended in 10 mM Tris, and re-pelleted at 24,400 × g for 15 min at 4°C in a new solution of 10 mM Tris.

The same *I. basta* individual sampled in October 2011 was used for proteomic analyses. However, in order to minimize protein degradation, biomass preparation methods were shortened and modified. The first cell fraction used for proteomic analysis involved rinsing in 1X CMF-ASW and tissue homogenization using a mortar/pestle and glass douncer as described above. However, the tissue homogenization steps occurred in 1X TE buffer with Roche Complete Protease Inhibitor (Roche). Supernatant from the glass douncer was immediately collected into 2 ml Eppendorf tubes and frozen at −80°C before shipment on dry ice to Greifswald, Germany, for protein extraction and downstream analyses. A second cell fraction comprised a 5 μm filtrate from the supernatant in the glass douncer. Finally, a crude sponge homogenate sample was obtained by direct grinding of sponge tissue (without rinsing in 1X CMF-ASW) in liquid N_2 followed by freezing at −80°C in 1X TE with Roche Complete. These three fractions could thus be characterized as a sponge homogenate without sponge skeleton, a sponge homogenate without sponge nuclei, and a direct sponge homogenate respectively.

DNA extraction, library preparation and sequencing

DNA was extracted from the individual sampled in October 2011 using the FastDNA spin kit for soil (MP Biomedicals, Solon, OH, USA) from the cell-rich layers above the 30% Gastrografin cushion and in the ultracentrifuged pellet according to the manufacturer's instructions. Sequencing libraries were prepared using the Nextera kit (Illumina) according to the manufacturer's instructions and concentrations measured using the QuantIT kit (Molecular Probes, Life Technologies, Naerum, Denmark). The libraries were paired-end (2 × 150 bp) sequenced on an Illumina HiSeq2000 using the TruSeq PE Cluster Kit v3-cBot-HS and TruSeq SBS kit v.3-HS sequencing kit and on an Illumina MiSeq using v3 2 × 300 bp kits.

For the individual sampled in October 2010, cell pellets were pooled from the 5 μm filtration step and DNA was extracted using the modified protocol of Zhou *et al*. (1996) described above. Metagenomic sequences were then generated at the Ramaciotti Sequencing Centre (Sydney, Australia) using the GS FLX instrument using Titanium chemistry (Roche) on a 454 half-sequencing-plate (454 Life Sciences, Branford, CT, USA).

Metagenome assembly and genome binning

Metagenome reads in fastq format, obtained from the Illumina sequencing runs, were end-trimmed at a minimum phred score of 15, a minimum length of 50 bp, allowing no ambiguous nucleotides and Illumina sequencing adaptors removed. Trimmed reads from each dataset were assembled using Spades version 3.11.0 (Bankevich *et al*., 2012), using default parameters and genomes were binned using Metabat v 2.12.0 (Kang *et al*., 2015). MAGs from multiple Illumina data sets were dereplicated using dRep (Olm *et al*., 2017) and provisionally classified using CheckM (Parks *et al*., 2015). A single high-quality archaeal MAG was recovered after dereplication and uploaded to MaGe (Vallenet *et al*., 2009) for annotation. This MAG has been submitted to the European Nucleotide Archive with the accession number UYNY01000000 under BioProject PRJEB29556.

For the 454-pyrosequencing run, artificially amplified reads were dereplicated using CD-HIT (Li and Godzik,

2006), assembled with MIRA (Chevreux *et al*.,) and binned with a genome-specific Phymm model (Brady and Salzberg, 2009), which was trained from the contigs that contained phylogenetic marker genes. The MAG derived from this pyrosequencing run was similarly GC-rich (64.2%), but considerably larger (6.58 Mb), slightly less complete (98.06%), and considerably more fragmented (2508 scaffolds). This MAG is available on the MaGe platform as '*Ianthella basta symbiont* thaum' at http://www.genoscope.cns.fr/agc/microscope/home/index.php.

Comparative genomics

The annotated archaeal MAG was downloaded from MaGe and compared to published thaumarchaeotal genomes (Supporting Information Table S5) using genomic average nucleotide identity (gANI), average amino acid identity (AAI), and through construction of orthologous gene families. For all analyses, annotated genes were supplemented with additional gene calls predicted by Prodigal (Hyatt *et al*., 2010). gANI was calculated with MiSI (Varghese *et al*., 2015). AAI (Konstantinidis and Tiedje, 2005) was calculated using bidirectional best blastp hits (Camacho *et al*., 2009) that aligned over at least 70% of gene length with average identity values weighted according to gene length. Orthologous gene families were constructed using Orthofinder (Emms and Kelly, 2015). For functional annotation of eukaryotic-like proteins (ELPs) and mobile/selfish genetic elements, predicted genes from all sequenced thaumarchaeal genomes were searched against the Protein Family A (Pfam-A) database (v31.0) (Finn *et al*., 2014) using Hmmer 3 (hmmer.org.) and the gathering threshold option (–cutga). Results were screened for the presence of domains associated with ELPs and mobile/selfish genetic elements.

Phylogenetic analyses

Bayesian trees were constructed using Phylobayes v 4.1c (Lartillot and Philippe, 2004) using the best model identified for each dataset by ModelFinder (Kalyaanamoorthy *et al*., 2017). Phylogenomic reconstruction was based on a concatenated amino-acid alignment of 34 marker genes constructed with CheckM (Parks *et al*., 2015) with 10 independent runs of 11,000 generations under the LG4 model. Six thousand generations of each independent run were discarded as burn-in and the remaining trees from each run were pooled for calculation of a consensus tree and for determining posterior branch support. For the 16S rRNA gene phylogenies, top representative hits in a blastn query against the Genbank nr database of the full-length sequence from *Ca*. N. ianthellae and *Ca*. *C. symbiosum*, along with sequences from sequenced thaumarchaeal genomes were aligned with SINA (Pruesse *et al*., 2012) and analysed further as described above. Ten independent runs of 30,000 generations under the GTR model were used. Then, 7500 generations of each independent run were discarded as burn-ins and the remaining trees from each run were pooled for calculation of a consensus tree and for determining posterior branch support.

Protein identification and proteome analyses

1D PAGE followed by liquid chromatography-based mass spectrometry (1D-PAGE-LC–MS/MS) were used for protein and peptide separation and identification as described previously (Washburn *et al*., 2001; Otto *et al*., 2010), with slight modifications. MS spectra and MS/MS spectra were acquired from eluting peptides ionized with electrospray ionization (ESI) and analysed in a LTQ Orbitrap Velos hybrid mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA), as described previously (Verberkmoes *et al*., 2009; Otto *et al*., 2010), with minor modifications. Samples of the sponge homogenate without sponge skeleton and crude sponge homogenate processed in liquid N_2 were analysed in technical duplicates, whereas the sponge homogenate without sponge nuclei was analysed only once. All MS/MS spectra were searched against predicted protein sequence databases composed of the *I. basta* symbiont-enriched metagenome bins and common laboratory contaminants using the Sorcerer SEQUEST (v.27, rev. 11) algorithm. The CD-HIT software (Li and Godzik, 2006) was used to remove redundancies from the database due to the potential for strain-level redundancies. Protein identifications were filtered with Scaffold version 3.5.1 applying the 'sequest' filter criteria described previously (Heinz *et al*. 2012). For protein identification only peptides identified with high mass accuracy (maximum $\pm$10 ppm difference between calculated and observed mass) were considered and at least two exclusively unique peptides were required to identify a protein. False-discovery rates (FDRs) were estimated with searches against a target-decoy database as described previously (Peng *et al*., 2003; Elias and Gygi, 2007). Peptide FDRs were between 2.5% and 3.1%, and protein FDRs were below 0.4% throughout all samples. For relative quantitation of proteins, normalized spectral abundance factor were calculated for each sample according to the method of Florens *et al*. (2006) and averaged for all replicates and samples. The proteomics raw data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD012484 and 10.6019/PXD012484.

Acknowledgements

This work was supported by the Marie Curie Initial Training Network - SYMBIOMICS. In addition, F.U.M., C.W.H, M.M. and M.W. were supported by the European Research Council Advanced Grant project NITRICARE 294343. M.A. and

P.H.N. were supported by research grants (15510 and 16578) from VILLUM FONDEN. Many thanks to Christian Hentschker for Orbitrap Velos measurements and database searches. We thank Bernhard Schink for valuable advice regarding provisional naming of the *I. basta* AOA symbiont.

References

Adams, M.D., Wagner, L.M., Graddis, T.J., Landick, R., Antonucci, T.K., Gibson, A.L., and Oxender, D.L. (1990) Nucleotide sequence and genetic characterization reveal six essential genes for the LIV-I and LS transport systems of *Escherichia coli*. *J Biol Chem* **265**: 11436–11443.

Agarwal, V., Blanton, J.M., Podell, S., Taton, A., Schorn, M. A., Busch, J., *et al*. (2017) Metagenomic discovery of polybrominated diphenyl ether biosynthesis by marine sponges. *Nat Chem Biol* **13**: 537–543.

Ahlgren, N.A., Chen, Y., Needham, D.M., Parada, A.E., Sachdeva, R., Trinh, V., *et al*. (2017) Genome and epigenome of a novel marine Thaumarchaeota strain suggest viral infection, phosphorothioation DNA modification and multiple restriction systems: genome and epigenome of a new marine thaumarchaeon. *Environ Microbiol* **19**: 2434–2452.

Alonso-Saez, L., Waller, A.S., Mende, D.R., Bakker, K., Farnelid, H., Yager, P.L., *et al*. (2012) Role for urea in nitrification by polar marine Archaea. *Proc Natl Acad Sci U S A* **109**: 17989–17994.

Amano, F., and Noda, T. (1995) Improved detection of nitric oxide radical (NO.) production in an activated macrophage culture with a radical scavenger, carboxy PTIO and Griess reagent. *FEBS Lett* **368**: 425–428.

Andersson, M., Brion, N., and Middelburg, J. (2006) Comparison of nitrifier activity versus growth in the Scheldt estuarya turbid, tidal estuary in northern Europe. *Aquat Microb Ecol* **42**: 149–158.

Arreguín, R., Arreguín, B., Soriano-García, M., Hernández-Arana, A., and Rodríguez-Romero, A. (1993) Isolation and characterization of a protease from the marine sponge *Spheciospongia vesparia*. *FEBS Lett* **320**: 235–238.

Astudillo-García, C., Bell, J.J., Webster, N.S., Glasl, B., Jompa, J., Montoya, J.M., and Taylor, M.W. (2017) Evaluating the core microbiota in complex communities: a systematic investigation. *Environ Microbiol* **19**: 1450–1462.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., *et al*. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**: 455–477.

Bayer, K., Schmitt, S., and Hentschel, U. (2007) Microbial nitrification in Mediterranean sponges: possible involvement of ammonium-oxidizing *Betaproteobacteria*. In *Porifera Research: Biodiversity, Innovation, Sustainability*, Custódio, M., Lôbo-Hajdu, G., Hajdu, E., and Muricy, G. (eds). Brazil: Museu Nacional: Série Livros. Rio de Janeiro, pp. 165–171.

Bayer, K., Schmitt, S., and Hentschel, U. (2008) Physiology, phylogeny and *in situ* evidence for bacterial and archaeal nitrifiers in the marine sponge *Aplysina aerophoba*. *Environ Microbiol* **10**: 2942–2955.

Bayer, B., Vojvoda, J., Offre, P., Alves, R.J.E., Elisabeth, N. H., Garcia, J.A., *et al*. (2016) Physiological and genomic characterization of two novel marine thaumarchaeal strains indicates niche differentiation. *ISME J* **10**: 1051–1063.

Beam, J.P., Jay, Z.J., Kozubal, M.A., and Inskeep, W.P. (2014) Niche specialization of novel Thaumarchaeota to oxic and hypoxic acidic geothermal springs of Yellowstone National Park. *ISME J* **8**: 938–951.

Beckers, G., Bendt, A.K., Kramer, R., and Burkovski, A. (2004) Molecular identification of the urea uptake system and transcriptional analysis of urea transporter- and urease-encoding genes in *Corynebacterium glutamicum*. *J Bacteriol* **186**: 7645–7652.

Bell, J.J. (2008) The functional roles of marine sponges. *Estuar Coast Shelf Sci* **79**: 341–353.

Bergquist, P.R., and Kelly-Borges, M. (1995) Systematics and biogeography of the genus *Ianthella* (Demospongiae: Verongida: Ianthellidae) in the south-west Pacific. *The Beagle, Records of the Museums and Art Galleries of the Northern Territory* **12**: 151–176.

Boardman, G.D., Starbuck, S.M., Hudgins, D.B., Li, X., and Kuhn, D.D. (2004) Toxicity of ammonia to three marine fish and three marine invertebrates. *Environ Toxicol* **19**: 134–142.

de Boer, W., and Laanbroek, H.J. (1989) Ureolytic nitrification at low pH by *Nitrosospira* spec. *Arch Microbiol* **152**: 178–181.

Brady, A., and Salzberg, S.L. (2009) Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. *Nat Methods* **6**: 673–676.

Brezonik, P.L., and Lee, G.F. (1968) Dentrification as a nitrogen sink in Lake Mendota, Wisconsin. *Environ Sci Technol* **2**: 120–125.

Brunori, M., Giuffre, A., Nienhaus, K., Nienhaus, G.U., Scandurra, F.M., and Vallone, B. (2005) Neuroglobin, nitric oxide, and oxygen: functional pathways and conformational changes. *Proc Natl Acad Sci U S A* **102**: 8483–8488.

Burton, S.A.Q., and Prosser, J.I. (2001) Autotrophic ammonia oxidation at low pH through urea hydrolysis. *Appl Environ Microbiol* **67**: 2952–2957.

Callebaut, I., Mornon, J.-P., Gilgès, D., and Vigon, I. (2000) HYR, an extracellular module involved in cellular adhesion and related to the immunoglobulin-like fold. *Protein Sci* **9**: 1382–1390.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Camargo, J.A., and Alonso, Á. (2006) Ecological and toxicological effects of inorganic nitrogen pollution in aquatic ecosystems: a global assessment. *Environ Int* **32**: 831–849.

Carini, P., Dupont, C.L., and Santoro, A.E. (2018) Patterns of thaumarchaeal gene expression in culture and diverse marine environments. *Environ Microbiol* **20**: 2112–2124.

Caranto, J.D., and Lancaster, K.M. (2017) Nitric oxide is an obligate bacterial nitrification intermediate produced by hydroxylamine oxidoreductase. *Proc Natl Acad Sci U S A* **114**: 8217–8222.

Chevreux, B., Wetter, T., and Suhai, S. (1999) Genome sequence assembly using trace signals and additional sequence information. German Conference on Bioinformatics GCB'99 GCB. *Comput Sci Biol Proc* pp. 45–56.

Conaco, C., Neveu, P., Zhou, H., Arcila, M., Degnan, S.M., Degnan, B.M., and Kosik, K.S. (2012) Transcriptome profiling of the demosponge *Amphimedon queenslandica*

reveals genome-wide events that accompany major life cycle transitions. *BMC Genomics* **13**: 209.
Corredor, J.E., Wilkinson, C.R., Vicente, V.P., Morell, J.M., and Otero, E. (1988) Nitrate release by Caribbean reef sponges. *Limnol Oceanogr* **33**: 114–120.
Cort, J.R., Yee, A., Edwards, A.M., Arrowsmith, C.H., and Kennedy, M.A. (2000) Structure-based functional classification of hypothetical protein MTH538 from *Methanobacterium thermoautotrophicum*. *J Mol Biol* **302**: 189–203.
Daims, H., Ramsing, N.B., Schleifer, K.-H., and Wagner, M. (2001) Cultivation-independent, semiautomatic determination of absolute bacterial cell numbers in environmental samples by fluorescence in situ hybridization. *Appl Environ Microbiol* **67**: 5810–5818.
Daims, H., Lebedeva, E.V., Pjevac, P., Han, P., Herbold, C., Albertsen, M., *et al*. (2015) Complete nitrification by Nitrospira bacteria. *Nature* **528**: 504–509.
Díez-Vives, C., Moitinho-Silva, L., Nielsen, S., Reynolds, D., and Thomas, T. (2017) Expression of eukaryotic-like protein in the microbiome of sponges. *Mol Ecol* **26**: 1432–1451.
Drapeau, G.R. (1980) Substrate specificity of a proteolytic enzyme isolated from a mutant of *Pseudomonas fragi*. *J Biol Chem* **255**: 839–840.
Elias, J.E., and Gygi, S.P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **4**: 207–214.
Ellis, A., Lu, H., Li, C.G., and Rand, M.J. (2001) Effects of agents that inactivate free radical NO (NO center dot) on nitroxyl anion-mediated relaxations, and on the detection of NO center dot released from the nitroxyl anion donor Angeli's salt. *Brit J Pharmacol* **134**: 521–528.
Emms, D.M., and Kelly, S. (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**: 157.
Essuman, K., Summers, D.W., Sasaki, Y., Mao, X., Yim, A. K.Y., DiAntonio, A., and Milbrandt, J. (2018) TIR domain proteins are an ancient family of NAD(+)-consuming enzymes. *Curr Biol* **28**: 421–430.
Fan, L., Reynolds, D., Liu, M., Stark, M., Kjelleberg, S., Webster, N.S., and Thomas, T. (2012) Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proc Natl Acad Sci U S A* **109**: E1878–E1887.
Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S.R., *et al*. (2014) Pfam: the protein families database. *Nucleic Acids Res* **42**: D222–D230.
Fiore, C.L., Labrie, M., Jarett, J.K., and Lesser, M.P. (2015) Transcriptional activity of the giant barrel sponge, *Xestospongia muta* Holobiont: molecular evidence for metabolic interchange. *Front Microbiol* **6**: 364.
Florens, L., Carozza, M., Swanson, S., Fournier, M., Coleman, M., Workman, J., and Washburn, M. (2006) Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods* **40**: 303–311.
Freckelton, M.L., Luter, H.M., Andreakis, N., Webster, N.S., and Motti, C.A. (2012) Qualitative variation in colour morphotypes of *Ianthella basta* (Porifera: Verongida). *Hydrobiologia* **687**: 191–203.
Freeman, M.F., Vagstad, A.L., and Piel, J. (2016) Polytheonamide biosynthesis showcasing the metabolic potential of sponge-associated uncultivated 'Entotheonella' bacteria. *Curr Opin Chem Biol* **31**: 8–14.
Gardner, P.R. (2012) Hemoglobin: a nitric-oxide dioxygenase. *Scientifica* **2012**, 1–34.
Gauthier, M.-E.A., Watson, J.R., and Degnan, S.M. (2016) Draft genomes shed light on the dual bacterial symbiosis that dominates the microbiome of the coral reef sponge *Amphimedon queenslandica*. *Front Mar Sci* **3**: 196.1–196.18.
de Goeij, J.M., van Oevelen, D., Vermeij, M.J.A., Osinga, R., Middelburg, J.J., de Goeij, A.F.P.M., and Admiraal, W. (2013) Surviving in a marine desert: the sponge loop retains resources within coral reefs. *Science* **342**: 108–110.
Guy, L., and Ettema, T.J.G. (2011) The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol* **19**: 580–587.
Hallam, S.J., Konstantinidis, K.T., Putnam, N., Schleper, C., Watanabe, Y., Sugahara, J., *et al*. (2006) Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc Natl Acad Sci U S A* **103**: 18296–18301.
Hatzenpichler, R., Lebedeva, E.V., Spieck, E., Stoecker, K., Richter, A., Daims, H., and Wagner, M. (2008) A moderately thermophilic ammonia-oxidizing crenarchaeote from a hot spring. *Proc Natl Acad Sci U S A* **105**: 2134–2139.
Heinz, E., Williams, T.A., Nakjang, S., Noël, C.J., Swan, D.C., Goldberg, A.V., *et al*. (2012) The genome of the obligate intracellular parasite *Trachipleistophora hominis*: new insights into microsporidian genome dynamics and reductive evolution. *PLoS Pathog* **8**: e1002979.
Hentschel, U., Piel, J., Degnan, S.M., and Taylor, M.W. (2012) Genomic insights into the marine sponge microbiome. *Nat Rev Microbiol* **10**: 641–654.
Herndl, G.J., Reinthaler, T., Teira, E., van Aken, H., Veth, C., Pernthaler, A., and Pernthaler, J. (2005) Contribution of *Archaea* to total prokaryotic production in the deep Atlantic Ocean. *Appl Environ Microbiol* **71**: 2303–2309.
Hoffmann, F., Larsen, O., Thiel, V., Rapp, H.T., Pape, T., Michaelis, W., and Reitner, J. (2005) An anaerobic world in sponges. *Geomicrobiol J* **22**: 1–10.
Hoffmann, F., Radax, R., Woebken, D., Holtappels, M., Lavik, G., Rapp, H.T., *et al*. (2009) Complex nitrogen cycling in the sponge *Geodia barretti*. *Environ Microbiol* **11**: 2228–2243.
Horn, H., Slaby, B.M., Jahn, M.T., Bayer, K., Moitinho-Silva, L., Förster, F., *et al*. (2016) An enrichment of CRISPR and other defense-related features in marine sponge-associated microbial metagenomes. *Front Microbiol* **7**: 1751.
Hoshino, T. (1979) Transport systems for branched-chain amino acids in *Pseudomonas aeruginosa*. *J Bacteriol* **139**: 705–712.
Hosie, A.H.F., Allaway, D., Galloway, C.S., Dunsby, H.A., and Poole, P.S. (2002) *Rhizobium leguminosarum* has a second general amino acid permease with unusually broad substrate specificity and high similarity to branched-chain amino acid transporters (Bra/LIV) of the ABC family. *J Bacteriol* **184**: 4071–4080.
Hosseinzadeh, P., Tian, S., Marshall, N.M., Hemp, J., Mullen, T., Nilges, M.J., *et al*. (2016) A purple cupredoxin from *Nitrosopumilus maritimus* containing a mononuclear

type 1 copper center with an open binding site. *J Am Chem Soc* **138**: 6324–6327.
Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.
Ingalls, A.E., Shah, S.R., Hansman, R.L., Aluwihare, L.I., Santos, G.M., Druffel, E.R.M., and Pearson, A. (2006) Quantifying archaeal community autotrophy in the mesopelagic ocean using natural radiocarbon. *Proc Natl Acad Sci U S A* **103**: 6442–6447.
Inselsbacher, E., Cambui, C.A., Richter, A., Stange, C.F., Mercier, H., and Wanek, W. (2007) Microbial activities and foliar uptake of nitrogen in the epiphytic bromeliad *Vriesea gigantea*. *New Phytol* **175**: 311–320.
Jompa, J., and McCook, L.J. (2002) The effects of nutrients and herbivory on competition between a hard coral (*Porites cylindrica*) and a brown alga (*Lobophora variegata*). *Limnol Oceanogr* **47**: 527–534.
Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**: 587–589.
Kang, D.D., Froula, J., Egan, R., and Wang, Z. (2015) MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**: e1165.
Kerou, M., Offre, P., Valledor, L., Abby, S.S., Melcher, M., Nagler, M., *et al*. (2016) Proteomics and comparative genomics of *Nitrososphaera viennensis* reveal the core genome and adaptations of archaeal ammonia oxidizers. *Proc Natl Acad Sci U S A* **113**: E7937–E7946.
Kim, J.-G., Park, S.-J., Sinninghe Damsté, J.S., Schouten, S., Rijpstra, W.I.C., Jung, M.-Y., *et al*. (2016) Hydrogen peroxide detoxification is a key mechanism for growth of ammonia-oxidizing archaea. *Proc Natl Acad Sci U S A* **113**: 7888–7893.
Kits, K.D., Jung, M.Y., Vierheilig, J., Pjevac, P., Sedlacek, C. J., Liu, S., *et al*. (2019) Low yield and abiotic origin of N_2O formed by the complete nitrifier *Nitrospira inopinata*. *Nat Commun* **10**: 1836.
Kitzinger, K., Padilla, C.C., Marchant, H.K., Hach, P.F., Herbold, C.W., Kidane, A.T., *et al*. (2019) Cyanate and urea are substrates for nitrification by Thaumarchaeota in the marine environment. *Nat Microbiol* **4**: 234–243.
Koch, H., Galushko, A., Albertsen, M., Schintlmeister, A., Gruber-Dorninger, C., Lucker, S., *et al*. (2014) Growth of nitrite-oxidizing bacteria by aerobic hydrogen oxidation. *Science* **345**: 1052–1054.
Könneke, M., Bernhard, A.E., de la Torre, J.R., Walker, C.B., Waterbury, J.B., and Stahl, D.A. (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**: 543–546.
Konstantinidis, K.T., and Tiedje, J.M. (2005) Towards a genome-based taxonomy for prokaryotes. *J Bacteriol* **187**: 6258–6264.
Konstantinidis, K.T., Rosselló-Móra, R., and Amann, R. (2017) Uncultivated microbes in need of their own taxonomy. *ISME J* **11**: 2399–2406.
Kozlowski, J.A., Stieglmeier, M., Schleper, C., Klotz, M.G., and Stein, L.Y. (2016) Pathways and key intermediates required for obligate aerobic ammonia-dependent chemolithotrophy in bacteria and Thaumarchaeota. *ISME J* **10**: 1836–1845.
Lam, P., Jensen, M.M., Kock, A., Lettmann, K.A., Plancherel, Y., Lavik, G., *et al*. (2011) Origin and fate of the secondary nitrite maximum in the Arabian Sea. *Biogeosciences* **8**: 1565–1577.
Lartillot, N., and Philippe, H. (2004) A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol* **21**: 1095–1109.
Lechauve, C., Jager, M., Laguerre, L., Kiger, L., Correc, G., Leroux, C., *et al*. (2013) Neuroglobins, pivotal proteins associated with emerging neural systems and precursors of metazoan globin diversity. *J Biol Chem* **288**: 6957–6967.
Lehtovirta-Morley, L.E., Stoecker, K., Vilcinskas, A., Prosser, J.I., and Nicol, G.W. (2011) Cultivation of an obligate acidophilic ammonia oxidizer from a nitrifying acid soil. *Proc Natl Acad Sci U S A* **108**: 15892–15897.
Li, B., Lopes, J.S., Foster, P.G., Embley, T.M., and Cox, C.J. (2014) Compositional biases among synonymous substitutions cause conflict between gene and protein trees for plastid origins. *Mol Biol Evol* **31**: 1697–1709.
Li, M., Baker, B.J., Anantharaman, K., Jain, S., Breier, J.A., and Dick, G.J. (2015) Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. *Nat Commun* **6**: 8933.
Li, Y., Ding, K., Wen, X., Zhang, B., Shen, B., and Yang, Y. (2016) A novel ammonia-oxidizing archaeon from wastewater treatment plant: its enrichment, physiological and genomic characteristics. *Sci Rep-UK* **6**: 23747.
Li, P.-N., Herrmann, J., Tolar, B.B., Poitevin, F., Ramdasi, R., Bargar, J.R., *et al*. (2018) Nutrient transport suggests an evolutionary basis for charged archaeal surface layer proteins. *ISME J* **12**: 2389–2402.
Lin, X., Handley, K.M., Gilbert, J.A., and Kostka, J.E. (2015) Metabolic potential of fatty acid oxidation and anaerobic respiration by abundant members of Thaumarchaeota and Thermoplasmata in deep anoxic peat. *ISME J* **9**: 2740–2744.
Liu, M., Fan, L., Zhong, L., Kjelleberg, S., and Thomas, T. (2012) Metaproteogenomic analysis of a community of sponge symbionts. *ISME J* **6**: 1515–1525.
Lloyd, K.G., Schreiber, L., Petersen, D.G., Kjeldsen, K.U., Lever, M.A., Steen, A.D., *et al*. (2013) Predominant archaea in marine sediments degrade detrital proteins. *Nature* **496**: 215–218.
Love, G.D., Grosjean, E., Stalvies, C., Fike, D.A., Grotzinger, J. P., Bradley, A.S., *et al*. (2009) Fossil steroids record the appearance of Demospongiae during the Cryogenian period. *Nature* **457**: 718–721.
Luo, C., Rodriguez-R, L.M., and Konstantinidis, K.T. (2014) MyTaxa: an advanced taxonomic classifier for genomic and metagenomic sequences. *Nucleic Acids Res* **42**: e73–e73.
Lurie-Weinberger, M.N., Gomez-Valero, L., Merault, N., Glöckner, G., Buchrieser, C., and Gophna, U. (2010) The origins of eukaryotic-like proteins in *Legionella pneumophila*. *Int J Med Microbiol* **300**: 470–481.
Luter, H.M., Whalan, S., and Webster, N.S. (2010) Exploring the role of microorganisms in the disease-like syndrome affecting the sponge *Ianthella basta*. *Appl Environ Microbiol* **76**: 5736–5744.

Luter, H.M., Whalan, S., and Webster, N.S. (2012) Thermal and sedimentation stress are unlikely causes of brown spot syndrome in the coral reef sponge, *Ianthella basta*. *PLoS ONE* **7**: e39779.

Maldonado, M., Ribes, M., and van Duyl, F.C. (2012) Nutrient fluxes through sponges: biology, budgets, and ecological implications. *Adv Mar Biol* **62**: 113–182.

Maixner, F., Noguera, D.R., Anneser, B., Stoecker, K., Wegl, G., Wagner, M., and Daims, H. (2006) Nitrite concentration influences the population structure of Nitrospira-like bacteria. *Environ Microbiol* **8**: 1487–1495.

Martens-Habbena, W., Berube, P.M., Urakawa, H., de la Torre, J.R., and Stahl, D.A. (2009) Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. *Nature* **461**: 976–979.

Martens-Habbena, W., Qin, W., Horak, R.E.A., Urakawa, H., Schauer, A.J., Moffett, J.W., *et al*. (2015) The production of nitric oxide by marine ammonia-oxidizing archaea and inhibition of archaeal ammonia oxidation by a nitric oxide scavenger: marine AOA produce NO and are inhibited by NO scavenger. *Environ Microbiol* **17**: 2261–2274.

Mehrshad, M., Amoozegar, M.A., Ghai, R., Shahzadeh Fazeli, S.A., and Rodriguez-Valera, F. (2016) Genome reconstruction from metagenomic data sets reveals novel microbes in the brackish waters of the Caspian Sea. *Appl Environ Microbiol* **82**: 1599–1612.

Miller, W.T. (2012) Tyrosine kinase signaling and the emergence of multicellularity. *BBA-Mol Cell Res* **1823**: 1053–1057.

Mohamed, N.M., Saito, K., Tal, Y., and Hill, R.T. (2010) Diversity of aerobic and anaerobic ammonia-oxidizing bacteria in marine sponges. *ISME J* **4**: 38–48.

Moitinho-Silva, L., Díez-Vives, C., Batani, G., Esteves, A.I., Jahn, M.T., and Thomas, T. (2017a) Integrated metabolism in sponge–microbe symbiosis revealed by genome-centered metatranscriptomics. *ISME J* **11**: 1651–1666.

Moitinho-Silva, L., Nielsen, S., Amir, A., Gonzalez, A., Ackermann, G.L., Cerrano, C., *et al*. (2017b) The sponge microbiome project. *GigaScience* **6**: 1–7.

Morley, S.A., Berman, J., Barnes, D.K.A., de Juan Carbonell, C., Downey, R.V., and Peck, L.S. (2016) Extreme phenotypic plasticity in metabolic physiology of Antarctic demosponges. *Front Ecol Evol* **3**: 157.

Müller, W.E.G., Ushijima, H., Batel, R., Krasko, A., Borejko, A., Müller, I.M., and Schröder, H.-C. (2006) Novel mechanism for the radiation-induced bystander effect: nitric oxide and ethylene determine the response in sponge cells. *Mutat Res* **597**: 62–72.

Murphy, D., Recous, S., Stockdale, E., Fillery, I.R., Jensen, L., Hatch, D., and Goulding, K.W. (2003) Gross nitrogen fluxes in soil: theory, measurement and application of 15N pool dilution techniques. *Adv Agron* **79**: 69–118.

Mußmann, M., Brito, I., Pitcher, A., Sinninghe Damste, J.S., Hatzenpichler, R., Richter, A., *et al*. (2011) Thaumarchaeotes abundant in refinery nitrifying sludges express *amoA* but are not obligate autotrophic ammonia oxidizers. *Proc Natl Acad Sci U S A* **108**: 16771–16776.

Nakagawa, T., and Stahl, D.A. (2013) Transcriptional response of the archaeal ammonia oxidizer *Nitrosopumilus maritimus* to low and environmentally relevant ammonia concentrations. *Appl Environ Microbiol* **79**: 6911–6916.

Newstead, S. (2015) Molecular insights into proton coupled peptide transport in the PTR family of oligopeptide transporters. *BBA - Gen Subjects* **1850**: 488–499.

Off, S., Alawi, M., and Spieck, E. (2010) Enrichment and physiological characterization of a novel Nitrospira-like bacterium obtained from a marine sponge. *Appl Environ Microbiol* **76**: 4640–4646.

Offre, P., Kerou, M., Spang, A., and Schleper, C. (2014) Variability of the transporter gene complement in ammonia-oxidizing archaea. *Trends Microbiol* **22**: 665–675.

Olm, M.R., Brown, C.T., Brooks, B., and Banfield, J.F. (2017) dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* **11**: 2864–2868.

Otto, A., Bernhardt, J., Meyer, H., Schaffer, M., Herbst, F.-A., Siebourg, J., *et al*. (2010) Systems-wide temporal proteomic profiling in glucose-starved *Bacillus subtilis*. *Nat Commun* **1**: 137.

Ouverney, C.C., and Fuhrman, J.A. (2000) Marine planktonic archaea take up amino acids. *Appl Environ Microbiol* **66**: 4829–4833.

Palatinszky, M., Herbold, C., Jehmlich, N., Pogoda, M., Han, P., von Bergen, M., *et al*. (2015) Cyanate as an energy source for nitrifiers. *Nature* **524**: 105–108.

Park, S.-J., Ghai, R., Martín-Cuadrado, A.-B., Rodríguez-Valera, F., Chung, W.-H., Kwon, K., *et al*. (2014) Genomes of two new ammonia-oxidizing archaea enriched from deep marine sediments. *PLoS ONE* **9**: e96449.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**: 1043–1055.

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., and Hugenholtz, P. (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* **36**: 996–1004.

Passmore, I.J., Nishikawa, K., Lilley, K.S., Bowden, S.D., Chung, J.C.S., and Welch, M. (2015) Mep72, a metzincin protease that is preferentially secreted by biofilms of *Pseudomonas aeruginosa*. *J Bacteriol* **197**: 762–773.

Patterson, N.J., Günther, J., Gibson, A.J., Offord, V., Coffey, T.J., Splitter, G., *et al*. (2014) Two TIR-like domain containing proteins in a newly emerging zoonotic *Staphylococcus aureus* strain sequence type 398 are potential virulence factors by impacting on the host innate immune response. *Front Microbiol* **5**: 662.

Peng, J., Elias, J.E., Thoreen, C.C., Licklider, L.J., and Gygi, S.P. (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res* **2**: 43–50.

Pester, M., Schleper, C., and Wagner, M. (2011) The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. *Curr Opin Microbiol* **14**: 300–306.

Phillips, G., Swairjo, M.A., Gaston, K.W., Bailly, M., Limbach, P.A., Iwata-Reuyl, D., and de Crécy-Lagard, V. (2012) Diversity of archaeosine synthesis in crenarchaeota. *ACS Chem Biol* **7**: 300–305.

Picossi, S., Montesinos, M.L., Pernil, R., Lichtlé, C., Herrero, A., and Flores, E. (2005) ABC-type neutral amino

acid permease N-I is required for optimal diazotrophic growth and is repressed in the heterocysts of *Anabaena* sp. strain PCC 7120. *Mol Microbiol* **57**: 1582–1592.

Pruesse, E., Peplies, J., and Glöckner, F.O. (2012) SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**: 1823–1829.

Qin, W., Martens-Habbena, W., Kobelt, J.N., and Stahl, D.A. (2016) *Candidatus* Nitrosopumilaceae. In *Bergey's Manual of Systematics of Archaea and Bacteria*, Whitman, W.B., Rainey, F., Kämpfer, P., Trujillo, M., Chun, J., DeVos, P., *et al*. (eds). Chichester, UK: John Wiley & Sons, pp. 1–2.

Qin, W., Meinhardt, K.A., Moffett, J.W., Devol, A.H., Virginia Armbrust, E., Ingalls, A.E., and Stahl, D.A. (2017) Influence of oxygen availability on the activities of ammonia-oxidizing archaea: influence of oxygen availability. *Envirol Microbiol Rep* **9**: 250–256.

Qin, W., Amin, S.A., Martens-Habbena, W., Walker, C.B., Urakawa, H., Devol, A.H., *et al*. (2014) Marine ammonia-oxidizing archaeal isolates display obligate mixotrophy and wide ecotypic variation. *Proc Natl Acad Sci U S A* **111**: 12504–12509.

Radax, R., Hoffmann, F., Rapp, H.T., Leininger, S., and Schleper, C. (2012a) Ammonia-oxidizing archaea as main drivers of nitrification in cold-water sponges: Archaeal nitrification in cold-water sponges. *Environ Microbiol* **14**: 909–923.

Radax, R., Rattei, T., Lanzen, A., Bayer, C., Rapp, H.T., Urich, T., and Schleper, C. (2012b) Metatranscriptomics of the marine sponge *Geodia barretti*: tackling phylogeny and function of its microbial community. *Environ Microbiol* **14**: 1308–1324.

Reveillaud, J., Maignien, L., Eren, A.M., Huber, J.A., Apprill, A., Sogin, M.L., and Vanreusel, A. (2014) Host-specificity among abundant and rare taxa in the sponge microbiome. *ISME J* **8**: 1198–1209.

Reynolds, D., and Thomas, T. (2016) Evolution and function of eukaryotic-like proteins from sponge symbionts. *Mol Ecol* **25**: 5242–5253.

Riesgo, A., Farrar, N., Windsor, P.J., Giribet, G., and Leys, S.P. (2014) The analysis of eight transcriptomes from all poriferan classes reveals surprising genetic complexity in sponges. *Mol Biol Evol* **31**: 1102–1120.

Sabina, J., and Söll, D. (2006) The RNA-binding PUA domain of archaeal tRNA-guanine transglycosylase is not required for archaeosine formation. *J Biol Chem* **281**: 6993–7001.

Santoro, A.E., Dupont, C.L., Richter, R.A., Craig, M.T., Carini, P., McIlvin, M.R., *et al*. (2015) Genomic and proteomic characterization of " *Candidatus* Nitrosopelagicus brevis": an ammonia-oxidizing archaeon from the open ocean. *Proc Natl Acad Sci U S A* **112**: 1173–1178.

Sauder, L.A., Engel, K., Lo, C.-C., Chain, P., and Neufeld, J.D. (2018) "Candidatus Nitrosotenuis aquarius," an ammonia-oxidizing archaeon from a freshwater aquarium biofilter. *Appl Environ Microbiol* **84**: e01430–e01418.

Sauder, L.A., Albertsen, M., Engel, K., Schwarz, J., Nielsen, P. H., Wagner, M., and Neufeld, J.D. (2017) Cultivation and characterization of Candidatus Nitrosocosmicus exaquare, an ammonia-oxidizing archaeon from a municipal wastewater treatment system. *ISME J* **11**: 1142–1157.

Schaefer, S.C., and Hollibaugh, J.T. (2017) Temperature decouples ammonium and nitrite oxidation in coastal waters. *Environ Sci Technol* **51**: 3157–3164.

Schläppy, M.-L., Schöttner, S.I., Lavik, G., Kuypers, M.M.M., de Beer, D., and Hoffmann, F. (2010) Evidence of nitrification and denitrification in high and low microbial abundance sponges. *Mar Biol* **157**: 593–602.

Schleper, C., DeLong, E.F., Preston, C.M., Feldman, R.A., Wu, K.Y., and Swanson, R.V. (1998) Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon *Cenarchaeum symbiosum*. *J Bacteriol* **180**: 5003–5009.

Schmitt, S., Tsai, P., Bell, J., Fromont, J., Ilan, M., Lindquist, N., *et al*. (2012) Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. *ISME J* **6**: 564–576.

Sharp, K.H., Eam, B., Faulkner, D.J., and Haygood, M.G. (2007) Vertical transmission of diverse microbes in the tropical sponge *Corticium* sp. *Appl Environ Microbiol* **73**: 622–629.

Shen, T., Stieglmeier, M., Dai, J., Urich, T., and Schleper, C. (2013) Responses of the terrestrial ammonia-oxidizing archaeon *Ca*. Nitrososphaera viennensis and the ammonia-oxidizing bacterium *Nitrosospira multiformis* to nitrification inhibitors. *FEMS Microbiol Let* **344**: 121–129.

Shigeno-Nakazawa, Y., Kasai, T., Ki, S., Kostyanovskaya, E., Pawlak, J., Yamagishi, J., *et al*. (2016) A pre-metazoan origin of the CRK gene family and co-opted signaling network. *Sci Rep-UK* **6**: 3439.

Simion, P., Philippe, H., Baurain, D., Jager, M., Richter, D.J., Di Franco, A., *et al*. (2017) A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. *Curr Biol* **27**: 958–967.

Simister, R.L., Deines, P., Botté, E.S., Webster, N.S., and Taylor, M.W. (2012) Sponge-specific clusters revisited: a comprehensive phylogeny of sponge-associated microorganisms. *Environ Microbiol* **14**: 517–524.

Sørensen, P., and Jensen, E.S. (1991) Sequential diffusion of ammonium and nitrate from soil extracts to a polytetrafluoroethylene trap for 15N determination. *Anal Chim Acta* **252**: 201–203.

Southwell, M.W., Popp, B.N., and Martens, C.S. (2008) Nitrification controls on fluxes and isotopic composition of nitrate from Florida keys sponges. *Mar Chem* **108**: 96–108.

Spang, A., Poehlein, A., Offre, P., Zumbrägel, S., Haider, S., Rychlik, N., *et al*. (2012) The genome of the ammonia-oxidizing *Candidatus* Nitrososphaera gargensis: insights into metabolic versatility and environmental adaptations. *Environ Microbiol* **14**: 3122–3145.

Steger, D., Ettinger-Epstein, P., Whalan, S., Hentschel, U., de Nys, R., Wagner, M., and Taylor, M.W. (2008) Diversity and mode of transmission of ammonia-oxidizing archaea in marine sponges. *Environ Microbiol* **10**: 1087–1094.

Stoecker, K., Dorninger, C., Daims, H., and Wagner, M. (2010) Double labeling of oligonucleotide probes for fluorescence in situ hybridization (DOPE-FISH) improves signal intensity and increases rRNA accessibility. *Appl Environ Microbiol* **76**: 922–926.

Tabares, P., Pimentel-Elardo, S.M., Schirmeister, T., Hünig, T., and Hentschel, U. (2011) Anti-protease and immunomodulatory activities of bacteria associated with Caribbean sponges. *Marine Biotechnol* **13**: 883–892.

Taylor, M.W., Radax, R., Steger, D., and Wagner, M. (2007) Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol Mol Biol Rev* **1**: 295–347.

Teira, E., van Aken, H., Veth, C., and Herndl, G.J. (2006) Archaeal uptake of enantiomeric amino acids in the meso- and bathypelagic waters of the North Atlantic. *Limnol Oceanogr* **51**: 60–69.

Thiaville, J.J., Kellner, S.M., Yuan, Y., Hutinet, G., Thiaville, P.C., Jumpathong, W., *et al*. (2016) Novel genomic Island modifies DNA with 7-deazaguanine derivatives. *Proc Natl Acad Sci U S A* **113**: E1452–E1459.

Thomas, T., Moitinho-Silva, L., Lurgi, M., Björk, J.R., Easson, C., Astudillo-García, C., *et al*. (2016) Diversity, structure and convergent evolution of the global sponge microbiome. *Nat Commun* **7**: 11870.

Tian, R.-M., Sun, J., Cai, L., Zhang, W.-P., Zhou, G.-W., Qiu, J.-W., and Qian, P.-Y. (2016) The deep-sea glass sponge *Lophophysema eversa* harbours potential symbionts responsible for the nutrient conversions of carbon, nitrogen and sulfur. *Environ Microbiol* **18**: 2481–2494.

Tong, T., Chen, S., Wang, L., Tang, Y., Ryu, J.Y., Jiang, S., *et al*. (2018) Occurrence, evolution, and functions of DNA phosphorothioate epigenetics in bacteria. *Proc Natl Acad Sci U S A* **115**: E2988–E2996.

Tourna, M., Stieglmeier, M., Spang, A., Könneke, M., Schintlmeister, A., Urich, T., *et al*. (2011) *Nitrososphaera viennensis*, an ammonia oxidizing archaeon from soil. *Proc Natl Acad Sci U S A* **108**: 8420–8425.

Trame, C.B., Chang, Y., Axelrod, H.L., Eberhardt, R.Y., Coggill, P., Punta, M., and Rawlings, N.D. (2014) New minizincin structures provide a minimal scaffold for members of this metallopeptidase superfamily. *BMC Bioinformatics* **15**: 1.

Ueda, N., Richards, G.S., Degnan, B.M., Kranz, A., Adamska, M., Croll, R.P., and Degnan, S.M. (2016) An ancient role for nitric oxide in regulating the animal pelagobenthic life cycle: evidence from a marine sponge. *Sci Rep-UK* **6**: 37546.

Vacelet, J., Boury-Esnault, N., Fiala-Medioni, A., and Fisher, C.R. (1995) A methanotrophic carnivorous sponge. *Nature* **377**: 296–296.

Valladares, A., Montesinos, M.L., Herrero, A., and Flores, E. (2002) An ABC-type, high-affinity urea permease identified in cyanobacteria. *Mol Microbiol* **43**: 703–715.

Vallenet, D., Engelen, S., Mornico, D., Cruveiller, S., Fleury, L., Lajus, A., *et al*. (2009) MicroScope: a platform for microbial genome annotation and comparative genomics. *Database* **2009**: bap021.

Varghese, N.J., Mukherjee, S., Ivanova, N., Konstantinidis, K. T., Mavrommatis, K., Kyrpides, N.C., and Pati, A. (2015) Microbial species delineation using whole genome sequences. *Nucleic Acids Res* **43**: 6761–6771.

Ventura, M., Turroni, F., Motherway, M.O., MacSharry, J., and van Sinderen, D. (2012) Host–microbe interactions that facilitate gut colonization by commensal bifidobacteria. *Trends Microbiol* **20**: 467–476.

Verberkmoes, N.C., Russell, A.L., Shah, M., Godzik, A., Rosenquist, M., Halfvarson, J., *et al*. (2009) Shotgun metaproteomics of the human distal gut microbiota. *ISME J* **3**: 179–189.

Wallner, G., Amann, R., and Beisker, W. (1993) Optimizing fluorescent in situ hybridization with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms. *Cytometry* **14**: 136–143.

Wanek, W., Mooshammer, M., Blöchl, A., Hanreich, A., and Richter, A. (2010) Determination of gross rates of amino acid production and immobilization in decomposing leaf litter by a novel 15N isotope pool dilution technique. *Soil Biol Biochem* **42**: 1293–1302.

Wang, L., Chen, S., Xu, T., Taghizadeh, K., Wishnok, J.S., Zhou, X., *et al*. (2007) Phosphorothioation of DNA in bacteria by dnd genes. *Nat Chem Biol* **3**: 709–710.

Washburn, M.P., Wolters, D., and Yates, J.R. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**: 242–247.

Weber, E.B., Lehtovirta-Morley, L.E., Prosser, J.I., and Gubry-Rangin, C. (2015) Ammonia oxidation is not required for growth of group 1.1c soil Thaumarchaeota. *FEMS Microbiol Ecol* **91**: fiv001.

Webster, N.S., and Thomas, T. (2016) The sponge Hologenome. *MBio* **7**: e00135–e00116.

Webster, N.S., Taylor, M.W., Behnam, F., Lücker, S., Rattei, T., Whalan, S., *et al*. (2009) Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environ Microbiol* **12**: 2070–2082.

Wilkesman, J., and Schröder, H.C. (2002) Heat-stable protease from the marine sponge *Geodia cydonium*. *Cell Mol Bio (Noisy-le-grand)* **48**: 379–383.

Wilkesman, J.G., and Schröder, H.C. (2007) Analysis of serine proteases from marine sponges by 2-D zymography. *Electrophoresis* **28**: 429–436.

Wilkinson, C.R. (1983) Net primary productivity in coral reef sponges. *Science* **219**: 410–412.

Wilkinson, C.R., and Fay, P. (1979) Nitrogen fixation in coral reef sponges with symbiotic cyanobacteria. *Nature* **279**: 527–529.

Wilson, M.C., Mori, T., Rückert, C., Uria, A.R., Helf, M.J., Takada, K., *et al*. (2014) An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* **506**: 58–62.

Wuchter, C., Abbas, B., Coolen, M.J.L., Herfort, L., van Bleijswijk, J., Timmers, P., *et al*. (2006) Archaeal nitrification in the ocean. *Proc Natl Acad Sci U S A* **103**: 12317–12322.

Yan, J., Haaijer, S.C.M., Op den Camp, H.J.M., van Niftrik, L., Stahl, D.A., Könneke, M., *et al*. (2012) Mimicking the oxygen minimum zones: stimulating interaction of aerobic archaeal and anaerobic bacterial ammonia oxidizers in a laboratory-scale model system. *Environ Microbiol* **14**: 3146–3158.

You, D., Wang, L., Yao, F., Zhou, X., and Deng, Z. (2007) A novel DNA modification by sulfur: DndA is a NifS-like cysteine desulfurase capable of assembling DndC as an iron-sulfur cluster protein in *Streptomyces lividans*. *Biochemistry* **46**: 6126–6133.

Zhalnina, K.V., Dias, R., Leonard, M.T., Dorr de Quadros, P., Camargo, F.A.O., Drew, J.C., *et al*. (2014) Genome sequence of *Candidatus* Nitrososphaera evergladensis from group I.1b enriched from Everglades soil reveals novel genomic features of the ammonia-oxidizing archaea. *PLoS ONE* **9**: e101648.

Zaikova, E. (2014) *Molecular Mechanisms of Host-Symbiont Recognition in a Highly Specific Sponge-Archaeal Symbiosis*. PhD Thesis, University of British Columbia

Zhou, J., Bruns, M.A., and Tiedje, J.M. (1996) DNA recovery from soils of diverse composition. *Appl Environ Microbiol* **62**: 316–322.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix. Supporting Information.

Supporting Information

Characterization of a thaumarchaeal symbiont that drives incomplete nitrification in the tropical sponge *Ianthella basta*

Florian U. Moeller, Nicole S. Webster, Craig W. Herbold, Faris Behnam, Daryl Domman, Mads Albertsen, Maria Mooshammer[1], Stephanie Markert, Dmitrij Turaev, Dörte Becher, Thomas Rattei, Thomas Schweder, Andreas Richter, Margarete Watzka, Per Halkjaer Nielsen, and Michael Wagner

Table of contents:

Supporting Results and Discussion

High GC content of *Ca.* Nitrosospongia ianthellae. Both thaumarchaeal symbiont MAGs from *I. basta* possess the highest GC content (64.8%) of any genome-sequenced thaumarchaeote (Fig. 2C). While the GC content in the genome of *Ca.* Cenarchaeum symbiosum is similarly high (57.4%), the thaumarchaeal MAGs from the sponges *Cymbastela concentrica* (Moitinho-Silva *et al*., 2017a) and a glass sponge (Tian *et al*., 2016) had a much lower GC content (Fig. 2B). The high GC content found in the *I. basta* thaumarchaeal symbiont is consistent with the high GC content evident in

Mediterranean sponge metagenomes (GC content 58-63%), particularly when compared to seawater metagenomes (GC content 41%) collected at the same location (Horn *et al*., 2016). Similarly, the GC content of six sponge microbiome metagenomes (Fan *et al*., 2012) that we queried had an average GC content of 57.8% ± 6.7(SD), with three great barrier reef sponge species having an average microbiome metagenomic GC content of 63.3% ± 2.2 (SD). In contrast, the average GC composition of selected marine and coral microbiomes are ~48% and ~45%, respectively (Reichenberger *et al*., 2015). While several environmental factors are thought to affect the genomic GC content of bacteria and archaea (Foerstner *et al*., 2005; Wang *et al*., 2006), increased rates of homologous recombination via GC-biased gene conversion has recently been proposed as a crucial factor universally influencing the nucleotide content of microbial genes and genomes (Lassalle *et al*., 2015). The possibility that sponge microbiomes are hot spots for input of novel genetic material via lateral gene transfer events (Fan *et al*., 2012; Horn *et al*., 2016), and also display increased homologous recombination rates should be assessed in future work.

Environmental distribution of *Ca.* Nitrosospongia ianthellae. The *Ca*. N. ianthellae 16S rRNA gene was queried against the Sponge Microbiome Project (SMP) database containing amplicon data sets from 268 sponge species including *I. basta* (Moitinho-Silva *et al*., 2017b). The top hits were inserted into our reference 16S rRNA gene tree (Fig. 2A) using the Evolutionary Placement Algorithm (EPA; Berger *et al*., 2011). 76 OTUs placed adjacent to the sponge-specific sequence cluster 174, including one with 100% identity to *Ca.* N. ianthellae. These 76 OTUs comprised, on average, 17.2% ± 7.1 (SD) of all reads obtained from *I. basta* individuals in the SMP dataset, consistent with abundances determined by FISH and qPCR (see main text). Interestingly, the *Ca*. N. ianthellae-adjacent OTUs were also found to comprise 0.25 – 7.5% of the total reads obtained from *Ancorina alata*, *Stellata maori*, *Stellata aremaria* sampled in New Zealand and *Xestospongia exigua* sampled on the Great Barrier Reef. In *X. exigua*, one low abundance OTU (0.17% of all reads) had 100% nucleotide identity with the V4 region of the *Ca*. N. ianthellae 16S rRNA gene.

Among all environmental samples covered by the EMP, sequences highly similar or identical to *Ca*. N. ianthellae, with abundances above 0.1%, were exclusively found in sponges. Consistent with a habitat restricted to a few sponge species, no hits above 97.6% similarity to the 16S rRNA gene of *Ca*. N. ianthellae were detected by the Integrated Microbial NGS platform that queries most publicly available 16S rRNA gene amplicon data sets (but not the SMP dataset) (Lagkouvardos *et al*., 2016).

CO_2-fixation by *Ca.* N. ianthellae. As expected, *Ca.* N. ianthellae encodes all key enzymes of the thaumarchaeal 3-hydroxypropionate/4-hydroxybutyrate pathway (Könneke *et al*., 2014; Otte *et al*., 2015) and three proteins catalyzing five steps of this thaumarchaeal CO_2 fixation pathway were also detected within the metaproteome (Supporting Information Table S3).

Lack of archaellum genes in *Ca.* N. ianthellae. In contrast to several free-living AOA (Blainey *et al*., 2011; Jung *et al.*, 2014; Mosier *et al*., 2012; Spang *et al*., 2012; Bayer *et al*., 2016) and the DSGS-AOA, *Ca.* N. ianthellae, CCThau, and *Ca* C. symbiosum do not encode an archaellum. Lack of chemotaxis and motility traits has been described for other microbial symbionts (Karimi *et al*., 2018) and could represent an adaptation to the sponge-associated lifestyle, although also some non-sponge associated AOA lack an archaellum. The absence of an archaellum in *Ca.* N. ianthellae would also be consistent with vertical transmission of this symbiont, although this remains to be demonstrated as soon as larvae of *I. basta* will become available for experimentation.

Eukaryotic-like proteins (ELPs) in *Ca.* Nitrosospongia ianthellae. Four types of ELPs are found in *Ca*. N. ianthellae: Proteins with tetratricopeptide repeats (TPR), the Toll-interleukin-1 receptor (TIR) -like domain PF08937 (DUF1863; Cort *et al*., 2000), immunoglobulin-like (Ig-like) domains (DUF5011; Shigeno-Nakazawa *et al*., 2016), and hyaline repeats (HYR; Callebaut *et al*., 2000). The TPR were enriched in *Ca*. N. ianthellae when compared to other genome sequenced

thaumarchaeotes, while the TIR and Ig-like domains were exclusive to *Ca*. N. ianthellae (Fig. 3, Supporting Information Fig. S6). Of the TPR containing proteins in *Ca*. N. ianthellae, 41 represent TPR gene families not previously detected in other thaumarchaeotes. TPR-containing proteins, which mediate protein-protein interactions in eukaryotes (Blatch and Lässle, 1999), are thought to be important for the survival of sponge symbionts in their phagocytic hosts. Two TPR-containing proteins from sponge microbiomes cloned into *E. coli* affected amoeba phagocytosis (Reynolds and Thomas, 2016), however, both proteins contained the TPR-Sel1 motif which was absent in the TPR-containing proteins from *Ca*. N. ianthellae. TIR-like proteins have also been found in other marine sponges (Wiens *et al*., 2006; Gauthier *et al*., 2010), with these proteins known to be key mediators of the metazoan innate immune response as well as playing a role in regulating metabolic and bioenergetic pathways through modulating NAD^+ levels (Essuman *et al*., 2018). TIR-like proteins were expressed when sponges were subjected to bacteria-analogue lipoproteins (Wiens *et al*., 2006) and lipopolysaccharides (Wiens *et al*., 2005), which in turn caused the expression of a caspase likely involved in apoptosis and a macrophage-expressed protein, respectively. In this context, it is interesting to note that a protein encoding the DUF1863 domain from a zoonotic *Staphylococcus aureus* can decrease the survivability of mice infected by this strain (Patterson *et al*., 2014). Furthermore, the DUF1863 domain was recently shown to be a critical component of a bacterial defense system against myophages and may be involved in recognizing specific phage patterns (Doron *et al*., 2018).

An evolutionary homology between choanoflagellates and sponge choanocytes has long been speculated (Maldonado, 2005; Mah *et al*., 2014; Laundon et al., 2018). Both have diverse and abundant receptor tyrosine kinases (RTKs) (Srivastava *et al*., 2010; Miller, 2012), which are crucial components of metazoan signal transduction systems. Interestingly, choanoflagellate proteins with HYR-like domains were recently predicted to act as receptor tyrosine kinases (RTKs) (Manning *et al*., 2008). As HYR domains are structurally related to Ig and FN3 domains (Callebaut *et al*., 2000) and choanoflagellates lack the Ig domains found in many metazoan RTKs, the HYR domains may

be fulfilling the role of the Ig domains in metazoan RTKs. Furthermore, the Ig-like DUF5011 domain has been found on the extracellular portion of two choanoflagellate homologues of the tyrosine kinase substrate BCAR1 (Shigeno-Nakazawa *et al*., 2016). Consequently, the numerous DUF5011 and HYR domain containing proteins of *Ca.* N. ianthellae may be interacting with the host signaling network as *I. basta* likely harbors similar extracellular domains as part of its signal transduction and gene regulatory processes. This host-symbiont interaction is further supported by observations that (i) many *Ca.* N. ianthellae proteins encoding these domains are predicted to be exported and (ii) five extracellular proteins of the *Ca.* N. ianthellae exclusive gene family containing the DUF5011 domain were detected in the metaproteome (Supporting Information Table S3). On the other hand it is noteworthy that regarding ELPs the *Ca.* N. ianthellae genome is lacking the ankyrin repeat proteins, leucine-rich repeat proteins, protein tyrosine kinases, and armadillo-repeat proteins frequently detected in other symbiotic and pathogenic microbes (Fan *et al*., 2012; Jernigan and Bordenstein, 2015).

Supporting Experimental Procedures

Phylogenetic analyses. For the *amoA* gene tree, Bayesian trees were constructed using Phylobayes v 4.1c (Lartillot and Philippe, 2004) using the best model identified for each dataset by ModelFinder (Kalyaanmoorthy *et al*., 2017). *Ca*. N. ianthellae and *Ca.* C. symbiosum *amoA* sequences were placed into a reference tree, representing all OTU representatives of the curated database provided in Alves *et al*. (2018), using the Evolutionary Placement Algorithm (EPA; Berger *et al*., 2011) implemented in RAxML-HPC 8.2.11 (Stamatakis, 2014). Representative sequences clustering with the *Ca*. N. ianthellae *amoA* gene sequence along with *amoA* sequences from the aforementioned genome-sequenced thaumarchaeota were then aligned with MUSCLE (Edgar, 2004) and analyzed

S-layer proteins were identified in all sequenced thaumarchaeota based on orthologous groups identified with Orthofinder containing members previously identified as SLPs (Li *et al*., 2018), or also classified as arCOG08647 in EggNOG version 4.5 (Huerta-Cepas *et al*., 2016). A phylogenetic reconstruction was performed with a thaumarchaeal-specific dataset with a minimal sequence length of 300 amino acids. Sequences were aligned using mafft (Katoh and Standley, 2013) and automatically trimmed using trimAl version 1.4 (Capella-Gutiérrez *et al*., 2009) and the -gappyout function. After model selection using ModelFinder (Kalyaanmoorthy *et al*., 2017) and a maximum-likelihood amino acid phylogenetic tree was generated using the model LG+F+G4 in IQ-Tree, version 1.6.2 (Nguyen *et al*., 2015) with 1,000 ultrafast bootstraps (UFBoot).

Genes encoding S08A family endopeptidases were identified in all sequenced thaumarchaeota if they contained the PF00082 (Peptidase_S8) domain when searched against the Pfam-A database. The complete sequence set was identified using thaumarchaeal amino acid sequences as individual queries for blastp searches against the Genbank nr database and only top hits containing the PF00082 domain were included. In total 277 representative sequences of the S08A family endopeptidases were used for phylogenetic analyses and sequences were trimmed according to the presence of the PF00082 domain before alignment using mafft (Katoh and Standley, 2013). ModelFinder (Kalyaanmoorthy *et al*., 2017) was used for model selection and maximum-likelihood phylogenetic analyses implemented with IQ-Tree, version 1.6.2 (Nguyen *et al*., 2015), using the LG+I+G4 model and 1,000 ultrafast bootstraps (UFBoot). Phylogenetic analyses for serpins (PF00079) were conducted in a similar fashion and 75 representative sequences were analyzed (after filtering out 46 distant homologues) using the WAG+I+G4 model and 1,000 ultrafast bootstraps.

Phylogenetic analyses were conducted on amino acid sequences of the *livK* periplasmic branched chain amino acid transporter subunit as well as a concatenated alignment of the *livFGHMK* operon, using the best model identified for each dataset by ModelFinder (Kalyaanmoorthy *et al*., 2017) and implemented in IQ-Tree, version 1.6.2 (Nguyen *et al*., 2015)

with 1,000 ultrafast bootstraps (UFBoot). The complete sequence set was identified using individual *Ca*. N. ianthellae *livFGHMK* operon subunits as individual queries for blastp searches against the Genbank nr database. Top hits were included along with additional sequences demonstrated to be functional for active substrate transport within the hydrophobic amino-acid uptake transporter (HAAT) family (TCDB:3.A.14; Saier *et al*., 2009). Sequences were aligned using mafft (Katoh and Standley, 2013) and automatically trimmed using trimAl version 1.4 (Capella-Gutiérrez *et al*., 2009) and the -gappyout function. Models used for maximum-likelihood phylogenetic analyses were LG+F+G4 and LG+F+I+G4 for the *livK* subunit and the concatenated *livFGHMK* operon, respectively.

Database screening for 16S rRNA gene amplicons. A reference set of 16S rRNA gene sequences (N=65) from sequenced thaumarchaeotal genomes and amplicons associated with sponges, including *Ca.* N. bastadensis were aligned with SINA (Pruesse *et al*., 2012) and used to construct a reference tree in RaXML (Stamatakis, 2014). Sequence tags identified by the Sponge Microbiome Project (Moitinho-Silva *et al*., 2017b) were mapped to the reference set using blastn (Camacho *et al*., 2009), requiring at least 70% alignment and 90% identity. Successfully mapped reads were then aligned to the reference alignment using SINA and placed into the reference tree using RaXML-EPA (Berger *et al*, 2011). In addition, 16S rRNA gene sequences related to *Ca.* N. bastadensis were identified in short read archive (SRA) datasets using IMNGS (www.imngs.org - Lagkouvardos *et al*., 2016) with default parameters.

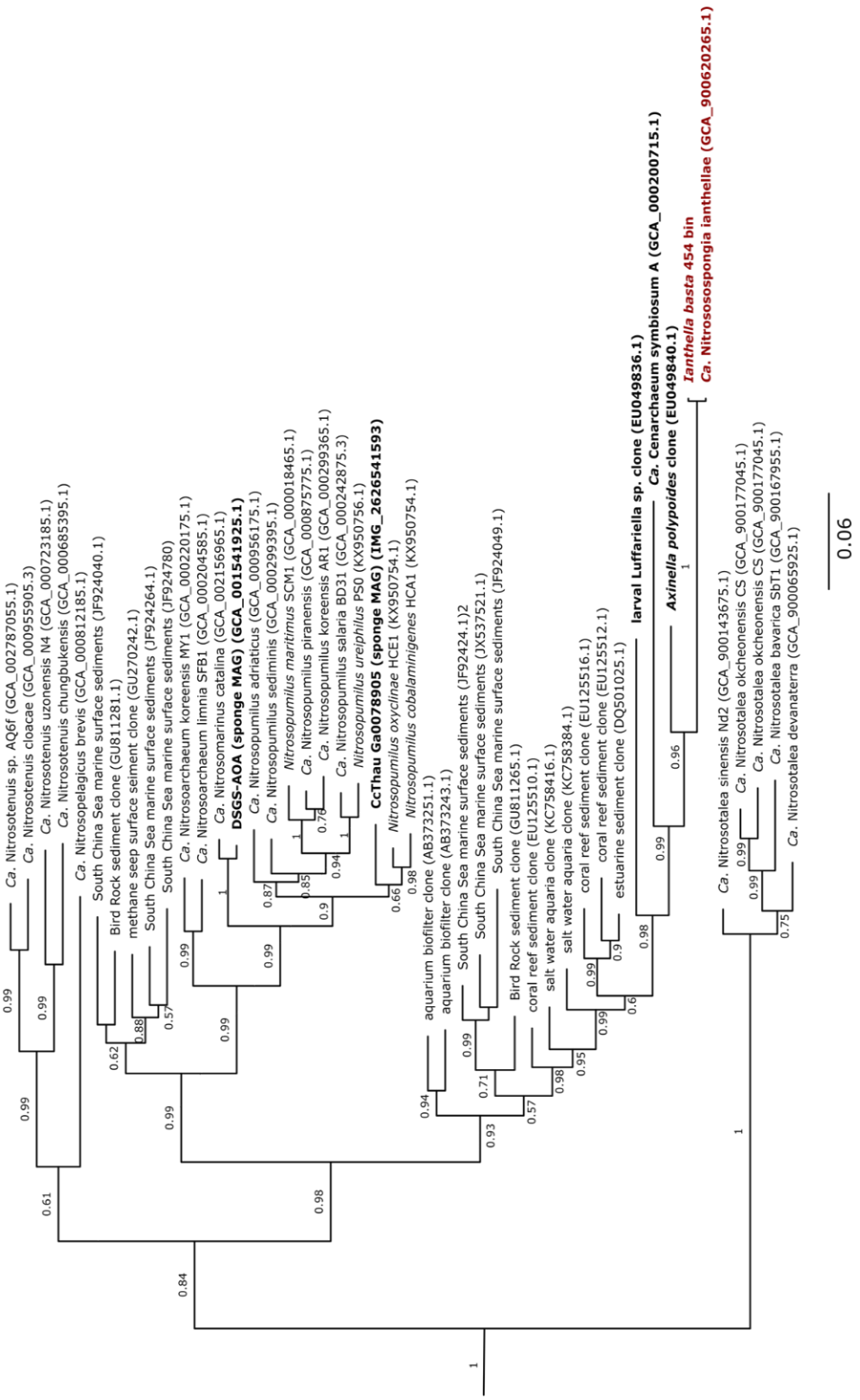


Figure S1. Bayesian *amoA* gene tree. For this analysis, *Ca*. N. ianthellae and *Ca.* C. symbiosum *amoA* sequences were placed into a reference tree, representing all OTU representatives of the curated database provided in Alves *et al.* (2018), using the Evolutionary Placement Algorithm (EPA; Berger *et al*., 2011) implemented in RAxML-HPC 8.2.11 (Stamatakis, 2014). Representative sequences clustering with the sponge symbiont *amoA* sequences were then analyzed along with *amoA* sequences from the other genome-sequenced AOA. Note that *Ca.* Nitrosotalea okcheonensis possesses two *amoA* gene copies (Herbold *et al*., 2017). Outgroups for both trees consisted of all three genome-sequenced members of the *Nitrososphaera* cluster, both members of the *Nitrosocosmicus* clade, and *Ca.* Nitrosocaldus icelandicus. In all trees, sequences obtained from sponges are depicted in bold.

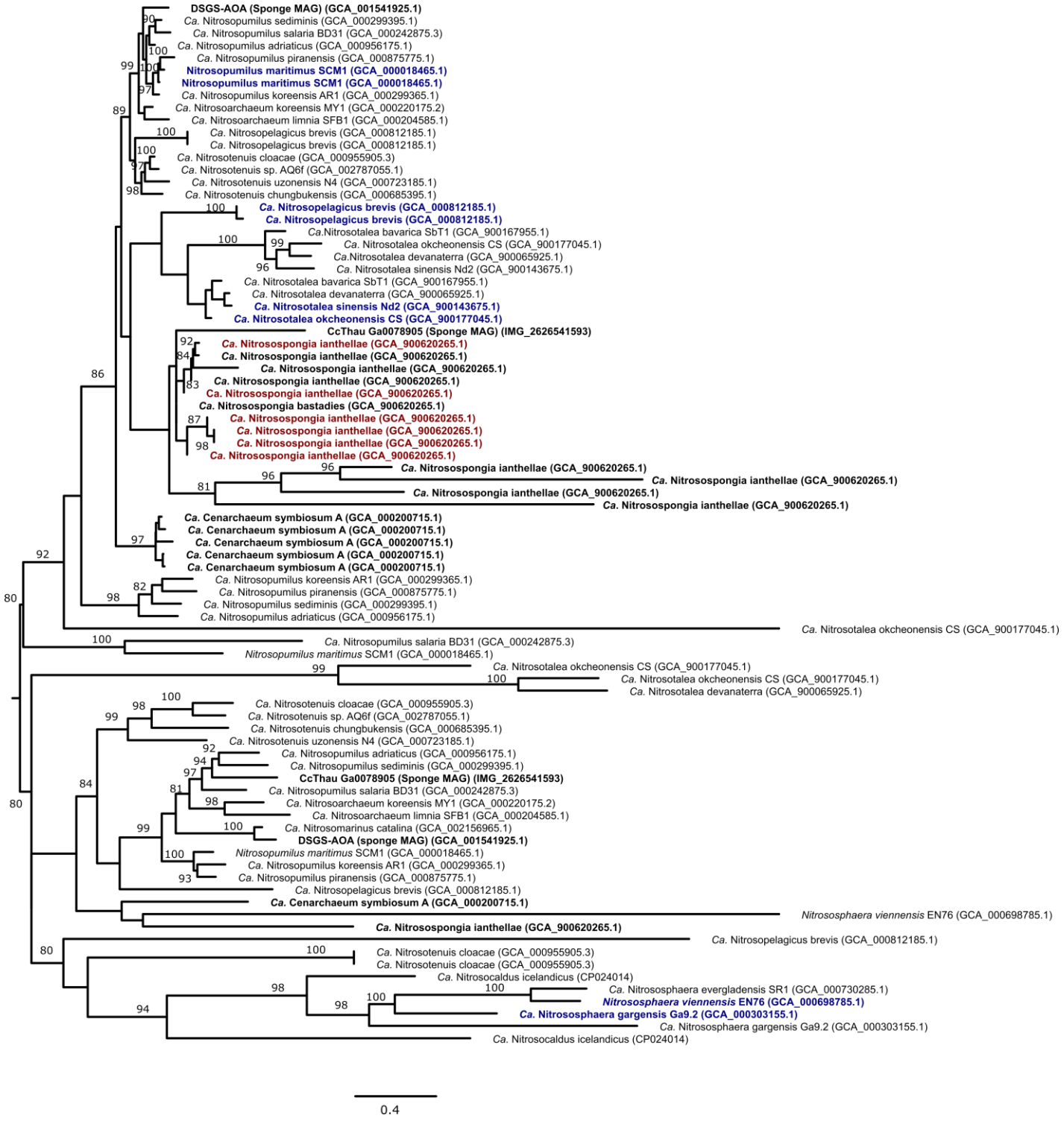


Figure S2. Maximum-likelihood amino acid phylogenetic tree (after automatic model selection with IQ-Tree, version 1.6.2) of putative S-layer proteins found in all analyzed thaumarchaeal genomes (28 in total). Sponge-derived sequences are depicted in bold black, while highly expressed proteins from *Ca*. N. ianthellae are depicted in bold red (6 in total). Highlighted in bold blue are putative S-layer proteins found to be highly expressed in previous proteomics studies (Santoro *et al*., 2015; Palatinszky *et al*., 2015; Kerou *et al*., 2016; Qin *et al*., 2017; Herbold *et al*., 2017). Values at nodes represent ultrafast bootstraps (UFBoot) with only values ≥80% shown for each branch.

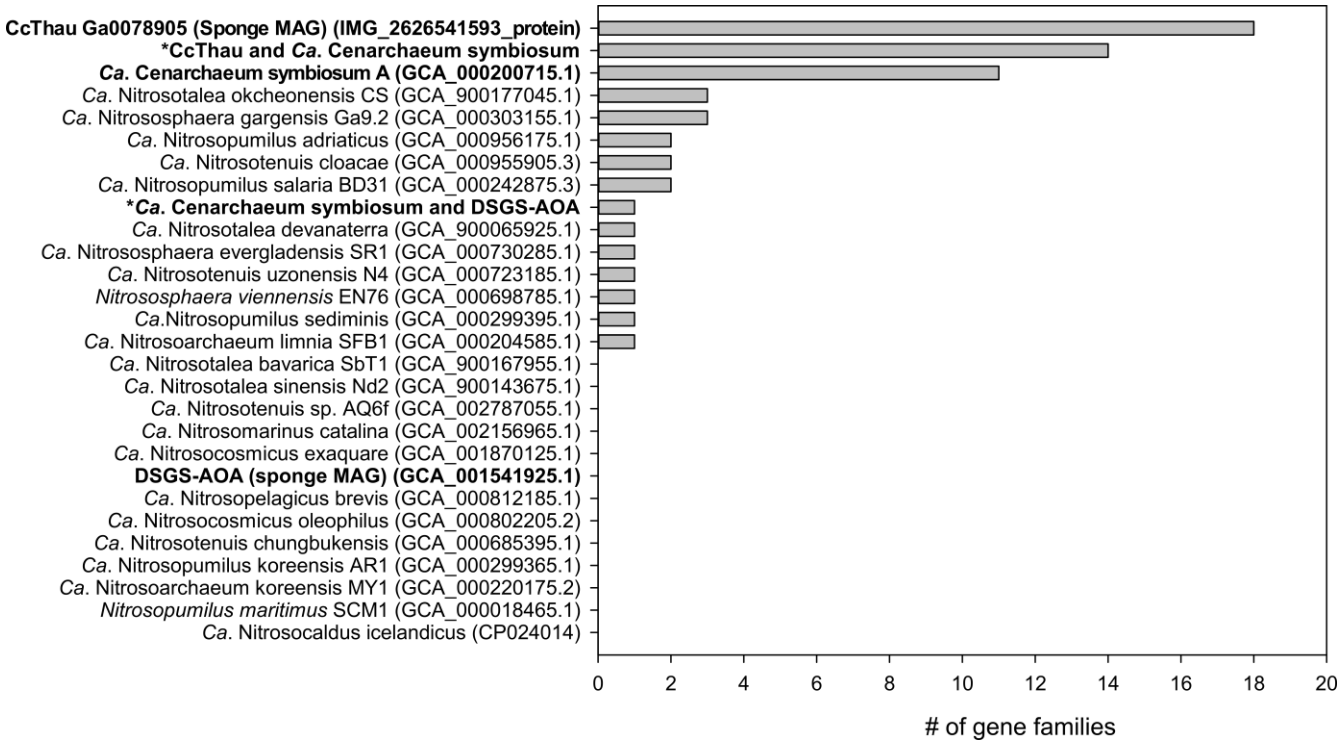


Figure S3. Number of gene families shared exclusively between *Ca*. N. ianthellae and each genome-sequenced member of the *Thaumarchaeota* including the sponge thaumarchaeal symbionts, CCThau, *Ca*. C. symbiosum, and the DSGS-AOA (in bold). (*) indicates gene families shared by *Ca*. N. ianthellae with *Ca.* C. symbiosum and CcThau exclusively, or with *Ca*. C. symbiosum and the glass sponge AOA exclusively.

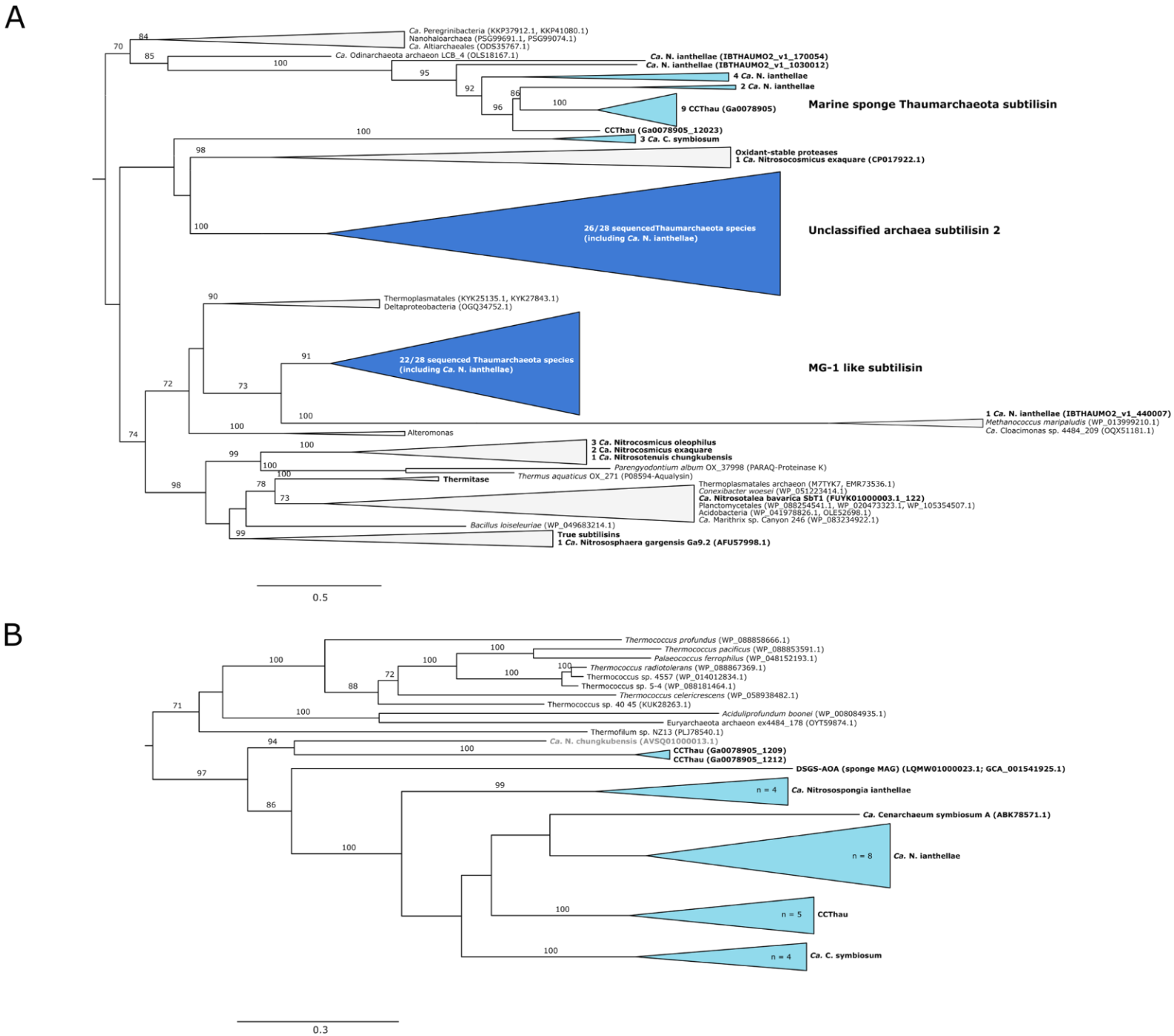
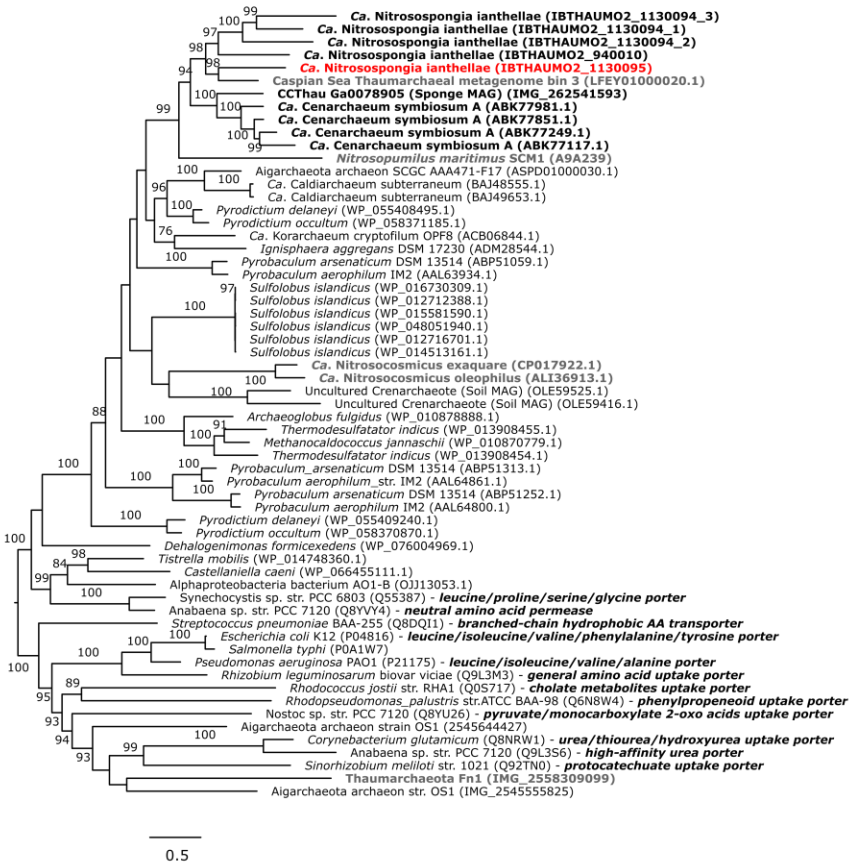

A
Ca. Peregrinibacteria (KKP37912.1, KKP41080.1)
Nanohaloarchaea (PSG99691.1, PSG99074.1)
Ca. Altiarchaeales (ODS35767.1)
Ca. Odinarchaeota archaeon LCB_4 (OLS18167.1)
Ca. N. ianthellae (IBTHAUMO2_v1_170054)
Ca. N. ianthellae (IBTHAUMO2_v1_1030012)
4 Ca. N. ianthellae
2 Ca. N. ianthellae
9 CCThau (Ga0078905)
Marine sponge Thaumarchaeota subtilisin
CCThau (Ga0078905_12023)
3 Ca. C. symbiosum
Oxidant-stable proteases
1 Ca. Nitrosocosmicus exaquare (CP017922.1)
26/28 sequencedThaumarchaeota species
(including Ca. N. ianthellae)
Unclassified archaea subtilisin 2
Thermoplasmatales (KYK25135.1, KYK27843.1)
Deltaproteobacteria (OGQ34752.1)
22/28 sequenced Thaumarchaeota species
(including Ca. N. ianthellae)
MG-1 like subtilisin
1 Ca. N. ianthellae (IBTHAUMO2_v1_440007)
Methanococcus maripaludis (WP_013999210.1)
Ca. Cloacimonas sp. 4484_209 (OQX51181.1)
Alteromonas
3 Ca. Nitrocosmicus oleophilus
2 Ca. Nitrocosmicus exaquare
1 Ca. Nitrosotenuis chungkubensis
Parengyodontium album OX_37998 (PARAQ-Proteinase K)
Thermitase
Thermus aquaticus OX_271 (P08594-Aqualysin)
Thermoplasmatales archaeon (M7TYK7, EMR73536.1)
Conexibacter woesei (WP_051223414.1)
Ca. Nitrosotalea bavarica SbT1 (FUYK01000003.1_122)
Planctomycetales (WP_088254541.1, WP_020473323.1, WP_105354507.1)
Acidobacteria (WP_041978826.1, OLE52698.1)
Ca. Marithrix sp. Canyon 246 (WP_083234922.1)
Bacillus loiseleuriae (WP_049683214.1)
True subtilisins
1 Ca. Nitrososphaera gargensis Ga9.2 (AFU57998.1)
0.5
B
Thermococcus profundus (WP_088858666.1)
Thermococcus pacificus (WP_088853591.1)
Palaeococcus ferrophilus (WP_048152193.1)
Thermococcus radiotolerans (WP_088867369.1)
Thermococcus sp. 4557 (WP_014012834.1)
Thermococcus sp. 5-4 (WP_088181464.1)
Thermococcus celericrescens (WP_058938482.1)
Thermococcus sp. 40 45 (KUK28263.1)
Aciduliprofundum boonei (WP_008084935.1)
Euryarchaeota archaeon ex4484_178 (OYT59874.1)
Thermofilum sp. NZ13 (PLJ78540.1)
Ca. N. chungkubensis (AVSQ01000013.1)
CCThau (Ga0078905_1209)
CCThau (Ga0078905_1212)
DSGS-AOA (sponge MAG) (LQMW01000023.1; GCA_001541925.1)
n = 4
Ca. Nitrosospongia ianthellae
Ca. Cenarchaeum symbiosum A (ABK78571.1)
n = 8
Ca. N. ianthellae
n = 5
CCThau
n = 4
Ca. C. symbiosum
0.3

Figure S4. Maximum-likelihood amino acid phylogenetic trees (after automatic model selection with IQ-Tree, version 1.6.2) of (A) S08A family endopeptidases and (B) serine protease inhibitors (serpins). Color coding donates the degree of homology among all sequenced *Thaumarchaeota*: light blue – shared exclusively among thaumarchaeal sponge symbionts; dark blue – ubiquitously found in *Thaumarchaeota*. In (A), 6 out of 8 *Ca.* N. ianthellae S08A family endopeptidases, within the "Marine sponge Thaumarchaeota subtilisin" clade, are predicted to be exported. On the other hand, the majority of the ubiquitous "Unclassified archaea subtilisin 2" (24/27) and all except one of the "MG-1 like subtilisin" (naming convention after Li *et al*., 2015) are predicted to be membrane anchored S08A endopeptidases. In (B), 3 of the 15 serpins found in *Ca*. N. ianthellae were excluded from phylogenetic analyses due to truncated length. Values at nodes represent ultrafast bootstraps (UFBoot) with only values ≥80% shown for each branch.

A


Ca. Nitrosospongia ianthellae (IBTHAUMO2_1130094_3)
Ca. Nitrosospongia ianthellae (IBTHAUMO2_1130094_1)
Ca. Nitrosospongia ianthellae (IBTHAUMO2_1130094_2)
Ca. Nitrosospongia ianthellae (IBTHAUMO2_940010)
Ca. Nitrosospongia ianthellae (IBTHAUMO2_1130095)
Caspian Sea Thaumarchaeal metagenome bin 3 (LFEY01000020.1)
CCThau Ga0078905 (Sponge MAG) (IMG_262541593)
Ca. Cenarchaeum symbiosum A (ABK77981.1)
Ca. Cenarchaeum symbiosum A (ABK77851.1)
Ca. Cenarchaeum symbiosum A (ABK77249.1)
Ca. Cenarchaeum symbiosum A (ABK77117.1)
Nitrosopumilus maritimus SCM1 (A9A239)
Aigarchaeota archaeon SCGC AAA471-F17 (ASPD01000030.1)
Ca. Caldiarchaeum subterraneum (BAJ48555.1)
Ca. Caldiarchaeum subterraneum (BAJ49653.1)
Pyrodictium delaneyi (WP_055408495.1)
Pyrodictium occultum (WP_058371185.1)
Ca. Korarchaeum cryptofilum OPF8 (ACB06844.1)
Ignisphaera aggregans DSM 17230 (ADM28544.1)
Pyrobaculum arsenaticum DSM 13514 (ABP51059.1)
Pyrobaculum aerophilum IM2 (AAL63934.1)
Sulfolobus islandicus (WP_016730309.1)
Sulfolobus islandicus (WP_012712388.1)
Sulfolobus islandicus (WP_015581590.1)
Sulfolobus islandicus (WP_048051940.1)
Sulfolobus islandicus (WP_012716701.1)
Sulfolobus islandicus (WP_014513161.1)
Ca. Nitrosocosmicus exaquare (CP017922.1)
Ca. Nitrosocosmicus oleophilus (ALI36913.1)
Uncultured Crenarchaeote (Soil MAG) (OLE59525.1)
Uncultured Crenarchaeote (Soil MAG) (OLE59416.1)
Archaeoglobus fulgidus (WP_010878888.1)
Thermodesulfatator indicus (WP_013908455.1)
Methanocaldococcus jannaschii (WP_010870779.1)
Thermodesulfatator indicus (WP_013908454.1)
Pyrobaculum_arsenaticum DSM 13514 (ABP51313.1)
Pyrobaculum aerophilum_str. IM2 (AAL64861.1)
Pyrobaculum arsenaticum DSM 13514 (ABP51252.1)
Pyrobaculum aerophilum IM2 (AAL64800.1)
Pyrodictium delaneyi (WP_055409240.1)
Pyrodictium occultum (WP_058370870.1)
Dehalogenimonas formicexedens (WP_076004969.1)
Tistrella mobilis (WP_014748360.1)
Castellaniella caeni (WP_066455111.1)
Alphaproteobacteria bacterium AO1-B (OJJ13053.1)
Synechocystis sp. str. PCC 6803 (Q55387) - leucine/proline/serine/glycine porter
Anabaena sp. str. PCC 7120 (Q8YVY4) - neutral amino acid permease
Streptococcus pneumoniae BAA-255 (Q8DQI1) - branched-chain hydrophobic AA transporter
Escherichia coli K12 (P04816) - leucine/isoleucine/valine/phenylalanine/tyrosine porter
Salmonella typhi (P0A1W7)
Pseudomonas aeruginosa PAO1 (P21175) - leucine/isoleucine/valine/alanine porter
Rhizobium leguminosarum biovar viciae (Q9L3M3) - general amino acid uptake porter
Rhodococcus jostii str. RHA1 (Q0S717) - cholate metabolites uptake porter
Rhodopseudomonas_palustris str.ATCC BAA-98 (Q6N8W4) - phenylpropeneoid uptake porter
Nostoc sp. str. PCC 7120 (Q8YU26) - pyruvate/monocarboxylate 2-oxo acids uptake porter
Aigarchaeota archaeon strain OS1 (2545644427)
Corynebacterium glutamicum (Q8NRW1) - urea/thiourea/hydroxyurea uptake porter
Anabaena sp. str. PCC 7120 (Q9L3S6) - high-affinity urea porter
Sinorhizobium meliloti str. 1021 (Q92TN0) - protocatechuate uptake porter
Thaumarchaeota Fn1 (IMG_2558309099)
Aigarchaeota archaeon str. OS1 (IMG_2545555825)
0.5

B

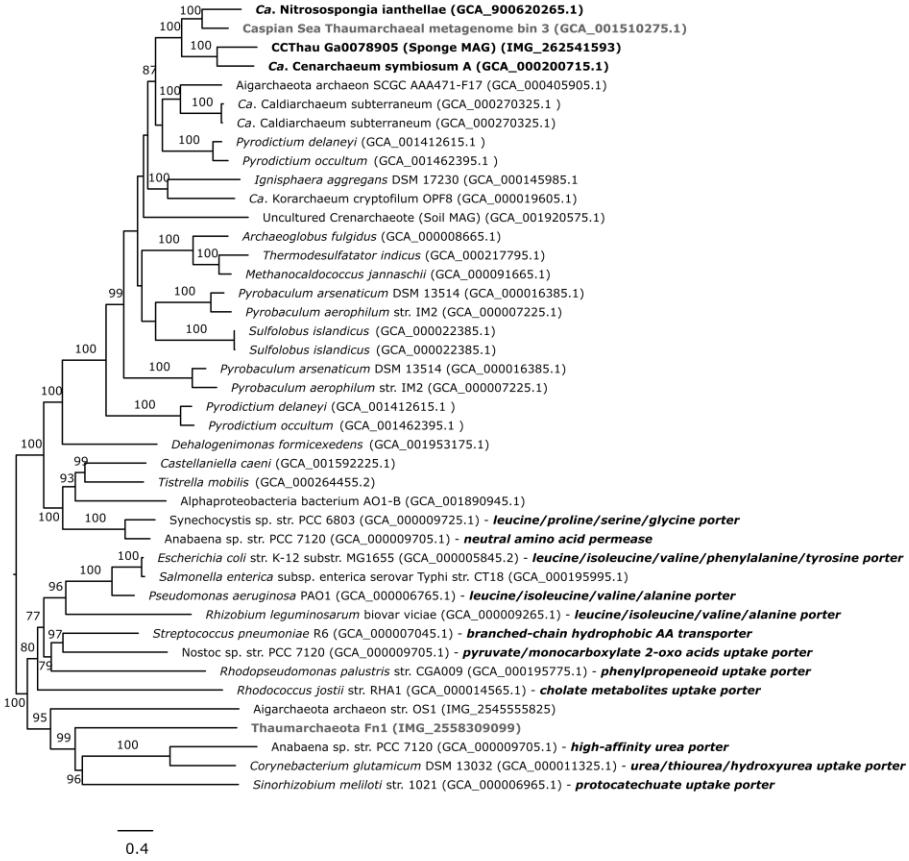

Ca. Nitrosospongia ianthellae (GCA_900620265.1)
Caspian Sea Thaumarchaeal metagenome bin 3 (GCA_001510275.1)
CCThau Ga0078905 (Sponge MAG) (IMG_262541593)
Ca. Cenarchaeum symbiosum A (GCA_000200715.1)
Aigarchaeota archaeon SCGC AAA471-F17 (GCA_000405905.1)
Ca. Caldiarchaeum subterraneum (GCA_000270325.1)
Ca. Caldiarchaeum subterraneum (GCA_000270325.1)
Pyrodictium delaneyi (GCA_001412615.1)
Pyrodictium occultum (GCA_001462395.1)
Ignisphaera aggregans DSM 17230 (GCA_000145985.1
Ca. Korarchaeum cryptofilum OPF8 (GCA_000019605.1)
Uncultured Crenarchaeote (Soil MAG) (GCA_001920575.1)
Archaeoglobus fulgidus (GCA_000008665.1)
Thermodesulfatator indicus (GCA_000217795.1)
Methanocaldococcus jannaschii (GCA_000091665.1)
Pyrobaculum arsenaticum DSM 13514 (GCA_000016385.1)
Pyrobaculum aerophilum str. IM2 (GCA_000007225.1)
Sulfolobus islandicus (GCA_000022385.1)
Sulfolobus islandicus (GCA_000022385.1)
Pyrobaculum arsenaticum DSM 13514 (GCA_000016385.1)
Pyrobaculum aerophilum str. IM2 (GCA_000007225.1)
Pyrodictium delaneyi (GCA_001412615.1)
Pyrodictium occultum (GCA_001462395.1)
Dehalogenimonas formicexedens (GCA_001953175.1)
Castellaniella caeni (GCA_001592225.1)
Tistrella mobilis (GCA_000264455.2)
Alphaproteobacteria bacterium AO1-B (GCA_001890945.1)
Synechocystis sp. str. PCC 6803 (GCA_000009725.1) - leucine/proline/serine/glycine porter
Anabaena sp. str. PCC 7120 (GCA_000009705.1) - neutral amino acid permease
Escherichia coli str. K-12 substr. MG1655 (GCA_000005845.2) - leucine/isoleucine/valine/phenylalanine/tyrosine porter
Salmonella enterica subsp. enterica serovar Typhi str. CT18 (GCA_000195995.1)
Pseudomonas aeruginosa PAO1 (GCA_000006765.1) - leucine/isoleucine/valine/alanine porter
Rhizobium leguminosarum biovar viciae (GCA_000009265.1) - leucine/isoleucine/valine/alanine porter
Streptococcus pneumoniae R6 (GCA_000007045.1) - branched-chain hydrophobic AA transporter
Nostoc sp. str. PCC 7120 (GCA_000009705.1) - pyruvate/monocarboxylate 2-oxo acids uptake porter
Rhodopseudomonas palustris str. CGA009 (GCA_000195775.1) - phenylpropeneoid uptake porter
Rhodococcus jostii str. RHA1 (GCA_000014565.1) - cholate metabolites uptake porter
Aigarchaeota archaeon str. OS1 (IMG_2545555825)
Thaumarchaeota Fn1 (IMG_2558309099)
Anabaena sp. str. PCC 7120 (GCA_000009705.1) - high-affinity urea porter
Corynebacterium glutamicum DSM 13032 (GCA_000011325.1) - urea/thiourea/hydroxyurea uptake porter
Sinorhizobium meliloti str. 1021 (GCA_000006965.1) - protocatechuate uptake porter
0.4

Figure S5. Maximum-likelihood amino acid phylogenetic trees (after automatic model selection with IQ-Tree, version 1.6.2) for the periplasmic subunit (A) *LivK* and for the concatenated (B) *LivFGHMK* operon. *LivK* was found to be highly expressed (expressed copy in red) in the *Ca*. N. ianthellae proteome and can occur in multiple copies among *livFGHMK* encoding microorganisms. *Ca*. N. ianthellae has two *livK* genes and a gene containing three fused *LivK* domains and the predicted proteins from these five *livK* genes/domains were included in the phylogenetic analysis. *LivK* is also found in some other thaumarchaeotes (bold grey), but these lack the *livFGHM* genes necessary for branched-chain amino acid transport. In a few cases in the concatenated *livFGHMK* tree, multiple operons were present in a given genome. For both trees, sponge thaumarchaeal sequences are highlighted in bold black while other thaumarchaeal sequences are highlighted in bold grey. Sequences from organisms where the specific transport functions have been identified, have those specific functions annotated. These annotations are taken from information collated in the Transporter Classification database (Saier *et al*., 2009). Values at nodes represent ultrafast bootstraps (UFBoot) with only values ≥80% shown for each branch.

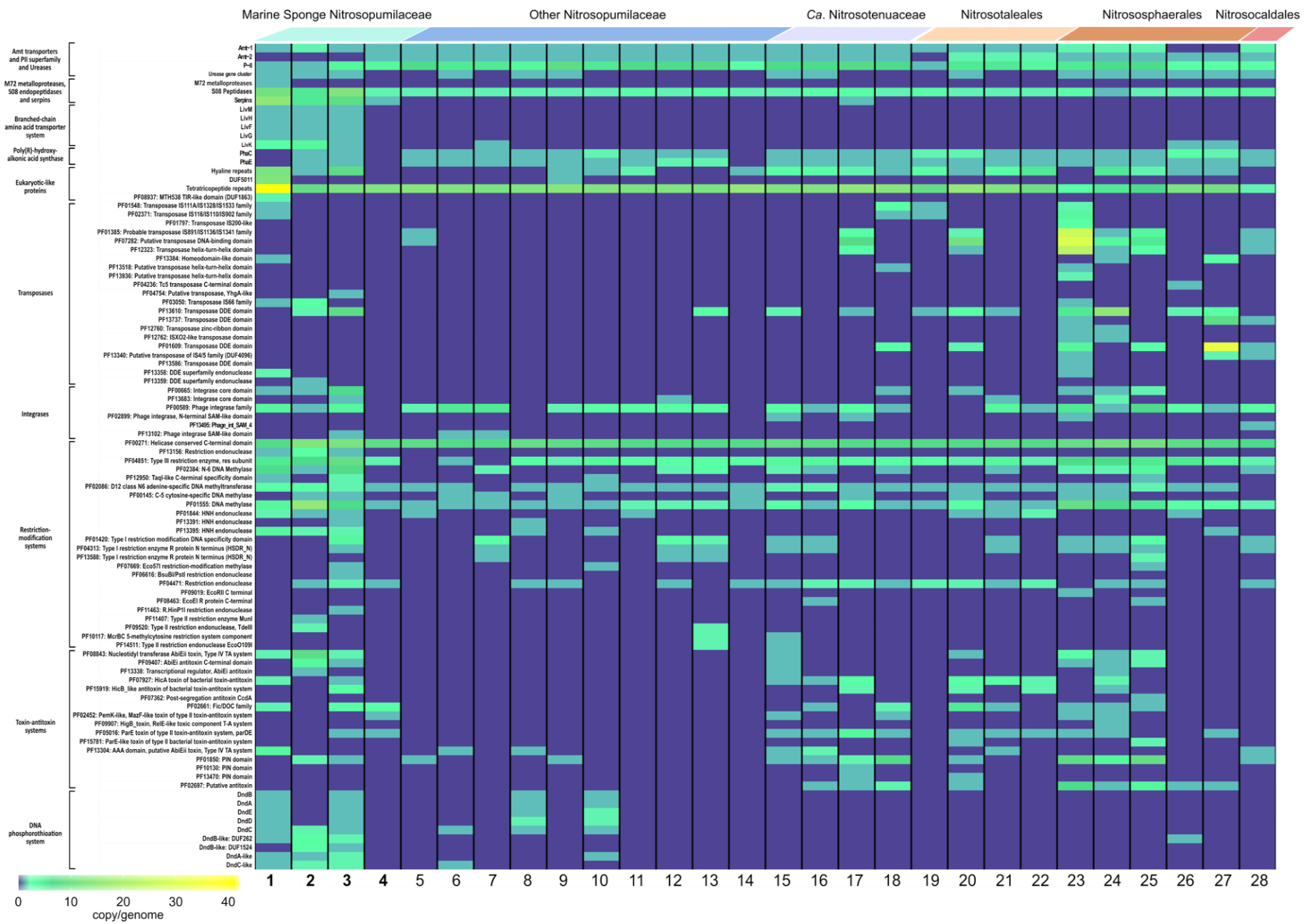

Marine Sponge Nitrosopumilaceae
Other Nitrosopumilaceae
Ca. Nitrosotenuaceae
Nitrosotaleales
Nitrososphaerales
Nitrosocaldales
Amt transporters and PII superfamily and Ureases
Amt-1
Amt-2
P-II
Urease gene cluster
M72 metalloproteases, S08 endopeptidases and serpins
M72 metalloproteases
S08 Peptidases
Serpins
Branched-chain amino acid transporter system
LivM
LivH
LivF
LivG
LivK
Poly(R)-hydroxy-alkonic acid synthase
PhaC
PhaE
Eukaryotic-like proteins
Hyaline repeats
DUF5011
Tetratricopeptide repeats
PF08837: MTH538 TIR-like domain (DUF1863)
Transposases
PF01548: Transposase IS111A/IS1328/IS1533 family
PF02371: Transposase IS116/IS110/IS902 family
PF01797: Transposase IS200-like
PF01385: Probable transposase IS891/IS1136/IS1341 family
PF07282: Putative transposase DNA-binding domain
PF12323: Transposase helix-turn-helix domain
PF13384: Homeodomain-like domain
PF13518: Putative transposase helix-turn-helix domain
PF13936: Putative transposase helix-turn-helix domain
PF04236: Tc5 transposase C-terminal domain
PF04754: Putative transposase, YhgA-like
PF03050: Transposase IS66 family
PF13610: Transposase DDE domain
PF13737: Transposase DDE domain
PF12760: Transposase zinc-ribbon domain
PF12762: ISXO2-like transposase domain
PF01609: Transposase DDE domain
PF13340: Putative transposase of IS4/5 family (DUF4096)
PF13586: Transposase DDE domain
PF13358: DDE superfamily endonuclease
PF13359: DDE superfamily endonuclease
Integrases
PF00665: Integrase core domain
PF13683: Integrase core domain
PF00589: Phage integrase family
PF02899: Phage integrase, N-terminal SAM-like domain
PF13495: Phage_int_SAM_4
PF13102: Phage integrase SAM-like domain
Restriction-modification systems
PF00271: Helicase conserved C-terminal domain
PF13156: Restriction endonuclease
PF04851: Type III restriction enzyme, res subunit
PF02384: N-6 DNA Methylase
PF12950: TaqI-like C-terminal specificity domain
PF02086: D12 class N6 adenine-specific DNA methyltransferase
PF00145: C-5 cytosine-specific DNA methylase
PF01555: DNA methylase
PF01844: HNH endonuclease
PF13391: HNH endonuclease
PF13395: HNH endonuclease
PF01420: Type I restriction modification DNA specificity domain
PF04313: Type I restriction enzyme R protein N terminus (HSDR_N)
PF13588: Type I restriction enzyme R protein N terminus (HSDR_N)
PF07669: Eco57I restriction-modification methylase
PF06616: BsuBI/PstI restriction endonuclease
PF04471: Restriction endonuclease
PF09019: EcoRII C terminal
PF08463: EcoEI R protein C-terminal
PF11463: R.HinP1II restriction endonuclease
PF11407: Type II restriction enzyme MunI
PF09520: Type II restriction endonuclease, TdeIII
PF10117: McrBC 5-methylcytosine restriction system component
PF14511: Type II restriction endonuclease EcoO109I
Toxin-antitoxin systems
PF08843: Nucleotidyl transferase AbiEii toxin, Type IV TA system
PF09407: AbiEi antitoxin C-terminal domain
PF13338: Transcriptional regulator, AbiEi antitoxin
PF07927: HicA toxin of bacterial toxin-antitoxin
PF15919: HicB_like antitoxin of bacterial toxin-antitoxin system
PF07362: Post-segregation antitoxin CcdA
PF02661: Fic/DOC family
PF02452: PemK-like, MazF-like toxin of type II toxin-antitoxin system
PF09907: HigB_toxin, RelE-like toxic component T-A system
PF05016: ParE toxin of type II toxin-antitoxin system, parDE
PF15781: ParE-like toxin of type II bacterial toxin-antitoxin system
PF13304: AAA domain, putative AbiEii toxin, Type IV TA system
PF01850: PIN domain
PF10130: PIN domain
PF13470: PIN domain
PF02697: Putative antitoxin
DNA phosphorothioation system
DndB
DndA
DndE
DndD
DndC
DndB-like: DUF262
DndB-like: DUF1524
DndA-like
DndC-like
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
0 10 20 30 40
copy/genome

Figure S6. Heat map showing the distribution and gene copy number per genome of selected genes, gene classes and PFAM annotations among genome-sequenced AOA. The color scale ranges from 0 (dark blue) to 40 (yellow) and indicates copies per genome. Sponge-derived genomes start on the left and are depicted in bold, followed by members of *Ca*. Nitrosopumilaceae, *Ca*. Nitrosotenuaceae, *Ca*. Nitrosotaleale, the *Nitrososphaerales*, and *Ca*. Nitrosocaldales, respectively. Numbers at the bottom of the figure refer to the AOA genomes: **1)** *Ca*. N. ianthellae, **2)** *Ca*. C. symbiosum, **3)** CcThau Ga007890, **4)** DSGS-AOA, **5)** *Ca*. N. koreensis AR1, **6)** *Ca*. N. piranensis, **7)** *N. maritimus* SCM1, **8)** *Ca*. N. catalina, **9)** *Ca*. N. sediminis, **10)** *Ca*. N. salaria, **11)** *Ca*. N. adriaticus, **12)** *Ca*. N. koreensis MY1, **13)** N. limnia SFB1, **14)** *Ca*. N. brevis, **15)** *Ca*. N sp. AQ6f, **16)** *Ca*. N. cloacae, **17)** *Ca*. N. chungbukensis, **18)** *Ca*. N. uzonensis, **19)** *Ca*. N. bavarica, **20)** *Ca*. N. okcheonensis, **21)** *Ca*. N. devanaterra, **22)** *Ca*. N. siniensis, **23)** *Ca*. N. gargensis, **24)** *N. viennensis*, **25)** *Ca*. N. evergladensis, **26)** *Ca*. N. exaquare, **27)** *Ca*. N. oleophilus, and **28)** *Ca*. N. icelandicus.

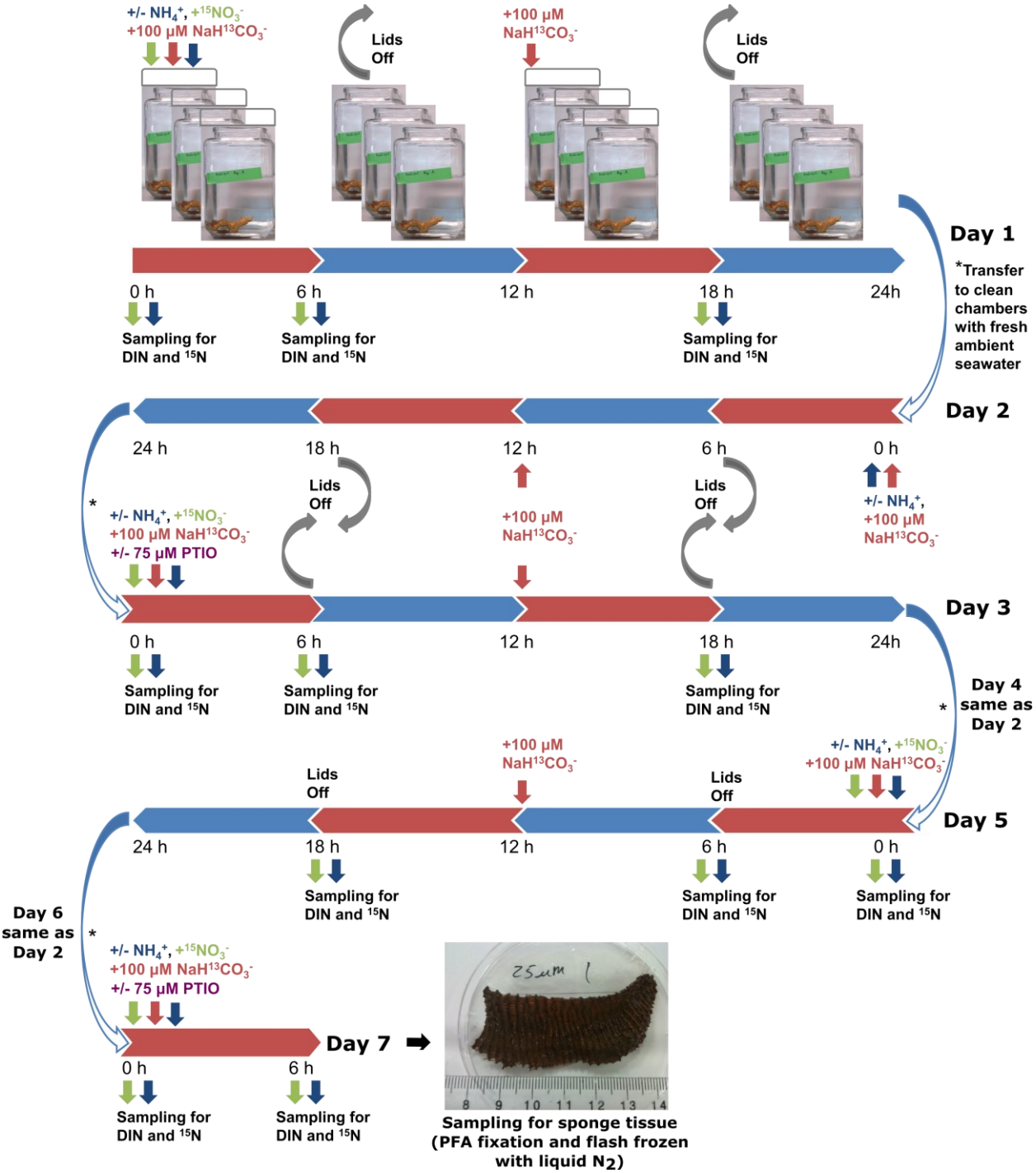


Figure S7. Experimental design for the *I. basta* holobiont nitrification incubations. The intermittently closed setup was employed to avoid oxygen depletion while minimizing loss of ^{13}C-labeled HCO_3^-. Green and blue arrows denote samples used for the calculation of net and gross nitrification rates, respectively. Days 4, 5 and 6 were recovery days for those incubations to which PTIO was added.

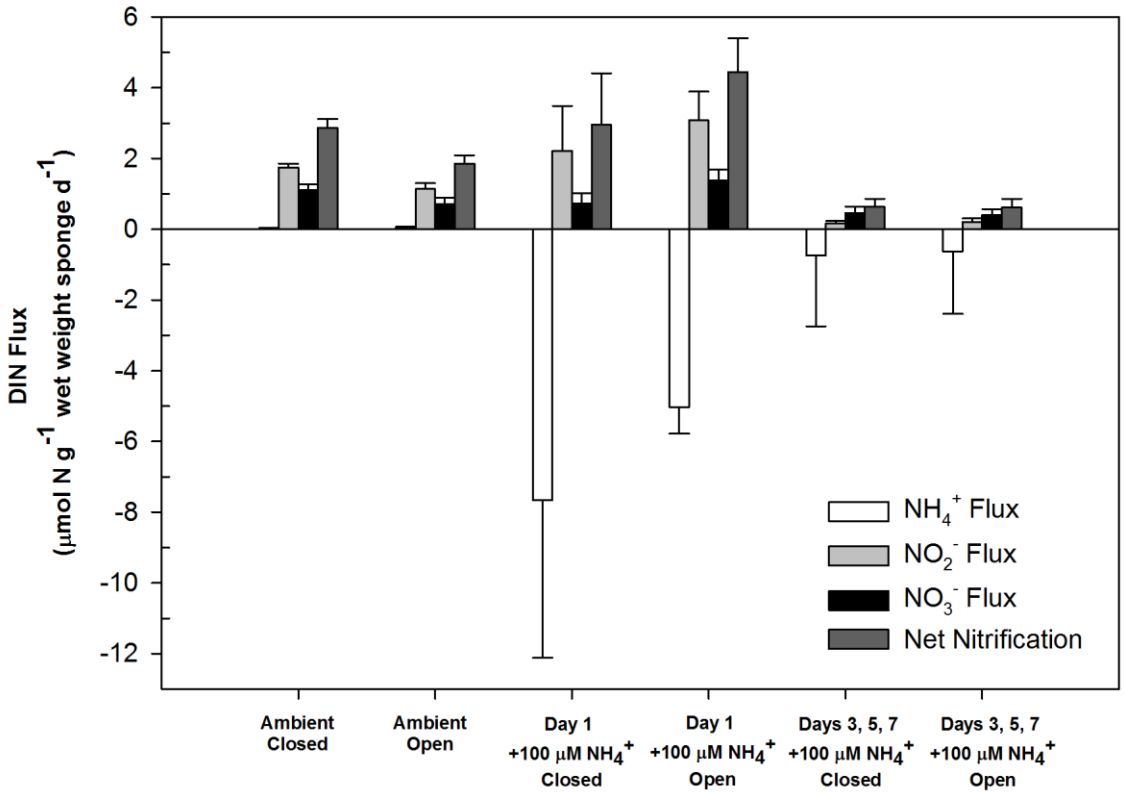


Figure S8. Comparison of average DIN (NH_4^+, NO_3^-, NO_2^-) flux and net nitrification rates exhibited by the *I. basta* holobiont in multiple-day incubations between intermittently closed (Supporting Information Figure 1) and completely open aquaria at ambient conditions and at 100 μM NH_4^+. Pair-wise comparisons of individual DIN species flux and net nitrification between open and intermittently closed aquaria, revealed significant differences in net NO_2^- flux and net nitrification between ambient treatments (both: $p < 0.01$, Mann-Whitney U-test).

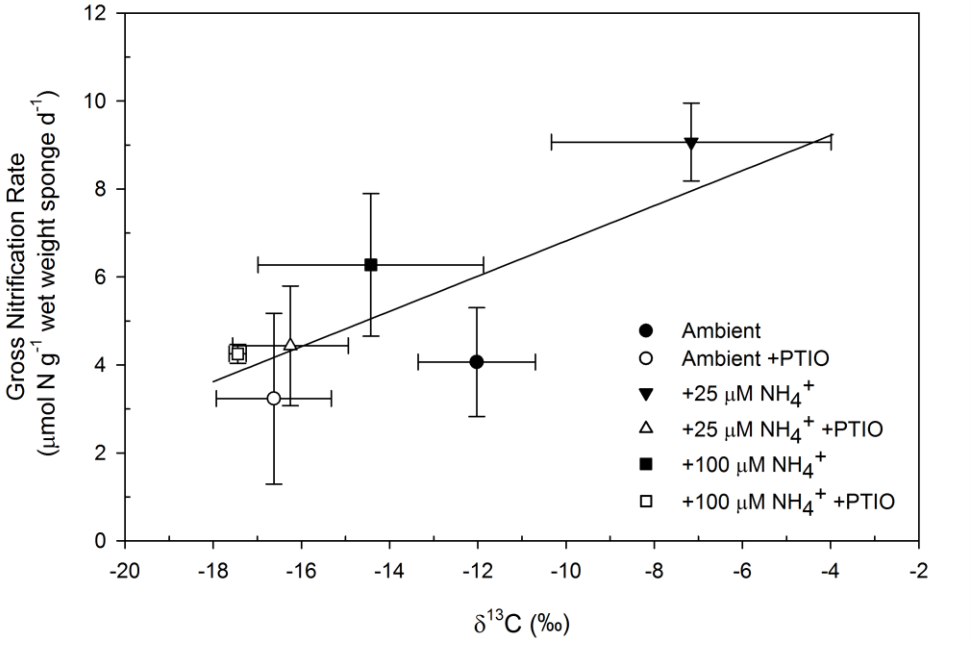


Figure S9. Nitrification activity of the *I. basta* holobiont during 7-day incubation experiments at ambient conditions and with added ammonium (25 or 100 μM) and with/without the AOA inhibitor PTIO. The relationship between $\delta^{13}C$ values of sponge clones (labeled from added ^{13}C-bicarbonate) sampled on day 7 and gross nitrification rates of the *I. basta* holobiont is depicted. To yield average nitrification activity over the course of the incubation, we used the weighted mean of the gross nitrification rates (i.e., days are weights). For comparison, the $\delta^{13}C$ value reflecting the natural abundance was determined from the sponge individual used for Illumina metagenome sequencing and found to be -21.7 ‰ (not shown). The $\delta^{13}C$ values derived from the *I. basta* holobiont nitrification experiments displayed a positive and significant correlation with gross nitrification rates ($R = 0.656$, $p < 0.005$). Error bars reflect the standard error of the sample mean, where $n = 3$ for the $\delta^{13}C$ values and for all weighted gross nitrification rates in all treatment conditions. PTIO additions started two days after the experiment was commenced; hence fixation of ^{13}C-labeled bicarbonate during the first two days was not influenced by inhibition.

Table S1. Overview of genome binning parameters, statistics and associated metadata for the *Ca.* N. ianthellae genome bin.

Bin ID	**IBThaumO2**
Analysis project type	metagenome-assembled genome (MAG)
Taxa_id	16S rRNA and multi-marker phylogenetics
Assembly software	Spades
Annotation	MaGe
Genome Quality	High Quality Draft
Completeness (%)	99.03
Contamination (%)	0.97
Completeness/Contamination Software	CheckM
Number of contigs	113
16S rRNA gene recovered	yes
16S rRNA gene recovery software	Rnammer / MaGe
Number of standard tRNAs extracted	20
tRNA extraction software	tRNA-scan / MaGe
Binning software	metabat2
Binning parameters	kmer
Genome size (Mbp)	1.99
N50 (bp)	35099
Longest contig (bp)	135,585
Average contig length (bp)	17,669.70
GC content (%)	64.8
Protein coding sequences	2,342
rRNAs	3
tRNAs	42

Sponge species	Net nitrification rates [µmol N (cm^{-3} or g^{-1} wet wt.) day^{-1}]	Net nitrite formation rates [µmol N (cm^{-3} or g^{-1} wet wt.) day^{-1}]	Marine area, depth, and season	Experimental Setup	AOB/AOA Diversity	Reference
Anthosigmella varians	0.003 – 0.105 (g^{-1}*)	N.D.	Caribbean coral reef (6m)	4h - 2.25 L batch incubations +/- light and 5 µM NH_4^+	N.D.	Corredor *et al.*, 1988
Chondrilla nucula	1.03 – 1.49 ◊ (g^{-1}*)					
Chondrilla nucula	0.864 – 6.36 (g^{-1}*)	0.014 – 0.24 (g^{-1}*)	Caribbean coral reef (20-40m) and mangroves (1-3m); June-Sept.	6 to 12h – 3 and 20L batch inc., +/- light	N.D.	Diaz & Ward, 1997
Pseudaxinella zeai	0 – 2.47 (g^{-1}*)	0 – 0.048 (g^{-1}*)				
Oligoceras violacea	0 – 1.37 (g^{-1}*)	0.408 – 1.39 (g^{-1}*)				
Plakortis halichondroides	0 – 0.768 (g^{-1}*)	0 – 0.192 (g^{-1}*)				
Aplysina aerophoba	3.6 – 9.2 ◊ (g^{-1})	N.D.	Med. Sea. 2-20m:	9 to 28h - 3L batch, +/- 100 or 200 µM NH_4^+	β-AOB 16S – 9 OTUs – 1 OTU	Bayer *et al.*, 2007
Dysidea avara	0					
Chondrosia reniformis	~0.3 (g^{-1})					
Aplysina aerophoba	0.214 – 0.826 (g^{-1}*)	N.D.	Med. Sea, 2-15m April - Sept.	21 to 28h – 3L batch inc., +/- 100 or 200 µM NH_4^+ +/- nitrapyrin	A- and β-*amoA*/16S – 5, 7, 9 OTUs	Bayer *et al.*, 2008
Aplysina aerophoba	1.13 (g^{-1}*)	N.D.	Med. Sea. 10-20m; Sept. for all except for *A. aroides* (July)	6h – 7L batch inc.	6 A-*amoA* OTUs in 3 clust. 12 β-*amoA* OTUs in 3 clust. 83 γ-*amoA* clones	Jiménez & Ribes, 2007; Ribes *et al.*, 2012
Agelas oroides	0.875 (g^{-1}*)					
Dysidea avara	N.S.					
Chondrosia reniformis	1.68 (g^{-1}*)					
Axinella polypoides	0.452 (g^{-1}*)					
Ircinia oros	0.544 (g^{-1}*)					
Aplysina cauliformis	4.08 (g^{-1}*)	N.D.	Florida Keys, ~4m	6 to 8h – 2-4L batches; *in situ* sampling	N.D.	Southwell *et al.*, 2008
Smenospongia aurea	4.32 (g^{-1}*)					
7 other species: *A. archeri, A. lacunose, I. felix, I. strobilina, P. crassa, V. rigida*, and *X. muta*	+ + + + + + +					
Chondrosia reniformis,	0.176 (g^{-1})	0.016 (g^{-1})	Med. Sea, 10-15m	24h – 1L batches with 10 µM NH_4^+	1 A-*amoA* OTU	Schläppy *et al.*, 2010
Dysidea avara	0.294 (g^{-1})	0.016 (g^{-1})				
Phakellia ventilabrum	0.14 – 2.26 (g^{-1}○)	0 – 0.192 (g^{-1}○)	Norwegian coast (fjords), 200-300m: Mar.-Nov.	24 to 48 h in 500 to 900 mL batches with 10-12 µM NH_4^+	3 A-*amoA* OTUs 3 A-*amoA* OTUs 1 A-*amoA* OTU	Radax *et al.*, 2012; Hoffmann *et al.*, 2009
Antho dichotoma	0.	0.360 (g^{-1}○)				
Geodia barretti	0.679 (g^{-1}○)	0.180 (g^{-1}○)				
Ianthella basta	1.67 – 11.3 (g^{-1})	0.64 – 6.07 (g^{-1})	Coral Sea, Australia, 10 m; Sept. – Oct.	7- day long 24 h batch incubations in 1.74 L +/- 25, 100 µM NH_4^+ +/- PTIO	1 A-*amoA* OTU	This study
	1.54 – 13.3 ◊ (g^{-1})	0.35 – 10.36 ◊ (g^{-1})				

Table S4. Comparison of net nitrite formation and nitrification rates from our study with previously published sponge studies. (*****) signifies that sponge dry wt. g^{-1} was converted by assuming dry weight = 10% of wet weight while a (○) signifies that sponge cm^{-3} was converted by assuming 1 cm^3 sponge = 1.2 g wet weight sponge (Schläppy *et al.*, 2010). A (◊) denotes that the rate is a potential rate (i.e. with added ammonium). A (+) indicates positive NO_x^- production where quantity was not assessed. "A-, β- , and γ-*amoA*" refer to archaeal, betaproteobacterial and gammaproteobacterial *amoA* OTUs, respectively.

Table S5. AOA used for comparative genome analyses.

Organism/Name	Assembly Accession	doi	Size (Mb)	GC%	Status	BioProject Accession	BioSample Accession	Strain	Completeness	Contamination
Candidatus Nitrosocaldus icelandicus	GCF_002906225.1	doi: https://doi.org/10.1101/235028	1.62	41.5	Complete Gen.	PRJNA413650	SAMN07759886	3F	100	0
Nitrosopumilus maritimus SCM1	GCA_000018465.1	doi: 10.1073/pnas.0913533107	1.65	34.2	Complete Gen.	PRJNA19265	SAMN00000032	SCM1	100	0.97
Candidatus Cenarchaeum symbiosum A	GCA_000200715.1	doi: 10.1073/pnas.0608549103	2.05	57.4	Chromosome	PRJNA202	SAMN02744041	-	99.03	0
Candidatus Nitrosoarchaeum limnia SFB1	GCA_000204585.1	doi: 10.1371/journal.pone.0016626	1.77	32.6	Chromosome	PRJNA52465	SAMN02471010	SFB1	98.06	0
Candidatus Nitrosoarchaeum koreensis MY1	GCA_000220175.2	doi: 10.1128/JB.05717-11	1.61	32.7	Contig	PRJNA67913	SAMN02470178	MY1	100	0
Candidatus Nitrosopumilus salaria BD31	GCA_000242875.3	doi: 10.1128/JB.00013-12	1.57	33.8	Contig	PRJNA50075	SAMN00016669	BD31	92.39	1.94
Candidatus Nitrosopumilus koreensis AR1	GCA_000299365.1	doi: 10.1128/JB.01857-12	1.64	34.2	Complete Gen.	PRJNA174387	SAMN02603137	AR1	94.66	0
Candidatus Nitrosopumilus sediminis	GCA_000299395.1	doi: 10.1128/JB.01869-12	1.69	33.6	Complete Gen.	PRJNA174388	SAMN02603138	AR2	97.09	0
Candidatus Nitrososphaera gargensis Ga9.2	GCA_000303155.1	doi: 10.1111/j.1462-2920.2012.02893.x	2.83	48.3	Complete Gen.	PRJNA60505	SAMN02603264	-	100	2.91
Candidatus Nitrosotenuis chungbukensis	GCA_000685395.1	doi: 10.1128/AEM.03730-13	1.76	41.8	Contig	PRJNA210247	SAMN02767256	MY2	99.03	0.97
Nitrososphaera viennensis EN76	GCA_000698785.1	doi: 10.1099/ijs.0.063172-0	2.53	52.7	Complete Gen.	PRJEA60103	SAMN02721150	EN76	100	0.97
Candidatus Nitrosotenuis uzonensis N4	GCA_000723185.1	doi: 10.1371/journal.pone.0080835	1.64	42.2	Contig	PRJEB4650	SAMEA3139018	N4	100	0.97
Candidatus Nitrososphaera evergladensis SR1	GCA_000730285.1	doi: 10.1371/journal.pone.0101648	2.95	50.1	Complete Gen.	PRJNA235208	SAMN03081530	-	100	2.91
Candidatus Nitrosocosmicus oleophilus	GCA_000802205.2	doi: 10.1111/1758-2229.12477	3.43	34.1	Complete Gen.	PRJNA210256	SAMN03074222	MY3	98.06	0.97
Candidatus Nitrosopelagicus brevis	GCA_000812185.1	doi: 10.1073/pnas.1416223112	1.23	33.2	Complete Gen.	PRJNA223412	SAMN03273964	CN25	99.51	0
Candidatus Nitrosopumilus piranensis	GCA_000875775.1	doi: 10.1038/ismej.2015.200	1.71	33.8	Complete Gen.	PRJNA269924	SAMN03257648	D3C	100	0.97
Candidatus Nitrosotenuis cloacae	GCA_000955905.3	doi: 10.1038/srep23747	1.62	41.0	Complete Gen.	PRJNA272771	SAMN03286947	SAT1	100	1.94
Candidatus Nitrosopumilus adriaticus	GCA_000956175.1	doi: 10.1038/ismej.2015.200	1.80	33.4	Complete Gen.	PRJNA269341	SAMN03253153	NF5	100	0
Candidatus Nitrosopumilus sp. Nsub (DSGS-AOA sponge MAG)	GCA_001541925.1	doi: 10.1128/mSystems.00184-16	1.38	31.4	Contig	PRJNA308059	SAMN04386402	Nsub	100	0
Candidatus Nitrosocosmicus exaquare	GCA_001870125.1	doi: 10.1038/ismej.2016.192	2.99	33.9	Complete Gen.	PRJNA317395	SAMN04606696	G61	99.03	2.91
Candidatus Nitrosomarinus catalina	GCA_002156965.1	doi: 10.1111/1462-2920.13768	1.36	31.4	Complete Gen.	PRJNA341864	SAMN05730076	SPOT01	100	0
Candidatus Nitrosotenuis aquarius AQ6f	GCA_002787055.1	doi: 10.1128/AEM.01430-18	1.70	42.2	Complete Gen.	PRJNA406986	SAMN07637255	AQ6F	99.68	1.94
Candidatus Nitrosotalea devanaterra	GCA_900065925.1	doi: 10.1128/AEM.04031-15	1.81	37.1	Complete Genon	PRJEB10948	SAMEA3577360	-	98.54	0
Candidatus Nitrosotalea sinensis Nd2	GCA_900143675.1	doi: 10.1111/1462-2920.13971	1.60	37.4	Contig	PRJEB15449	SAMEA20449918	-	99.51	0.97
Candidatus Nitrosotalea bavarica SbT1	GCA_900167955.1	doi: 10.1111/1462-2920.13971	1.55	36.0	Scaffold	PRJEB15449	SAMEA101071168	-	97.57	1.94
Candidatus Nitrosotalea okcheonensis CS	GCA_900177045.1	doi: 10.1111/1462-2920.13971	1.97	37.5	Chromosome	PRJEB15449	SAMEA20449168	-	99.51	0
CcThau Ga0078905 (Sponge MAG)	IMG_2626541593_protein	doi:10.1038/ismej.2017.25	2.16	38.4	Contig	NA	NA	-	97.12	4.01
Candidatus Nitrosospongia ianthellae	GCA_900620265.1	This publication	1.99	64.8	Scaffold	PRJEB29556	SAMEA5126984	O2	99.03	0.97

References

Alves, R.J.E., Minh, B.Q., Urich, T., von Haeseler, A., and Schleper, C. (2018) Unifying the global phylogeny and environmental distribution of ammonia-oxidising archaea based on *amoA* genes. *Nat Commun* **9**: 1517.

Bayer, K., Schmitt, S., and Hentschel, U. (2007) Microbial nitrification in Mediterranean sponges: possible involvement of ammonium-oxidizing *Betaproteobacteria*. In *Porifera Research: Biodiversity, Innovation, Sustainability.*Custódio, M., Lôbo-Hajdu, G., Hajdu, E., and Muricy, G. (eds). Série Livros. Rio de Janeiro, Brazil: Museu Nacional, pp. 165–171.

Bayer, K., Schmitt, S., and Hentschel, U. (2008) Physiology, phylogeny and *in situ* evidence for bacterial and archaeal nitrifiers in the marine sponge *Aplysina aerophoba*. *Environ Microbiol* **10**: 2942–2955.

Bayer, B., Vojvoda, J., Offre, P., Alves, R.J.E., Elisabeth, N.H., Garcia, J.A., *et al*. (2016) Physiological and genomic characterization of two novel marine thaumarchaeal strains indicates niche differentiation. *ISME J* **10**: 1051–1063.

Berger, S.A., Krompass, D., and Stamatakis, A. (2011) Performance, accuracy, and web Server for evolutionary placement of short sequence reads under maximum-likelihood. *Syst Biol* **60**: 291–302.

Blainey, P.C., Mosier, A.C., Potanina, A., Francis, C.A., and Quake, S.R. (2011) Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. *PLoS ONE* **6**: e16626.

Blatch, G.L. and Lässle, M. (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays* **21**: 932–939.

Callebaut, I., Mornon, J.-P., Gilgès, D., and Vigon, I. (2000) HYR, an extracellular module involved in cellular adhesion and related to the immunoglobulin-like fold. *Protein Sci* **9**: 1382–1390.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.

Cort, J.R., Yee, A., Edwards, A.M., Arrowsmith, C.H., and Kennedy, M.A. (2000) Structure-based functional classification of hypothetical protein MTH538 from *Methanobacterium thermoautotrophicum*. *J Mol Biol* **302**: 189–203.

Corredor, J.E., Wilkinson, C.R., Vicente, V.P., Morell, J.M., and Otero, E. (1988) Nitrate release by Caribbean reef sponges. *Limnol Oceanogr* **33**: 114–120.

Diaz, M. and Ward, B. (1997) Sponge-mediated nitrification in tropical benthic communities. *Mar Ecol Prog Ser* **156**: 97–107.

Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., *et al*. (2018) Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**: eaar4120.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 17921797.

Fan, L., Reynolds, D., Liu, M., Stark, M., Kjelleberg, S., Webster, N.S., and Thomas, T. (2012) Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proc Natl Acad Sci USA* **109**: E1878–E1887.

Foerstner, K.U., von Mering, C., Hooper, S.D., and Bork, P. (2005) Environments shape the nucleotide composition of genomes. *EMBO Rep* **6**: 1208–1213.

Gauthier, M.E.A., Du Pasquier, L., and Degnan, B.M. (2010) The genome of the sponge Amphimedon queenslandica provides new perspectives into the origin of Toll-like and interleukin 1 receptor pathways: The origin of Toll-like and IL1 receptor pathways. *Evol Dev* **12**: 519–533.

Herbold, C.W., Lehtovirta-Morley, L.E., Jung, M.-Y., Jehmlich, N., Hausmann, B., Han, P., *et al*. (2017) Ammonia-oxidising archaea living at low pH: Insights from comparative genomics: Comparative genomics of ammonia-oxidising archaea. *Environ Microbiol* **19**: 4939–4952.

Hoffmann, F., Radax, R., Woebken, D., Holtappels, M., Lavik, G., Rapp, H.T., *et al*. (2009) Complex nitrogen cycling in the sponge *Geodia barretti*. *Environ Microbiol* **11**: 2228–2243.

Horn, H., Slaby, B.M., Jahn, M.T., Bayer, K., Moitinho-Silva, L., Förster, F., *et al*. (2016) An enrichment of CRISPR and other defense-related features in marine sponge-associated microbial metagenomes. *Front Microbiol* **7**: 1751.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., *et al*. (2016) eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* **44**: D286–D293.

Jernigan, K.K. and Bordenstein, S.R. (2015) Tandem-repeat protein domains across the tree of life. *PeerJ* **3**: e732.

Jiménez, E. and Ribes, M. (2007) Sponges as a source of dissolved inorganic nitrogen: Nitrification mediated by temperate sponges. *Limnol Oceanogr* **52**: 948–958.

Jung, M.-Y., Park, S.-J., Kim, S.-J., Kim, J.-G., Sinninghe Damsté, J.S., Jeon, C.O., and Rhee, S.-K. (2014) A mesophilic, autotrophic, ammonia-oxidizing archaeon of thaumarchaeal group I.1a cultivated from a deep oligotrophic soil horizon. *Appl Environ Microbiol* **80**: 3645–3655.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**: 587–589.

Karimi, E., Slaby, B.M., Soares, A.R., Blom, J., Hentschel, U., and Costa, R. (2018) Metagenomic binning reveals versatile nutrient cycling and distinct adaptive features in alphaproteobacterial symbionts of marine sponges. *FEMS Microbiol Ecol* **94**: 6.

Katoh, K. and Standley, D.M. (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* **30**: 772–780.

Kerou, M., Offre, P., Valledor, L., Abby, S.S., Melcher, M., Nagler, M., *et al*. (2016) Proteomics and comparative genomics of *Nitrososphaera viennensis* reveal the core genome and adaptations of archaeal ammonia oxidizers. *Proc Natl Acad Sci USA* **113**: E7937–E7946.

Könneke, M., Schubert, D.M., Brown, P.C., Hugler, M., Standfest, S., Schwander, T., *et al*. (2014) Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO2 fixation. *Proc Natl Acad Sci USA* **111**: 8239–8244.

Lagkouvardos, I., Joseph, D., Kapfhammer, M., Giritli, S., Horn, M., Haller, D., and Clavel, T. (2016) IMNGS: A comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. *Sci Rep-UK* **6**: 33721.

Lartillot, N. and Philippe, H. (2004) A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol BiolEvol* **21**: 1095–1109.

Lassalle, F., Périan, S., Bataillon, T., Nesme, X., Duret, L., and Daubin, V. (2015) GC-Content evolution in bacterial genomes: the biased gene conversion hypothesis expands. *PLoS Genet* **11**: e1004941.

Laundon, D., Larson, B., McDonald, K., King, N., and Burkhardt, P. (2018) The architecture of cell differentiation in choanoflagellates and sponge choanocytes. *bioRxiv*.

Li, P.-N., Herrmann, J., Tolar, B.B., Poitevin, F., Ramdasi, R., Bargar, J.R., *et al*. (2018) Nutrient transport suggests an evolutionary basis for charged archaeal surface layer proteins. *ISME J* **12**: 2389–2402.

Mah, J.L., Christensen-Dalsgaard, K.K., and Leys, S.P. (2014) Choanoflagellate and choanocyte collar-flagellar systems and the assumption of homology. *Evol Dev* **16**: 25–37.

Maldonado, M. (2005) Choanoflagellates, choanocytes, and animal multicellularity. *Invertebr Biol* **123**: 1–22.

Manning, G., Young, S.L., Miller, W.T., and Zhai, Y. (2008) The protist, *Monosiga brevicollis*, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan. *Proc Natl Acad Sci USA* **105**: 9674–9679.

Moitinho-Silva, L., Díez-Vives, C., Batani, G., Esteves, A.I., Jahn, M.T., and Thomas, T. (2017) Integrated metabolism in sponge–microbe symbiosis revealed by genome-centered metatranscriptomics. *ISME J* **11**: 1651–1666.

Moitinho-Silva, L., Nielsen, S., Amir, A., Gonzalez, A., Ackermann, G.L., Cerrano, C., et al. (2017) The sponge microbiome project. *GigaScience* **6**: 1–7.

Mosier, A.C., Allen, E.E., Kim, M., Ferriera, S., and Francis, C.A. (2012) Genome sequence of '*Candidatus* Nitrosopumilus salaria' BD31, an ammonia-oxidizing archaeon from the San Francisco Bay Estuary. *J Bacteriol* **194**: 2121–2122.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**: 268–274.

Otte, J., Mall, A., Schubert, D.M., Könneke, M., and Berg, I.A. (2015) Malonic semialdehyde reductase from the archaeon *Nitrosopumilus maritimus* is involved in the autotrophic 3-hydroxypropionate/4-hydroxybutyrate cycle. *Appl Environ Microbiol* **81**: 1700–1707.

Palatinszky, M., Herbold, C., Jehmlich, N., Pogoda, M., Han, P., von Bergen, M., *et al*. (2015) Cyanate as an energy source for nitrifiers. *Nature* **524**: 105–108.

Patterson, N.J., Günther, J., Gibson, A.J., Offord, V., Coffey, T.J., Splitter, G., *et al*. (2014) Two TIR-like domain containing proteins in a newly emerging zoonotic *Staphylococcus aureus* strain sequence type 398 are potential virulence factors by impacting on the host innate immune response. *Front Microbiol* **5**: 662.

Pruesse, E., Peplies, J., and Glöckner, F.O. (2012) SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**: 1823–1829.

Qin, W., Meinhardt, K.A., Moffett, J.W., Devol, A.H., Virginia Armbrust, E., Ingalls, A.E., and Stahl, D.A. (2017) Influence of oxygen availability on the activities of ammonia-oxidizing archaea. *Envirol Microbiol Rep* **9**: 250–256.

Radax, R., Hoffmann, F., Rapp, H.T., Leininger, S., and Schleper, C. (2012) Ammonia-oxidizing archaea as main drivers of nitrification in cold-water sponges. *Environ Microbiol* **14**: 909–923.

Reichenberger, E.R., Rosen, G., Hershberg, U., and Hershberg, R. (2015) Prokaryotic nucleotide composition is shaped by both phylogeny and the environment. *Genome Biol Evol* **7**: 1380–1389.

Reynolds, D. and Thomas, T. (2016) Evolution and function of eukaryotic-like proteins from sponge symbionts. *Mol Ecol* **25**: 5242–5253.

Ribes, M., Jiménez, E., Yahel, G., López-Sendino, P., Diez, B., Massana, R., *et al*. (2012) Functional convergence of microbes associated with temperate marine sponges. *Environ Microbiol* **14**: 1224–1239.

Saier, M.H., Yen, M.R., Noto, K., Tamang, D.G., and Elkan, C. (2009) The Transporter Classification Database: recent advances. *Nucleic Acids Res* **37**: D274–D278.

Santoro, A.E., Dupont, C.L., Richter, R.A., Craig, M.T., Carini, P., McIlvin, M.R., *et al*. (2015) Genomic and proteomic characterization of " *Candidatus* Nitrosopelagicus brevis": An ammonia-oxidizing archaeon from the open ocean. *Proc Natl Acad Sci USA* **112**: 1173–1178.

Schläppy, M.-L., Schöttner, S.I., Lavik, G., Kuypers, M.M.M., de Beer, D., and Hoffmann, F. (2010) Evidence of nitrification and denitrification in high and low microbial abundance sponges. *Mar Biol* **157**: 593–602.

Shigeno-Nakazawa, Y., Kasai, T., Ki, S., Kostyanovskaya, E., Pawlak, J., Yamagishi, J., *et al*. (2016) A pre-metazoan origin of the CRK gene family and co-opted signaling network. *Sci Rep-UK* **6**: 3439.

Southwell, M.W., Popp, B.N., and Martens, C.S. (2008) Nitrification controls on fluxes and isotopic composition of nitrate from Florida Keys sponges. *Mar Chem* **108**: 96–108.

Spang, A., Poehlein, A., Offre, P., Zumbrägel, S., Haider, S., Rychlik, N., *et al*. (2012) The genome of the ammonia-oxidizing *Candidatus* Nitrososphaera gargensis: insights into metabolic versatility and environmental adaptations. *Environ Microbiol* **14**: 3122–3145.

Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M.E.A., Mitros, T., *et al*. (2010) The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature* **466**: 720–726.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Tian, R.-M., Sun, J., Cai, L., Zhang, W.-P., Zhou, G.-W., Qiu, J.-W., and Qian, P.-Y. (2016) The deep-sea glass sponge *Lophophysema eversa* harbours potential symbionts responsible for the nutrient conversions of carbon, nitrogen and sulfur. *Environ Microbiol* **18**: 2481–2494.

Wang, H.-C., Susko, E., and Roger, A.J. (2006) On the correlation between genomic G+C content and optimal growth temperature in prokaryotes: Data quality and confounding factors. *Biochem Biophys Res Commun* **342**: 681–684.

Wiens, M., Korzhev, M., Perovic-Ottstadt, S., Luthringer, B., Brandt, D., Klein, S., and Muller, W.E.G. (2006) Toll-like receptors are part of the innate immune defense system of sponges (demospongiae: Porifera). *Mol Biol Evol* **24**: 792–804.

Wiens, M., Korzhev, M., Krasko, A., Thakur, N.L., Perović-Ottstadt, S., Breter, H.J., *et al*. (2005) 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* **280**: 27949–27959.

CHAPTER IV

Taurine as a key intermediate for host-symbiont interaction in the tropical sponge Ianthella basta

Authors:

Florian U. Moeller, Craig W. Herbold, Arno Schintlmeister, Maria Mooshammer, Cherie Motti, Faris Behnam, Margarete Watzka, Thomas Schweder, Mads Albertsen, Andreas Richter, Nicole S. Webster, and Michael Wagner

Published as a preprint publication in *bioRxiv* (2022; https://doi.org/10.1101/2022.09.23.509140) and currently in review at *The ISME Journal*.

Taurine as a key intermediate for host-symbiont interaction in the tropical sponge *Ianthella basta*

Florian U. Moeller[1], Craig W. Herbold[1], Arno Schintlmeister[1,2], Maria Mooshammer[1], Cherie Motti[3], Faris Behnam[1], Margarete Watzka[4], Thomas Schweder[5,6], Mads Albertsen[9], Andreas Richter[4], Nicole S. Webster[3,7,8], and Michael Wagner[1,2,9,*]

[1]Centre for Microbiology and Environmental Systems Science, Division of Microbial Ecology, University of Vienna, Austria

[2]Large-Instrument Facility for Environmental and Isotope Mass Spectrometry, Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria

[3]Australian Institute of Marine Science, Townsville, Queensland, Australia

[4]Centre for Microbiology and Environmental Systems Science, Division of Terrestrial Ecosystem Research, University of Vienna, Austria

[5]Institute of Marine Biotechnology e.V., Greifswald, Germany

[6]Institute of Pharmacy, Pharmaceutical Biotechnology, University of Greifswald, Greifswald, Germany

[7]Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, University of Queensland, St Lucia, Queensland, Australia

[8]Australian Antarctic Division, Kingston, Tasmania, Australia

[9]Center for Microbial Communities, Department of Chemistry and Bioscience, Aalborg University, Aalborg, Denmark

*Correspondence to: Michael Wagner
Email: michael.wagner@univie.ac.at

Abstract

Marine sponges are critical components of marine benthic fauna assemblages where their filter-feeding and reef-building capabilities provide bentho-pelagic coupling and crucial habitat. As potentially the oldest representation of a metazoan-microbe symbiosis, they also harbor dense, diverse, and species-specific communities of microbes, which are increasingly recognized for their contributions to dissolved organic matter (DOM) processing. Recent omics-based studies of marine sponge microbiomes have proposed numerous pathways of dissolved metabolite exchange between the host and symbionts within the context of the surrounding environment, but few studies have sought to experimentally interrogate these pathways. By using a combination of metaproteogenomics and laboratory incubations coupled with isotope-based functional assays, we showed that the dominant gammaproteobacterial symbiont '*Candidatus* Taurinisymbion ianthellae' residing in the marine sponge, *Ianthella basta*, expresses a pathway for the import and dissimilation of taurine, a ubiquitously occurring sulfonate metabolite in marine sponges. '*Candidatus* Taurinisymbion ianthellae' incorporates taurine-derived carbon and nitrogen while, at the same time, oxidizing the dissimilated sulfite into sulfate for export. Furthermore, we found that taurine-derived ammonia is exported by the symbiont for immediate oxidation by the dominant ammonia-oxidizing thaumarchaeal symbiont '*Candidatus* Nitrosospongia ianthellae'. Metaproteogenomic analyses also indicate that '*Candidatus* Taurinisymbion ianthellae' likely imports DMSP and possesses both pathways for DMSP demethylation and cleavage, enabling it to use this compound as a carbon and sulfur source for biomass, as well as for energy conservation. These results highlight the important role of biogenic sulfur compounds in the interplay between *Ianthella basta* and its microbial symbionts.

Introduction

Marine sponges are important components of coral reefs with a range of critical ecological functions [1]. Sponges play a crucial role in maintaining biomass in reef ecosystems by filtering dissolved organic matter (DOM) from large quantities of water and capturing it as biomass, which is then shed as particulate organic matter (POM) and passed to higher trophic levels [2–4]. Marine sponges often harbor dense, diverse and stable communities of microorganisms that are thought to play a role in DOM processing as well as a broad suite of other nutrient transformations [3, 5, 6]. These microbial symbionts have been shown to photosynthetically fix carbon [7], fix N_2 [8], nitrify [9, 10], as well as carry out anammox, denitrification, and sulfate respiration [11, 12]. Furthermore, isotope labeling experiments have proven that symbiont-derived carbon and nitrogen can be found in sponge cells [7, 11, 13, 14], with more recent studies revealing an emerging view that not only microbial symbionts can incorporate DOM, but that sponge cells do as well, and can even translocate DOM-derived carbon and nitrogen to their symbionts [15–17].

Whilst the specificity, biogeography and environmental sensitivity of the relationship between sponges and their microbiome has been intensively studied [18–20], our understanding of the physiological interactions between sponges and their symbiotic microorganisms is still very limited. Genome-based investigations of the sponge microbiome have revealed an array of diverse metabolic potential encoded in symbiont genomes, including the processing of organic carbon, nitrogen, and sulfur compounds [21–24], autotrophic sulfur oxidation [25, 26], and vitamin synthesis [27, 28]. These metabolic processes may be mediated by a diverse array of transporters, encoded in the symbionts, to facilitate exchange with the host, as well as the surrounding environment (reviewed in [29]). In a few studies, metaproteomic analyses have revealed the expression of proteins indicative of metabolic interactions between the host and symbionts, including transport functions for typical sponge metabolites, and the potential mediation of host-symbiont interactions via the expression of eukaryotic-like proteins and cell-cell mediators [10, 30]. Despite these insights obtained from a variety of approaches, the high complexity of sponge holobionts, coupled with the inherent difficulty in culturing sponge symbionts, has made it challenging to irrevocably link host interactions and symbiont physiology with symbiont phylogeny.

One way of examining host-symbiont interaction, is by tracking compounds that may be translocated from the host to the symbiont, such as taurine. Taurine has been reported from a broad suite of sponge

species, oftentimes as one of the most abundant free amino acids (up to 39% of the free amino acid pool; [31]). Free-living marine bacteria can utilize taurine produced by phytoplankton [32–34] and the genetic capacity for taurine utilization has been identified in many marine sponge microbiomes [25, 26, 35–37]. Taurine has been proposed to be an important intermediate driving metabolic exchange between marine bacteria and eukaryotic phytoplankton [32–34, 38], and yet, to date, there is no direct experimental evidence for the utilization of host-derived taurine by marine symbionts and of its role in mediating holobiont function, despite the abundance and widespread distribution of this compound in marine animals. Taurine, has however, been recently demonstrated to confer resistance to pathogens in mammalian gut systems, wherein taurine-metabolizing gut microbiota shield the host against infections by converting taurine to sulfide, an inhibitor of cellular respiration [39, 40]. Taurine is an ideal osmoregulator, as it is transported by a Na^+-dependent transport system unique to β-amino acids and so, it is responsive to ionic activities and other osmotic agents [41, 42]. As a zwitterionic compound, extremely high intra- to extracellular concentration gradients for taurine can be maintained, and due to its lipophobic properties, it is not readily lost by diffusion [43]. Taurine also likely functions as a storage and transport compound, providing nitrogen and sulfur to symbionts in the gutless clams *Solemya reidi* [44] and *Solemya velum* [45]. The synthesis of sulfur-containing free amino acids has long been proposed as a strategy for sulfide detoxification and transport in chemoautotrophic symbioses [46–50] and taurine is the dominant free amino acid in many of these species [44, 46–50].

Ianthella basta [51] is a widely distributed sponge species found throughout the Indo-Pacific region [52]. In contrast to many marine sponge species with high microbial abundance of microbial symbionts [53], *I. basta* hosts only a low diversity of microbial symbionts and is populated by three dominant phylotypes including an Alphaproteobacterium, a Gammaproteobacterium and a Thaumarchaeote, which all occur at relatively high abundances in the sponge mesohyl [10, 54, 55]. This “low complexity” community is highly stable across different host morphotypes [56], across the large geographic range of the host, under different host health states [55] and across a wide suite of environmental conditions [57]. Thus, the *I. basta* holobiont represents a comparatively tractable model system for the investigation of sponge-symbiont physiological interactions. Recently, we were able to demonstrate that the thaumarchaeal symbiont of *I. basta* — ’*Candidatus* Nitrosospongia ianthellae’ — is an active ammonia oxidizer and that, surprisingly, this sponge does not harbor a bacterial nitrite oxidizer [10]. However, the physiology of the remaining two symbiont

phylotypes remains unknown. In this study, we elucidate selected physiological traits of the gammaproteobacterial *I. basta* symbiont, which we named 'Candidatus Taurisymbion ianthellae'. Based on annotation of the metagenome-assembled genome (MAG) of this symbiont, we focused on the important biogenic sulfonate intermediate, taurine. We document the expression of a previously described [58], but infrequently observed pathway for the transport and dissimilation of taurine into sulfite, followed by sulfite oxidation into sulfate encoded by the symbiont, and provide experimental physiological support for this sponge-microbe interaction. For this purpose, we performed whole sponge incubations with added taurine and detected the predicted secreted end product, sulfate. We also employed the use of stable isotope-labeled taurine to track the incorporation of taurine-derived carbon and nitrogen into *I. basta* holobiont biomass, as well as the dissimilation of the nitrogen moiety to ammonium and its subsequent oxidation to nitrite.

Materials and methods

Sponge collection for metaproteogenomics and* Ianthella basta *holobiont taurine incubations. Two specimens of the purple thick morphotype of the sponge *Ianthella basta* were collected in waters offshore from Orpheus Island (18°36.878'S, 146°29.990'E), Queensland, Australia, in October 2010 and 2011, and processed as described in [10] for metaproteogenomic analysis. In addition, sponge explants were used for 48-hour batch incubations either in the presence or absence of taurine. The sponge explants used for the taurine incubations were derived from two large yellow thin morph *Ianthella basta* individuals (approx. 0.5 m x 0.5 m x 0.5 cm) collected from Davies Reef (18°49.354 S, 147°38.253 E) on 21st November 2014 and 25th August 2015. Upon collection, the individuals were maintained in flow-through reef seawater onboard the RV Cape Ferguson until its return to the Australian Institute of Marine Science (AIMS) where the individual was cut into small sponge explants between 10 cm x 10 cm x 0.5 cm and 5 cm x 5 cm x 0.5 cm and left to heal in AIMS aquaria overnight. Since *I. basta* individuals are thin, erect and laminar, while also exhibiting incurrent (ostia) and excurrent holes (oscula) that are spaced approximately 1 cm apart and on opposing sides of the sponge, relatively small samples can be representative of the entire aquiferous system [59]. The explants were prepared by cutting rectangular-like shapes, while preserving both sides of the fan-shaped sponge, thereby retaining multiple oscula and ostia on both sides. All explants were monitored at the beginning and end of the incubations for tissue health, and no major signs of tissue health regression or

necrosis were observed. Furthermore, the edges of the cut sides are estimated to represent 4–8% of the volume of each explant (by assuming an edge thickness of 0.1 cm) and thus, the area of tissue wounds are small relative to the size of each explant. Live sponge explants were transferred to sulfate-free artificial seawater (SFASW; see Supplementary Note for media details) and rinsed 3 times to remove as much sulfate from seawater as possible. For all incubation experiments, *I. basta* explants were incubated at 25ºC in the dark in continuously stirred acid-washed glass tanks either filled with 0.2 µm natural filtered seawater (FSW) or SFASW, and with or without taurine added at different time points and concentrations. All incubations were conducted in the dark to preclude light inhibition of nitrification. SFASW incubations were conducted to more readily discern the production of sulfate from the taurine-dissimilation/sulfite oxidation pathway, since natural seawater contains sulfate at high concentrations. In the first set of experiments, 1 mM unlabeled taurine was added at the beginning of the incubations and another 0.6 mM pulse was added at t = 36 h. For stable isotope incubations in standard seawater, ^{15}N and ^{13}C labeled taurine (2-Amino-^{15}N-ethanesulfonic acid: Taurine-^{15}N, 98 atom%; 2-Aminoethane-$^{13}C2$-sulfonic Acid: Taurine-$^{13}C2$, 99 atom%; Campro Scientific GmbH, Berlin, Germany) were added in equimolar concentrations (i.e. a total of 0.8 mM of ^{15}N-taurine and 0.8 mM of ^{13}C-taurine were added by the end of the incubations) simultaneously, so as to monitor potential incorporation into *I. basta* holobiont biomass as well as to track dissimilation into NH_4^+, followed by successive oxidation to NO_2^- by the dominant thaumarchaeal symbiont [10]. For these two sets of experiments, incubations were conducted in closed 1 l acid-washed glass aquaria at room temperature and water sampling occurred at t = 0, 12, and 48 h, while sampling of *I. basta* biological material occurred at the end of the experiment at t = 48h. These two experiments utilized 12 explants (approx. 10 cm x 10 cm 0.5 cm; 9.14 ± 2.4 g sponge wet wt) from the individual collected in November 2014 (3 replicates per treatment with *I. basta* explants: SFASW + *I. basta*; SFASW + taurine + *I. basta*; FSW + *I. basta*; FSW + $^{13}C^{15}N$-labeled taurine + *I. basta*). An additional follow-up experiment was conducted with explants derived from the individual collected in October 2015, in SFASW in closed 250 ml acid-washed glass aquaria but with higher temporal resolution (t = 0, 12, 24, 36, and 48 h) and 1 mM or 100 µM isotopically unlabelled taurine added at the beginning of the incubation. This set of experiments utilized 45 explants (approx. 5 cm x 5 cm 0.5 cm; 2.98 ± 0.82 g sponge wet wt), as each explant was sacrificially sampled at the end of each time point. For all sets of incubations parallel control incubations were performed under identical conditions, but without *I. basta* explants. All incubations were conducted in triplicates. We treated the explants from the same *I. basta*

individual as non-technical replicates since differences in symbiont distribution within the relatively large sponge individuals used here, may occur.

For chemical analyses, seawater samples (10 ml) were taken from the containers (see above-described set of experiments) and filtered using 0.45 µm Sartorius Minisart cellulose acetate filters (Göttingen, Germany). Duplicate samples for dissolved inorganic nitrogen (NH_4^+, NO_2^-, NO_3^-) were measured on a Seal AA3 segmented flow analyser and referenced against OSIL standards and in-house reference samples. Sulfate was measured via ion chromatography, followed by detection via suppressed conductivity. Confirmation of taurine uptake by *I. basta* was established following a two-step pre-column fluorometric derivatization HPLC method adapted from [60]. o-Phthalaldehyde-ethanethiol (OPA-ET; Sigma Aldrich; 750 µl), prepared daily in methanol (MeOH; Sharlau) buffered with 0.8 M borate (Sigma Aldrich; pH11) and aged at 4°C for 2 h, was added to 100 µl of seawater sample diluted in 900 µl MilliQ H_2O. After 1 min, 50 µl of fluorenylmethyloxycarbonyl chloride (FMOC-Cl) [Sigma Aldrich], prepared in acetonitrile [Sharlau] and aged for at least 12 h, was added; the higher concentration of the derivatizing reagents was required to ensure complete reaction of the taurine (~mM). After a further 1 min, 10 µl of the derivatized sample was injected onto an Agilent 1100 HPLC system (comprising a binary pump, autosampler, column oven and fluorescence detector, and operated using Chemstation for LC 3D Rev. A.10.01 and 735 Sampler Software v6.10). Samples (n = 2 samples per treatment per timepoint; n = 5 replicate injections per sample) were chromatographed on a 250 mm× 4.6 mm i.d., 5 µm particle size, HILIC (Phenomenex) column under the following conditions: 1 ml min^{-1} flow rate; column maintained at 40°C; mobile phases comprising MeOH-sodium acetate (0.36 M; pH 7.0)-H_2O (55:8:37, v/v/v) (A) and acetonitrile-sodium acetate (0.36 M; pH 7.0)-H_2O (10:5:85, v/v/v) (B); gradient 15% A at 0–2 min, 15–20% A at 2–6 min, 40% A at 6–9 min, 85% A at 9–17 min then held until 20 min; the column was flushed with 100% A at 20–21 min and held for a further 4 min before re-equilibration at 15% A for 5 min; fluorescence detection (excitation: 330 nm, emission: 460 nm). For calibration, a taurine stock solution (1 M; >99.9% Sigma Aldrich) was prepared in Milli-Q water and a dilution series prepared (10 mM–1 nM). γ-Aminobutyric acid (5 M GABA; Sigma Aldrich) was used as the internal standard. The dilution series was stored at 4°C. The calibration curve (R^2 = 0.99) was used to quantify taurine in seawater. ^{15}N-analyses of ammonium and nitrite from the incubation experiments with stable-isotope labeled taurine are described below.

Metagenome assembly and genome binning. Sequencing libraries were prepared from the same DNA extracted from the individual collected in October 2010 and previously reported in [10] using the Nextera kit (Illumina) according to the manufacturer's instructions and concentrations measured using the QuantIT kit (Molecular Probes, Life Technologies, Naerum, Denmark). The libraries were paired-end (2 × 150 bp) sequenced on an Illumina HiSeq2000 using the TruSeq PE Cluster Kit v3-cBot-HS and TruSeq SBS kit v.3-HS sequencing kit and on an Illumina MiSeq using v3 2 × 300 bp kits. Metagenome reads in fastq format, obtained from the Illumina sequencing runs, were end-trimmed at a minimum phred score of 15, a minimum length of 50 bp, allowing no ambiguous nucleotides and Illumina sequencing adaptors removed. Trimmed reads from each dataset were assembled using Spades version 3.11.0 [61], using default parameters and genomes were binned using Metabat v 2.12.0 [62]. MAGs from multiple Illumina datasets were dereplicated using dRep [63] and provisionally classified using CheckM [64]. A single high-quality gammaproteobacterial MAG was recovered after dereplication and uploaded to MaGe [65] for annotation. The mapped read pairs (at >98% identity) for the assembly of this MAG represented 3.02% of total QC paired reads and read coverage was approx. 26X. The MAG was assembled from run ERR884062 under BioProject PRJEB9378.

Two additional MAGs were assembled and binned from paired end reads from SRA runs derived from two sponge species, belonging to the same family as *Ianthella basta* (*Hexadella dedritifera* and *Hexadella* cf. *dedritifera*) [66]. Paired end reads for SRA runs SRR8088660, SRR8088661, SRR8088662 and SRR8088663 were downloaded from NCBI. BBduk (bbmap v. 37.61) was used to remove adapters and residual phiX (adapters.fa, phix174_ill.ref.fa.gz, ktrim=r k=21 mink=11 hdist=2), then quality filtered (minlen=99 qtrim=r trimq=15). Read sets were denoised using BayesHammer (SPAdes v3.15.3) [61] and merged using bbmerge (bbmap v. 37.61). Reads were assembled using SPAdes v3.15.3 (-k 21,33,55,77 --only-assemble) and metaspades (-k 21,33,55,77 --only-assemble) resulting in 8 assemblies (2 assemblers, 4 datasets). Each read set was mapped against each assembly using bbmap v. 37.61(minid=0.99 idfilter=0.99 ambiguous=best pairedonly=t killbadpairs=t mappedonly=t) for and MAGs were generated using metabat1 and metabat2 algorithms (metabat v 2.15). MAGs from each SRA run were then dereplicated with dRep v 1.4.3 dereplicate_wf (-comp 40 -con 10 -l 200000) [63]. "Winning" MAGs from each SRA run were

combined and dereplicated using dRep v 1.4.3 (-comp 40 -con 10 -l 200000 -pa 0.6 --S_algorithm gANI --S_ani 0.965). Reads from each SRA run were remapped to each MAG using bbmap v. 37.61(minid=0.99 idfilter=0.99 ambiguous=best pairedonly=t killbadpairs=t mappedonly=t and relative abundance was calculated by dividing mapped fragments by total fragments for each metagenomic dataset.

Cryosectioning and FISH for relative symbiont quantification. The aforementioned *I. basta* individual collected for metagenomic sequencing in October 2010 was assayed for its microbiological composition using fluorescence *in situ* hybridization (FISH). Briefly, after sample collection, the *I. basta* specimen was cut into tissue strips of approximately 4 cm^3, fixed in 4% paraformaldehyde (PFA) for 1 h at room temperature and stored in ethyl alcohol (EtOH)-phosphate-buffered saline (PBS) (1:1) at -20ºC. For FISH, PFA-fixed samples of *I. basta* were embedded in Neg-50 (Richard-Allan Scientific), a cryogenic water-soluble sectioning medium, and cut to 5-µm thin sections (Leica CM3050 S). We designed specific 16S rRNA-targeted FISH probes for the visualization of '*Ca*. Taurinisymbion ianthellae' (GamD1137, 5'-CTC AAA GTC CCC GCC ATT-3') and the dominant alphaproteobacterial symbiont (AlfD729, 5`-CGGACCTGGCGGCCGCTT-3'). To calculate the relative abundance of '*Ca*. Taurinisymbion ianthellae', we applied the double-labeled FISH probe GamD1137 (Cy3), along with the double-labeled FISH probe for the α-symbiont (AlfD729 in Fluos) and the double-labeled probe mix EUB338-I, EUB338-II, and EUB338-III (all in Cy5; [67]) in equimolar concentrations. *I. basta* sponge section hybridizations were carried out in the presence of 25% formamide in the hybridization buffer (after formamide concentration optimization), and the stringency of the washing buffer was adjusted accordingly [68]. As a negative control, the nonsense probe nonEUB338-I (reverse complementary probe to EUB338-I; double labeled in FLUOS, Cy3, and Cy5) was applied on one section per well per slide hybridized [69]. Probe-conferred visualizations of fluorescent cells were carried out by a confocal laser scanning microscope (CLSM) (LSM 510 Meta; Zeiss, Oberkochen, Germany). Specifically, labelled bacterial cells were counted by eye in 10 randomly selected fields of view (each field contained an average of 260.6 ± 73 SD cells of '*Ca*. Taurinisymbion ianthellae' at 630X magnification with an optical section thickness of 1 µm) on a 146.25 × 146.25 µm ocular grid. Based on these parameters and measurements we were able to estimate the average density of '*Ca*.

Taurinisymbion ianthellae' within the *I. basta* mesohyl as well as the fraction of '*Ca*. Taurinisymbion ianthellae' relative to total bacterial cells stained by the EUB338 probe mix.

DNA extraction from whole sponge tissue and qPCR for symbiont quantification. Between 80–150 mg of *I. basta* sponge tissue was thawed (if previously frozen at -80 ºC), rinsed successively (3x) in 1X calcium- and magnesium-free artificial seawater (CMF-ASW; see Supplementary Note for media details) and immediately ground into a paste with a mortar and pestle in liquid N_2. After resuspension in TE buffer, DNA was extracted from the suspension using the same methods as described in [10] for DNA extraction prior to symbiont quantification via quantitative PCR (qPCR). DNA was extracted from the two individuals used for metaproteogenomics (see above) as well as three additional yellow thin morphotype healthy individuals from a previous study examining the persistence of microbial symbionts in diseased and healthy specimens of *I. basta* (Luter *et al*., 2010). qPCR was used to quantify the number of 16S rRNA genes of '*Ca*. Taurinisymbion ianthellae', the thaumarchaeal ammonia-oxidizer '*Ca.* Nitrosospongia ianthellae', and the dominant α-proteobacterial symbiont using specific primers designed for each symbiont phylotype in [10].

Phylogenetic analyses. Sequences for the '*Ca*. Taurinisymbion ianthellae' 16S rRNA species tree were assembled by merging a subset of sequences described by [70], along with top BLAST hits from representative gammaproteobacterial sequences, and the 16S rRNA sequences from the most closely related members with sequenced genomes. These sequences were aligned via the SINA aligner and the phylogenetic trees constructed in IQ-Tree, version 1.6.2 [71] using the model TN+F+R6 (after model selection using Model Finder; [72] with 10,000 ultrafast bootstraps (UFBoot, [73]). In addition, 16S rRNA gene sequences related to '*Ca*. Taurinisymbion ianthellae' were identified in short read archive (SRA) datasets using IMNGS (www.imngs.org, [74]) with default parameters. Successfully mapped reads were then aligned to the reference alignment using SINA and placed into the reference tree using the Evolutionary Placement Algorithm (EPA; [75]) implemented in RAxML-HPC 8.2.11 [76]. Phylogenomic reconstruction was based on a concatenated amino-acid alignment of 43 marker genes constructed with CheckM [64] and trees were constructed using the model LG+F+R7 (after model selection using Model Finder; [72]) with 1000 ultrafast bootstraps (UFBoot, [73]).

Proteomics. Proteome analysis was performed as done previously [10]. In brief, protein extracts were separated by 1D PAGE followed by liquid chromatography-based mass spectrometry (1D-PAGE-LC-MS/MS). MS spectra and MS/MS spectra were acquired from eluting peptides in a LTQ Orbitrap Velos hybrid mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA), as described previously [77, 78] and searched against predicted protein sequence databases composed of the *I. basta* symbiont-enriched metagenome bins and common laboratory contaminants using the Sorcerer SEQUEST (v.27, rev. 11) algorithm. Protein identifications were filtered with Scaffold version 3.5.1 as described previously [79]. For protein identification only peptides identified with high mass accuracy (maximum ± 10 ppm difference between calculated and observed mass) were considered and at least two exclusively unique peptides were required to identify a protein. False-discovery rates (FDRs) were estimated with searches against a target-decoy database [80, 81]. For relative quantitation of proteins, normalized spectral abundance factors (NSAF) were calculated for each sample according to the method of [82] and averaged for all replicates and samples. The proteomics raw data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD012484 and 10.6019/PXD012484.

***NMR-, LC-MS-based detection and quantification of taurine in* Ianthella basta.**

To establish the presence of taurine in *I. basta*, a section of tissue from a yellow thin morphotype was excised by scalpel, weighed wet, and exhaustively extracted with 90% MeOH in MilliQ water. The MeOH extract was lyophilized overnight (Dynavac Freeze-drier). The dry MeOH extract (~5 mg) was dissolved in 1 ml deuterated methanol (CD_3OD; Cambridge Isotope Laboratories) for analysis by proton nuclear magnetic resonance (^{1}H NMR; Bruker Avance 600 MHz, 5 mm CPTXI inverse ^{1}H-^{13}C/^{15}N Z-gradient cryogenically cooled probe, referenced to CD_3OD 3.31 ppm, standard 1D zg pulse sequence). Given the crude nature of the extract and the likelihood that the diagnostic signals would be unable to be distinguished from similar chemistry in the extract, 20 mg (dissolved in ~1 ml 50% MeOH) was chromatographed over a pre-equilibrated Strata C18-E 55 μm, 70Å, 1000 mg solid phase extraction cartridge (C18-SPE; Phenomenex) and eluted with 10 ml of H_2O, 10% MeOH, 50% MeOH and 100% MeOH. The four fractions (Fractions 1–4) were evaporated under nitrogen stream at 40°C and lyophilized to dryness. Fractions 1–2 were prepared in

100% deuterated H_2O (D_2O; Cambridge Isotope Laboratories; referenced at 4.39 ppm) and Fractions 3–4 in 10% CD_3OD: 90% CD_3OD (referenced to 3.31 ppm) and analysed by ^1H NMR. ^1H and COSY NMR spectra of 100 μM taurine standard, prepared in D_2O, was used to confirm identification of taurine in Fraction 1. Final confirmation was provided by spiking the ^1H NMR sample with 32 μl and 128 μl 100 μM taurine and measuring the increase via integration of the two triplet signals.

To corroborate the NMR assignment, Fraction 1 was also analysed by LC-MS. A 10 μl aliquot of the aqueous Fraction 1 was injected onto an Agilent 1100 HPLC system coupled to a Bruker Esquire 3000 Quadrupole Ion Trap LC-MS and eluted from a 150 mm x 3 mm Luna 5μ NH_2 (Phenomenex) column under the following conditions: 0.8 ml min^{-1} flow rate; column maintained at 40°C; mobile phases comprised of H_2O + 0.1 % formic acid (A) and MeOH + 0.1% formic acid (B); isocratic elution at 90% A: 10% B for 10 min. The ionspray voltage was set to -4500 V, declustering potential -500V, curtain gas flow 9 l min^{-1}, ion source gas 40 PSI, and spray temperature at 350°C. Data was acquired over the mass range 70–200 *m/z*; product ion MS/MS data was acquired over 50–200 *m/z* to establish the fragmentation pattern. Retention time and total and extracted ion chromatographs (positive mode) were compared against a 100 μM taurine standard prepared in MilliQ.

The concentration of taurine in sponge tissue was measured by quantitative NMR (qNMR) following [83]. Briefly, five sponge explant samples were extracted and chromatographed over C18–SPE as described above. ^1H NMR spectra of Fraction 1 (~16 mg dry extract) from each of the five sponge extracts were acquired in 800 μl D_2O (δH 4.39) at 298 K, with sweep width 12 ppm (7184 Hz), 90° pulse, 35 sec relaxation delay, receiver gain 16, 2 dummy scans, 16 acquisition scans and 64 k data points. The external reference signal was calibrated using a stock solution of 2 mM taurine in D_2O. The concentration of taurine in the NMR samples was determined by comparing the signal integrals of well resolved non-exchangeable protons (2H) centred at 3.26 ppm in a 0.20 ppm window with that of the reference signal. Concentration was normalized to wet weight of sponge (wet wt).

Isotopic analysis of whole sponge tissue, total nucleic acids, and ^{15}N in NH_4^+ & NO_2^- using EA-IRMS. For the determination of ^{13}C- and ^{15}N-enrichment within the holobiont, sponge tissue samples were rinsed briefly

3X in 1X CMF-ASW, freeze-dried, ground to a fine powder in a mortar, and stored dry prior to analysis. $\delta^{13}C$- and $\delta^{15}N$-values of sponge tissues were determined by an elemental analyzer (EA 1110, CE Instruments, Milan, Italy) coupled to an isotope ratio mass spectrometer (IRMS; Finnigan MAT Delta^{Plus} IRMS with a Finnigan MAT ConFlo III interface). For determination of ^{13}C- and ^{15}N-enrichment in total nucleic acids, *I. basta* whole sponge tissue was homogenized using the above procedure employing liquid N_2, followed by nucleic acid extraction according to the procedures outlined in [84]. The extracted nucleic acids were then dried overnight in an oven, and $\delta^{13}C$- and $\delta^{15}N$-values were determined via the same procedure as above. For quantification of the amount of ^{13}C- and ^{15}N-derived taurine that was incorporated into *I. basta* holobiont biomass, the following formulas were used:

$$n_{C,taurine} = \frac{\left(a_{13C,\text{sponge}} - a_{13C,ctrl}\right)}{\left(a_{13C,taurine} - a_{13C,ctrl}\right)} n_{\text{C},total}$$

$$n_{N,taurine} = \frac{\left(a_{15N,\text{sponge}} - a_{15N,ctrl}\right)}{\left(a_{15N,taurine} - a_{15N,ctrl}\right)} n_{\text{N},total}$$

where the displayed symbols, abbreviations and designations refer to: *a* — isotope fraction $^{13}C/(^{12}C+^{13}C)$ or $^{15}N/(^{14}N+^{15}N)$ given in at%; *n*: number of carbon or nitrogen elemental atoms; *sponge* — sponge material; *ctrl* — natural isotope abundance control; *total* — total amount of carbon or nitrogen in the analyzed sample. The denominator in both equations ($a_{^{15}C,N,taurine} - a_{^{15}C,N,ctrl}$) were calculated via a standard two-pool mixing model [85]. For comparability of the measurement values obtained from distinct incubations, *n* was normalized to the wet weight of the utilized *I. basta* explants.

For ^{15}N-analysis of NH_4^+ and NO_2^-, sampled seawater from the glass tanks with the labeled taurine incubations was filtered through pre-combusted GF/Fs (Whatman International; treated for 4 h at 450˚C), and subsequently through 0.2 µm polycarbonate filters (Sartorius). All samples were immediately frozen at -20 ºC for later analysis. Prior to shipment to the University of Vienna, samples were thawed at room temperature and the microbial inhibitor phenylmercuric acetate was added (to a final concentration of 10 µM). Upon arrival in Vienna the samples were promptly stored at -80ºC. For determination of the ^{15}N content in ammonia, NH_4^+ was extracted from samples via microdiffusion as described in [10] and following the protocol from [86]. Briefly, 100 mg MgO and an acid trap (acidified cellulose filter disc enclosed in a semi-permeable Teflon membrane) were added to 9 ml of sample and 1.5 ml of 3 M KCl. After 5 days

shaking at 35°C, the acid traps were removed, dried over concentrated sulfuric acid and analyzed for the ^{15}N content by an elemental analyzer (EA 1110, CE Instruments, Milan, Italy) coupled to an isotope ratio mass spectrometer (IRMS; Finnigan MAT Delta^{Plus} IRMS with a Finnigan MAT ConFlo III interface). The nitrogen isotope composition of NO_2^- was determined by a method based on the reduction of NO_2^- to N_2O by using sodium azide under acidified conditions following the protocol of [87]. Briefly, 1 ml sample or standard was transferred to 12 ml exetainer and 1ml 1M HCl was added. After purging the vials with helium to expel air-N_2O from the sample headspace, 150 µl 1M sodium azide buffer (in 10% acetic acid solution) was injected and the vials were placed on a shaker at 37 °C for 18 h. The reaction was quenched by injecting 250 µl of 10M NaOH. Derived N_2O was analyzed using a purge-and-trap GC/IRMS system (PreCon, GasBench II headspace analyzer, Delta Advantage V IRMS; Thermo Fischer). Equivalently to the above calculations, the amount of taurine-derived NH_4^+ and NO_2^- were calculated as follows:

$$n_{NH_4^+,taurine} = \frac{\left(a_{15N,\,NH_4^+} - a_{15N,ctrl}\right)}{\left(a_{15N,taurine} - a_{15N,ctrl}\right)} n_{NH_4^+,total}$$

$$n_{NO_2^-,taurine} = \frac{\left(a_{15N,\,NO_2^-} - a_{15N,ctrl}\right)}{\left(a_{15N,taurine} - a_{15N,ctrl}\right)} n_{NO_2^-,total}$$

By taking into account that the isotopically labeled nitrite emerges from oxidation of ammonia obtained from taurine dissimilation, the total amount of NH_4^+ released by the holobiont through taurine metabolization was calculated via:

$$n_{NH_4^+,released} = \left(n_{NH_4^+,taurine,sponge} - n_{NH_4^+,taurine,SW}\right) + \left(n_{NO_2^-,taurine,sponge} - n_{NO_2^-,taurine,SW}\right)$$

where SW refers to the values determined from control incubations utilizing seawater without sponge explants. For comparability of the measurement values obtained from distinct incubations, *n* was normalized to the wet weight of the utilized *I. basta* explants.

NanoSIMS analysis of **Ianthella basta**. NanoSIMS measurements were undertaken on two of the three *I. basta* explant samples obtained from the $^{13}C^{15}N$-taurine incubations. Specimens were selected for maximum isotope enrichment, based on results from EA-IRMS bulk analysis (0.62 and 0.55 at% ^{15}N; 1.19 and 1.15 at% ^{13}C). Briefly, approximately 0.5 cm^3 sections were cut from *I. basta* samples, which had been fixed in 4% PFA and stored in a 1:1 mixture of PBS and 96% ethanol at -20º C. The thus obtained specimens were

then rinsed successively (3x) in 1X calcium- and magnesium-free artificial seawater (CMF-ASW) and gently dissociated in 1X CMF-ASW with a scalpel, mortar and pestle, and a final step of sonication for two 30 second cycles at 35% power using a sonication probe [Sonopuls HD2200 equipped with probe MS73 (output frequency 20 kHz, power 200 W, amplitude: 310 µm), Bandelin, Germany]. The suspensions were then filtered through 0.45 µm cellulose acetate filters (type Minisart, Sartorius, Göttingen, Germany) to remove a large portion of remaining sponge cells and nuclei (microscopic visualizations showed a lack of sponge nuclei after filtration), centrifuged at 10,844 × g for 10 min at 4°C to pellet cells, and resuspended in 50% EtOH/50% 1X PBS solution. To visualize all DNA-containing cells present in the sample, a 30 µL aliquot of the suspension was spotted onto antimony-doped silicon wafer platelets (7.1 x 7.1 x 0.75 mm; Active Business Company, Germany) and dried at 46º C, before staining with a 0.1 % (w/v) solution of the DNA-binding dye 4,6-diamidino-2-phenylindole (DAPI) (Sigma, Deisenhofen, Germany) at room temperature for 5 min. Excess DAPI was removed by rinsing with distilled water and the slides were air-dried. Samples were then imaged on an epi-fluorescence laser microdissection microscope (LMD, Leica LMD 7000) using a 40× air objective for pre-selection of suitable NanoSIMS measurement areas, which were subsequently marked utilizing the LMD laser. High resolution imaging of fluorescent cells in marked regions was carried out by a confocal laser scanning microscope (CLSM) (LSM 510 Meta; Zeiss, Oberkochen, Germany). NanoSIMS measurements were performed on a NanoSIMS 50L (Cameca, Gennevilliers, France) at the Large-Instrument Facility for Environmental and Isotope Mass Spectrometry at the University of Vienna. Further details on NanoSIMS data acquisition and analyses are presented in the Supplementary Information.

Results and Discussion

A metagenome assembled genome and phylogenetic analyses of the dominant γ-proteobacterial symbiont of Ianthella basta

From the extensive metagenomic data set used in [10] for genomically characterizing the thaumarchaeal symbiont of *I. basta*, we retrieved a 2.35 Mbp MAG consisting of 97 contigs, representing a nearly complete genome (94.8%) with a very low level of contamination (Table S1). Phylogenetic analysis of the only 16S rRNA gene copy in the MAG demonstrated that this gene specifically clustered at 99% nucleotide identity with previously recovered sequences of the dominant gammaproteobacterial symbionts from *I. basta* (Fig.

S1) [10, 55, 56] and fall into the neighborhood of sponge specific clusters 144–148 (SC) [70]. A phylogenomic reconstruction using concatenated single copy conserved marker genes (Fig. 1) [64], confirmed the affiliation of this MAG with the Gammaproteobacteria and revealed that it can be further classified into the Arenicellales order and within the LS-SOB family (originally named after the bacterial MAG extracted from the glass sponge *Lophophysema eversa,* which encodes a sulfur oxidation pathway [88]; based on the Genome Taxonomy Database [89]).

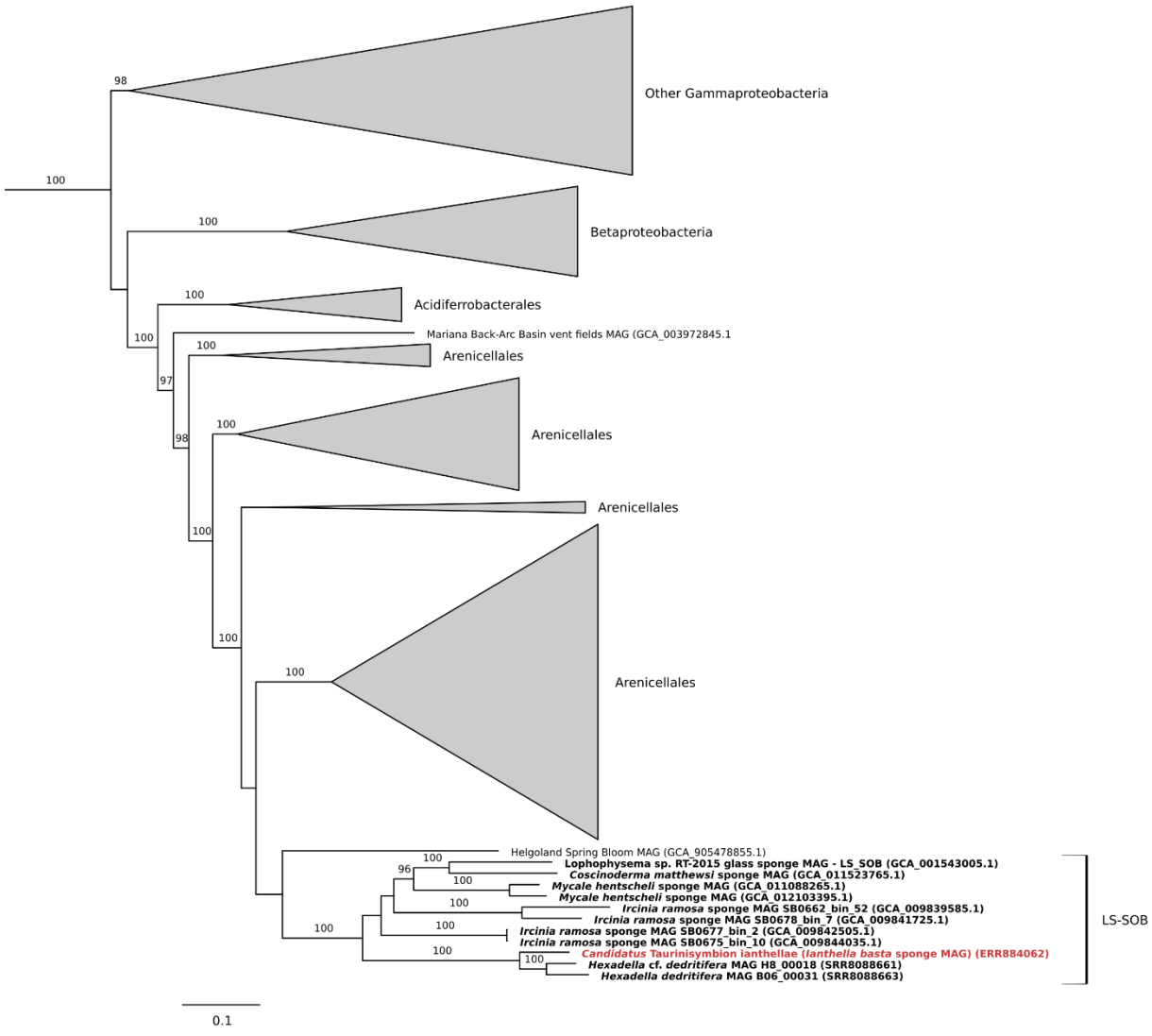


Figure 1. Maximum-likelihood phylogenomic tree (after automatic model selection with IQ-Tree) of '*Ca*. Taurinisymbion ianthellae' (red and bold) along with other selected proteobacterial symbionts (sponge symbionts in bold font). The phylogenomic tree is based on 43 concatenated universal, single-copy marker genes identified with CheckM.

Based on the concatenated gene tree analyses, we further queried representative 16S rRNA gene sequences pulled from LS-SOB genome bins, against the IMNGS 16S rRNA gene amplicon database [74],

and placed the hits into the reference 16S rRNA gene tree using the Evolutionary Placement Algorithm [75]. As a result, 3 sets of short reads were placed into the branch leading to the dominant γ-proteobacterial symbiont sequences obtained from *I. basta* (Fig. S1). The closest two groups (I4, I6) comprised 5 (out of 6 total) OTU reads derived from unspecified coral and sponge samples, 4 of which were OTU reads from corals sampled from the same study site as *I. basta* for this study (Orpheus Island, Queensland, Australia). 141 OTU reads were placed into the I7 group, of which 29 and 4 were derived from sponge and coral metagenomes, respectively (the other 112 OTU reads were derived from other metagenomic data sets spanning a variety of environmental matrices). Most interestingly, out of this group, three OTU reads were obtained from two sympatric marine sponge species (NE Atlantic) within the same genus, (*Hexadella dedritifera* and *Hexadella* cf. *dedritifera*), both of which are also affiliated with the same sponge family as *I. basta*, the Ianthellidae. These two closely related phylotypes, from *H. dedritifera* and *H*. cf. *detritifera*, respectively, exhibited 16S rRNA gene identities of ~97% (over 418 bp) with each other, 95–97% with the *I. basta* γ-symbiont, and were found to comprise 33–42% of the total amplicon dataset derived from these sponges [90]. Noting that these two Hexadella species harbor abundant and closely related γ-phylotypes, and also contain AOA-symbionts closely related to the AOA in *I. basta* [66], we assembled and binned two new γ-symbiont MAGs from the published raw read data set of these two Hexadella sponges.

The phylogenomic reconstruction confirmed that these two Hexadella γ-symbiont MAGs were most closely related to the *I. basta* γ-symbiont (Fig. 1), and AAI analyses indicated that the *I. basta* γ-symbiont shares AAI values of 76–80% with these two Hexadella γ-symbionts. Among the other LS-SOB genomes retrieved from marine sponges [37, 88, 91–93], the *I. basta* γ-symbiont shares AAI values of 54–59%. Since these identity values signify that the γ-symbiont represents a new species within a new genus along with the other two Hexadella γ-symbionts — 60–80% AAI is typical for organisms grouped at the genus level [94, 95] — we propose the name '*Candidatus* Taurinisymbion ianthellae'. Taurinisymbion gen. nov. (Tau.ri.ni.sym'bi.on. N.L. n. taurinum, taurin; Gr. pref. sym, together; Gr. masc. part. n. bion, a living being; N.L. masc. n. Taurinisymbion, a taurin-degrading living partner) for the *I. basta* γ-symbiont. The name *Taurinisymbion* describes this organism's ability to utilize taurine as a source of energy and biomass whilst residing in symbiosis within a sponge, and we were able to identify the same pathways encoded by '*Ca*. Taurinisymbion ianthellae' for taurine utilization in the two Hexadella γ-symbiont MAGs, thereby suggesting a conserved evolutionary function that may be adaptive for this proposed genus (see below; Table

S2). The species name *ianthellae* refers to its discovery and description as a symbiont of the marine sponge, *I. basta*.

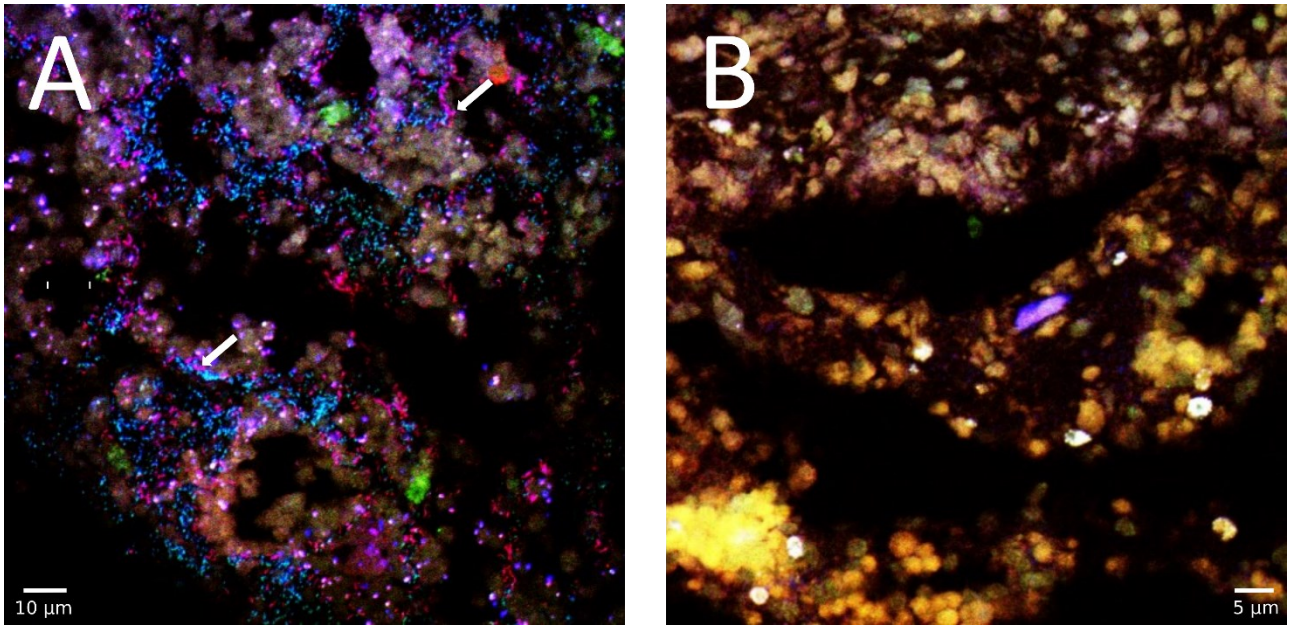


Figure 2. Fluorescence *in situ* hybridization of 5 µm thick cryo-sections of *Ianthella basta* using double-labeled probe sets [55]. A) '*Ca.* Taurinisymbion ianthellae' (red), *Alphaproteobacteria* symbiont-probe (green), and *Bacteria*-probe (EUB338-I-III; blue); B) negative controls using the double-labeled non-EUB338 probe set are depicted in all three color channels. The combination of the '*Ca.* Taurinisymbion ianthellae' specific probe signals in red with the general *Bacteria* probe signals in blue, results in the visualization of '*Ca.* Taurinisymbion ianthellae' as magenta signals and representative cells are indicated with white arrows. The α-symbiont appears in cyan. Green structures represent autofluorescence.

'Candidatus Taurisymbion ianthellae' is an abundant symbiont of Ianthella basta

Consistent with the high coverage of reads and peptides assigned to '*Ca*. Taurinisymbion ianthellae' in the metaprotoegenomic data sets from *I. basta* and the dominance of this phylotype in other *I. basta* studies [54–56], we observed a high abundance of this symbiont (Fig. 2) in the mesohyl of *I. basta* using fluorescence *in situ* hybridization (FISH) with a newly designed specific 16S rRNA-targeted probe. Quantitative FISH revealed that '*Ca*. Taurinisymbion ianthellae' achieved average cell densities of 1.22 ± 0.34 (SD) × 10^{10} cells cm^{-3} sponge and accounted for 24 ± 5.6% (SD) of all cells hybridizing with the probe set EUB338-I-III targeting most bacteria (which does not cover the thaumarchaeal symbiont of *I. basta*). Moreover, absolute quantification of '*Ca*. Taurinisymbion ianthellae' using qPCR with newly designed symbiont specific primers, showed an average absolute abundance of 2.78 ± 1.5 (SD) × 10^{10} 16S rRNA gene copies g^{-1} sponge wet wt, which is in the same range as the population density values observed from the FISH images (assuming that 1.2 g sponge wet wt = 1 cm^{-3}; [96]). The absolute abundances obtained for the dominant

thaumarchaeal and alphaproteobacterial symbionts are also in the 10^{10} 16S rRNA gene copies g^{-1} sponge wet wt range (Fig. S2). The ratio of '*Ca*. Taurinisymbion ianthellae' 16S rRNA sequences to the total 16S rRNA genes derived from qPCR assays (targeting the γ-, α-, and thaumarchaeal symbionts of *I. basta*) was 41.1 ± 34 (SD)%. Nevertheless, it should be noted that qPCR abundances cannot directly be translated to cell numbers as (i) symbionts can have multiple rRNA gene copies (although the '*Ca*. Taurinisymbion ianthellae' MAG has only one 16S rRNA gene copy), (ii) genome polyploidy might occur in sponge symbionts and (iii) extracellular DNA of the symbionts will also be detected. Similarly, using FISH for quantification may undercount microbial cells with low ribosome contents.

'Ca. Taurinisymbion ianthellae' ***expresses proteins for taurine catabolism and subsequent NH_4^+ and SO_4^{2-} export***

To analyse the metabolic potential of '*Ca*. Taurinisymbion ianthellae', we annotated its MAG by employing the metaproteogenomic data set presented in [10] and were able to assign 149 proteins to this symbiont MAG. These proteins represented 35.7%, in NSAF values, of all proteins assigned to the *I. basta* microbiome (Table S3), highlighting the physiological dominance of '*Ca*. Taurinisymbion ianthellae'. From this metaproteogenomic dataset, we concluded that '*Ca*. Taurinisymbion ianthellae' is a facultative anaerobe with an ability to metabolize a variety of organic compounds. The genomic potential for the major central metabolic pathways, such as glycolysis, the pentose phosphate pathway, the tricarboxylic acid (TCA) cycle, the glyoxylate bypass, a conventional respiratory chain (complex I -IV; two respiratory complex IV were found to be encoded, a cbb_3- and a aa_3-type cytochrome c oxidase; complex V), as well as the genes for denitrification of NO_2^- to N_2O, and DMSP catabolism, were all present in the recovered MAG (Fig. 3, Table S2). A further discussion of a few of these genomic features, with a focus on DMSP metabolism, is provided in the Supplementary Information.

Most notably, '*Ca*. Taurinisymbion ianthellae' encodes a previously described pathway for taurine import and transamination (*tauABC, tpa, xsc*) [97] into sulfite via a desulfonated intermediate sulfoacetaldehyde, as well as a pathway for the subsequent oxidation of the resultant toxic and highly reactive sulfite into sulfate via the cytoplasmic enzymes APS reductase (a*prAB*) and ATP sulfurylase (*sat*) [58]. This pathway consumes AMP and sulfite, and releases ATP via substrate-level phosphorylation and sulfate, while channeling two electrons from the oxidation of sulfite, with its very negative redox potential,

to the quinone pool via the membrane-bound QmoAB. Consistent with other sulfur oxidizing bacterial genomes, the *qmoC* gene is replaced by *hdrBC* genes ([98] and references therein), with the *hdrBC* genes found to be adjacent to *qmoAB* in '*Ca*. Taurinisymbion ianthellae', and the HdrB subunit expressed in the holobiont (Fig. 3, Tables S2 & S3). For the potential export of sulfate, two putative sulfate permeases (COG0730 permease family, [99]; COG0659, SulP family permease), were detected in the '*Ca*. Taurinisymbion ianthellae' MAG. Interestingly, the gene encoding taurine dioxygenase (*tauD*), an alpha-ketoglutarate-dependent dioxygenase, which desulfonates taurine into aminoacetaldehyde (the fate of which is currently unknown) and sulfite, was also found encoded by '*Ca*. Taurinisymbion ianthellae'. TauD is involved in assimilation of sulfur from taurine in several aerobic bacteria [100]. Furthermore, we found many of these genes (*tauABC*, *tpa*, *tauD*, *xsc*, *aprAB*, *qmoABC*) encoded in the previously mentioned Hexadella γ-symbiont MAGs, which belong to the same proposed genus as '*Ca*. Taurinisymbion ianthellae' (Table S2).

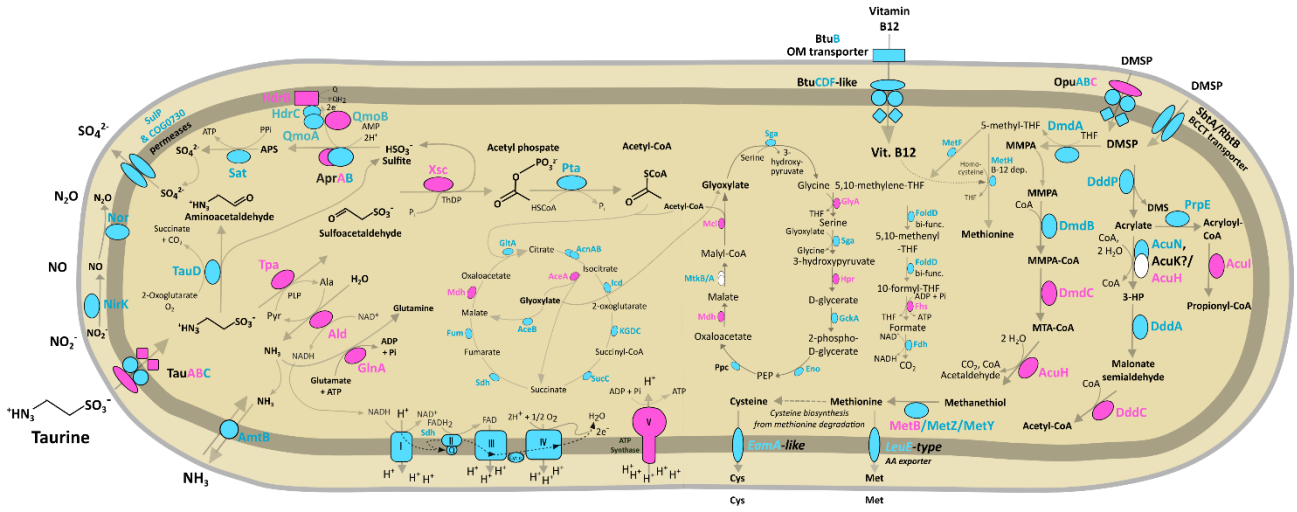


Figure 3. Taurine and DMSP metabolic pathways identified and expressed in '*Ca.* Taurinisymbion ianthellae'. Central carbon and nitrogen metabolism in '*Ca.* Taurinisymbion ianthellae' based on metaproteogenomic analyses with a focus on taurine and DMSP metabolism pathways along with the relevant import and export of exchanged metabolites. Depicted are proteins necessary for taurine import, followed by taurine dissimilation, energy conservation via sulfite oxidation and the electron transport chain, along with carbon incorporation via the glyoxylate bypass. Similarly depicted are the proteins necessary for DMSP import as well as the corresponding DMSP cleavage and demethylation pathways, including the oxidation of 5-methyl-THF and its role as a carbon donor for the conversion of glycine to serine. Enzymes detected via metaproteomics are highlighted in pink, while enzymes only identified in the '*Ca.* Taurinisymbion ianthellae' MAG are highlighted in turquoise. Missing pathway genes are depicted as white shapes. AMP, adenosine monophosphate; APS, adenosine-5'-phosphosulfate, Cys, cysteine; DMSP, dimethylsulfoniopropionate; DMS, dimethylsulfide; 3-HP, 3-hydroxypropanoate; Met, methionine; MMPA, methylmercaptopropionate; MTA-CoA, methylthioacryloyl-Coenzyme A; PEP, phosphoenolpyruvate; PLP, pyridoxal 5'-phosphate; Q, quinone; ThDP, thiamine diphosphate; THF, tetrahydrofolate. Additional information for genes and pathways depicted in the figure can be found in Supplementary Tables S2–S5.

Several of the genes involved in taurine degradation were also found to be expressed after being mapped to the '*Ca*. Taurinisymbion ianthellae' MAG. Specifically, we detected the expression of the taurine transporter TauABC (with TauAB detected only when the 454 metagenome dataset was included as a database for initial metaproteomic analyses), as well as the two enzymes, taurine-pyruvate aminotransferase (Tpa) and sulfoacetaldehyde acetyltransferase (Xsc), involved in the sequential dissimilation of taurine to sulfite via sulfoacetaldehyde. Additionally, 3 out of 7 proteins necessary for the subsequent oxidation of sulfite to sulfate (AprA, QmoB and HdrB; while Sat was not detected in the metaproteome data set) were expressed. A phosphate acetyltransferase is present in the '*Ca*. Taurinisymbion ianthellae' MAG, which would convert the acetyl phosphate produced from the activity of Xsc (also producing sulfite) into acetyl-CoA, for entry into the TCA and the glyoxylate bypass cycle. We detected the expression of the key enzymes malate dehydrogenase (Mdh) and isocitrate lyase (AceA), from each cycle, respectively, suggesting that the gyoxylate bypass is used for anabolic purposes, while sulfite oxidation is used for energy conservation by '*Ca*. Taurinisymbion ianthellae'. In addition to the expressed proteins detected for the processing of taurine-derived sulfur and carbon compounds, we detected the expression of alanine dehydrogenase (Ald), which would replenish the pyruvate needed by the Tpa enzyme for taurine dissimilation (while also providing $NADH/H^+$ for respiration) and provide NH_3 for assimilation into glutamine, by a highly expressed glutamine synthetase (GlnA; 1.30% NSAF). Ammonia can also be transported out of the cell via an encoded ammonium transporter (*amtB* ; Fig. 3, Table S2), where it can be readily oxidized to nitrite, as previously demonstrated [10], by the dominant thaumarchaeal symbiont. Excluding the expression of glutamine synthetase, the other detected 6 proteins involved in the catabolism of taurine (Tpa, Xsc, AprA, QmoB, and HdrB) amounted to 0.97% NSAF of the 149 proteins assigned to the '*Ca.* Taurinisymbion ianthellae' MAG. Nevertheless, the detection of the high affinity GlnA suggests that '*Ca.* Taurinisymbion ianthellae' was nitrogen limited at the time that the *I. basta* holobiont individual was sampled *in situ*. It has been previously observed that the *I. basta* holobiont, along with its thaumarchaeal ammonia-oxidizing symbiont, '*Ca.* Nitrosospongia ianthellae', is nitrogen limited at ambient conditions[10], and this is in line with apparent nitrogen limitation in marine systems [101] and mammalian large intestines [102]. The observations that the *I. basta* holobiont readily dissimilated added taurine in *I. basta* explant incubations (see below) and incorporated the taurine-derived carbon and nitrogen at C:N ratios indicative of nitrogen limitation (see Supplementary Information), further corroborate these findings.

While the expression of the taurine transporter has been observed in the γ-symbiont of *Olavius algarvensis* [103], the recent observations that marine sponge symbionts have the capacity for taurine utilization, is primarily based on gene presence [25, 26, 35–37], with the exception of metaproteomic evidence for *Ca*. Entotheonella phylotypes residing in the marine sponge *Theonella swinhoei* Y. However, this symbiont expressed multiple encoded copies of TauD-related proteins postulated to possibly function in the breakdown of small aromatic compounds [104]. A recent study by our group performed deep sequencing runs on more *I. basta* individuals — while also conducting metaproteomics — and was able to assemble additional low abundant microbial symbiont MAGs that were previously overlooked [105]. One of these low abundant symbionts is an Alphaprotebacterium (Alphaproteobacteria, g_UBA2767) that also possesses the capability to fully convert taurine into ammonium and sulfate via the same pathways as '*Ca*. Taurinisymbion ianthellae'. However, this low abundance Alphaproteobacterium only represents 1–5% of the microbial community and did not express any of the proteins involved in taurine import, taurine dissimilation or sulfite oxidation (it only expressed 0.4% of all detected proteins in the metaproteomic dataset). Moreover, this set of metaproteomic analyses was able to detect the TauABC and AprA proteins and definitively assign them to the '*Ca*. Taurinisymbion ianthellae' MAG.

Oxidation of sulfite to sulfate by taurine-degrading bacteria occurs via different pathways. Frequently periplasmic and cytoplasmic sulfite dehydrogenases are used [106, 107], but these enzymes were not detected in the '*Ca*. Taurinisymbion ianthellae' MAG. The genetic capability to carry out the two-step AMP-dependent indirect oxidation of sulfite to sulfate via APS in the cytoplasm employed by this symbiont, is infrequently detected in taurine degraders, with examples found among the ubiquitous, pelagic SAR11 ([58]; but most SAR11 lack ATP sulfurylase), the γ-3 symbiont in the gutless oligochaete, *Olavius algarvensis* [108], and two marine sponge symbionts [25, 26]. Although a metaproteomic study in Antarctic waters has previously revealed the expression of SAR11-associated peptides for this particular pathway of taurine dissimilation (Tpa, Xsc) followed by sulfite oxidation (AprAB) [58], SAR11 in culture has not yet been shown to excrete sulfate in the presence of taurine [109, 110].

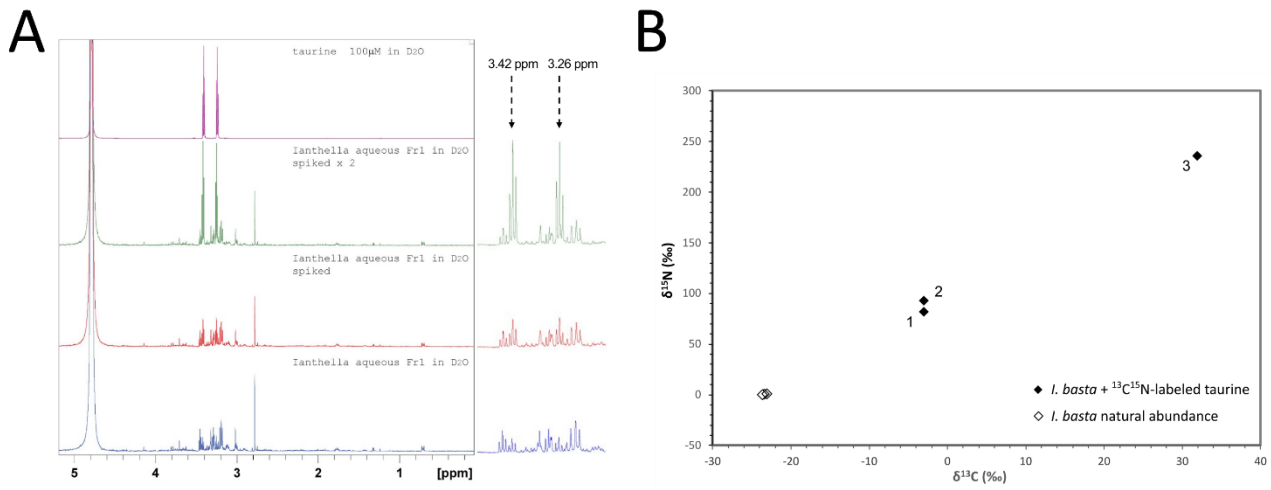


Figure 4. (A) Detection of taurine in the aqueous fraction Fraction 1 of the 90% methanol extract of *Ianthella basta* via ^{1}H NMR. ^{1}H NMR spectra of Fraction 1 from C18 of crude methanolic extract (blue), Fraction 1 spiked with 32 µl (red) and 128 µl (green) of 100 µM taurine, as well as 100 µM taurine (purple), all acquired in D_2O. Zoomed regions are depicted of diagnostic triplet signals at 3.26 and 3.42 ppm, and the increase in taurine signal intensity with spiking. (B) Incorporation of ^{13}C- and ^{15}N-labeled taurine into sponge holobiont biomass at the total nucleic acid level as determined by EA-IRMS. Data from three sponge explants incubated with labeled taurine (numbered for reference) and three control sponge explants incubated with unlabeled taurine are depicted.

Taurine metabolism in the Ianthella basta *holobiont*

Based on the findings of the metaproteogenomic data set, we investigated taurine metabolism in the *I. basta* holobiont experimentally. Firstly, NMR and LC-MS analyses of the 90% MeOH extract and the aqueous Fraction 1 confirmed that taurine was present in this sponge (Fig. 4a, Fig. S3)[111]. The natural concentration of taurine, determined by qNMR, was 5.48 ± 0.69 µmol g^{-1} sponge wet wt, which is in the mid-range of previously reported values for six other marine sponges (1.05–12.5 µmol g^{-1} sponge wet wt sponge)[112]. Congruently, we were able to detect ^{13}C and ^{15}N enrichment in whole sponge and total nucleic acids elemental analyzer isotope ratio mass spectrometry (EA-IRMS) analyses of *I. basta* sponge explants that underwent additions of an equimolar mixture of ^{13}C- and ^{15}N-labeled taurine in 48 h incubations (Fig. 4b, Fig. S4). While it cannot be excluded that sorption of labeled taurine contributed to the isotopic enrichment in the whole sponge experiments, the detection of labeled total nucleic acids clearly indicated assimilation and incorporation of labelled taurine into *I. basta* holobiont biomass. Thus, from the whole sponge EA-IRMS data, we were able to estimate uptake rates of taurine-derived C and N assimilated into whole *I. basta* holobiont biomass and obtained values of 1.38 ± 0.3 nmol C min^{-1} g^{-1} sponge wet wt (± SE)

and 0.74 ± 0.2 nmol N min^{-1} g^{-1} sponge wet wt (± SE). These *I. basta* holobiont taurine-derived carbon uptake rates correspond to approximately 4–10% of the bulk DOC uptake rates reported for the more diverse microbiome of the marine sponge *Aplysina aerophoba* [17], thereby highlighting the importance of taurine in the *I. basta* holobiont. Moreover, NanoSIMS analyses conducted on dissociated sponge extracellular material revealed accumulation of taurine-derived C and N within regions (Fig. S5 & S6) that were similar in size and shape to typical sponge microbial symbiont cells previously visualized in *I. basta* by transmission electron microscopy [55, 113].

Taurine catabolism by* Ianthella basta *holobiont results in the accumulation of inorganic S and N species

Some microbial pure cultures catabolizing taurine, oxidize the dissimilated sulfonate moiety to sulfate, which is subsequently excreted [114–116]. Considering that metaproteogenomic analyses predicted sulfate as a secreted end product of '*Ca*. Taurinisymbion ianthellae' taurine metabolism, we analyzed whether the *I. basta* holobiont produced sulfate (SO_4^{2-}) in experiments with and without added taurine. Specifically, we conducted 48 h incubations with *I. basta* explants amended with taurine (in total 1.6 mM) and without taurine in sulfate-free artificial seawater (SFASW). SFASW was used in these incubations to facilitate the detection of sulfate formation by the holobiont. Pairwise comparisons in incubations with sponges in SFASW without the addition of taurine (SFASW + *I. basta*), resulted in a small, but significant production of SO_4^{2-} after 12 h, indicating *in vivo* production of SO_4^{2-} ($p < 0.05$, one-tailed t-test) (Fig. 5a). In incubations with *I. basta* explants and added taurine (SFASW + *I. basta* + taurine), we observed a similarly small increase at 12 h ($p < 0.05$, one-tailed t-test), with SO_4^{2-} production strongly increasing between the 12 h and 48 h time points, suggesting a metabolic lag effect (Fig. 5a). Moreover, the final SO_4^{2-} concentrations of 590 ± 89 (SE) µM SO_4^{2-} (59.0 ± 2.0 µmol g^{-1} sponge wet wt in Fig. 5a) at 48 h were significantly higher than the SO_4^{2-} concentrations at t = 0 and t = 12 h (one-way ANOVA followed by Holm-Sidak tests, $p < 0.001$) and also significantly higher than all control incubations, including the incubations with sponge explants but without taurine at t = 48 h (one-way ANOVA followed by Holm-Sidak pairwise multiple comparison tests, $p < 0.001$). Two additional incubation experiments, with refined temporal resolution (t = 0, 12, 24, 36 and 48 h) using 1 mM and 100 µM unlabeled taurine amendments, respectively, resembled the results of the previous experiment, and displayed a significant SO_4^{2-} increase after 48 h for both taurine treatments (one-

way ANOVA followed by Tukey pairwise multiple comparison tests, $p < 0.05$), with essentially a full recovery of the +100 µM taurine sulfonate in the form of sulfate$^-$ after 48h (Fig. S7).

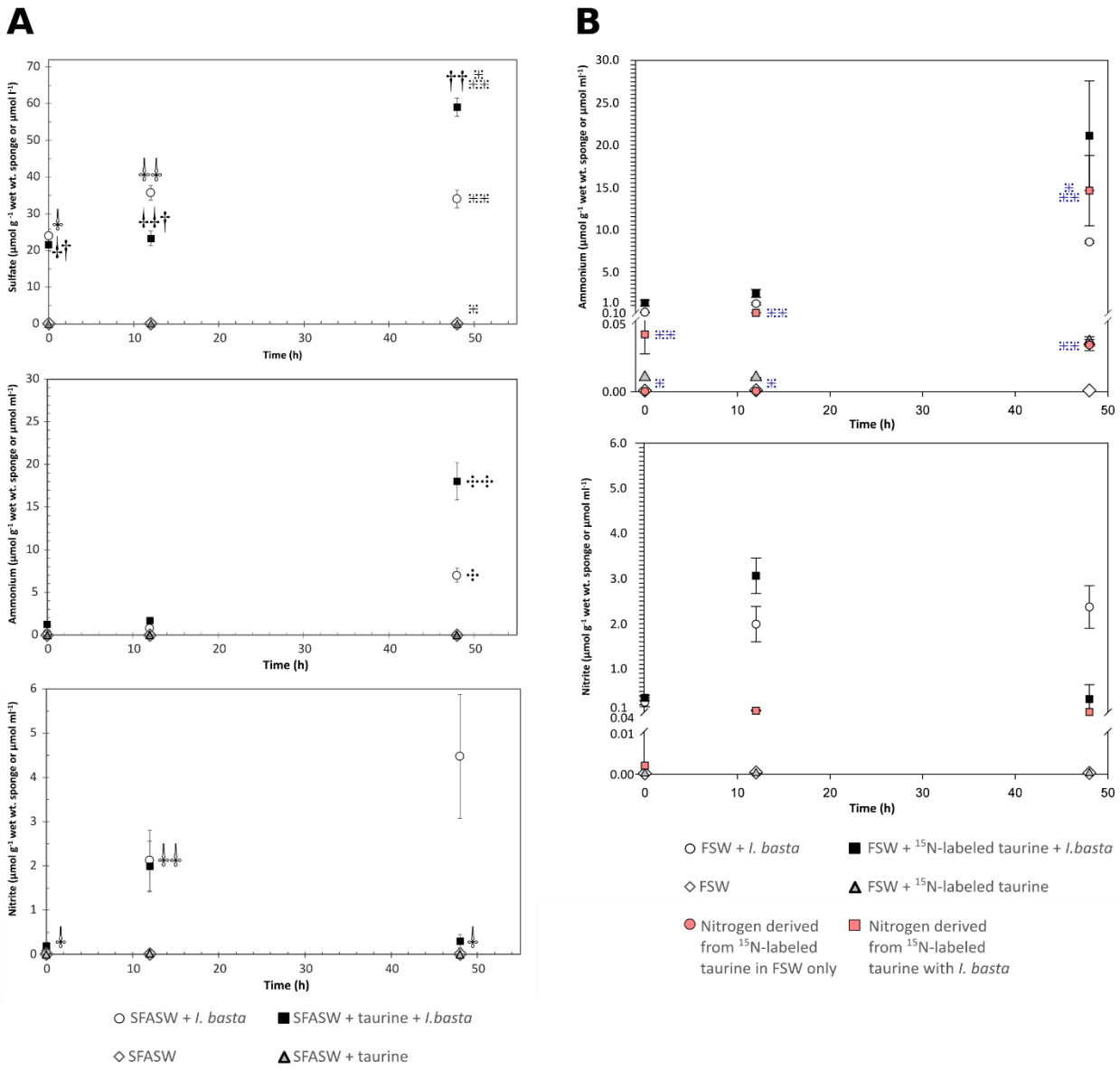


Figure 5. Time series of the concentrations of dissolved sulfate, ammonium, and nitrite in the incubation media of *I. basta* holobiont batch incubations (**A**) 48 h incubation in sulfate-free artificial seawater (SFASW) performed either without (white circle) or with added unlabeled taurine (filled square: t = 0, +1 mM taurine; t = 36 h, +0.6 mM taurine). Concentrations are normalized to the wet weight of the respective *I. basta* explants applied in the incubations (8.83 ± 2.38 SD g sponge wet wt.; range: 6.5–13.4 g sponge wet wt.). The data from the corresponding control experiments with SFASW without *I. basta* holobiont are displayed as white diamonds (without added taurine) and grey triangles (with added taurine) and plotted on the same axes, given in µmol ml^{-1}. A starting average concentration of 0.2 mM sulfate was measured in all incubation treatments containing *I. basta* explants most likely representing residual carryover from the sponge material. Incubations were performed in biological triplicates and water samples from each incubation were taken at each time point. In A), ⸸ symbols for sulfate concentrations denote significant difference in pairwise comparisons of incubations from t = 0 (⸸) to t = 12 (⸸⸸) for both the incubations with *I. basta* and taurine and without added taurine (one-tailed t-test, $p < 0.05$); †† symbols in the sulfate concentration panel reflect significant differences for sulfate concentrations for *I. basta* and taurine incubations at t = 48 (††) when compared to the earlier time points (†) (one-way ANOVA followed by Holm-Sidak tests, $p < 0.001$);

similarly, the number of ⁜ symbols reflects significant differences for sulfate concentrations for incubations with *I. basta* and taurine when compared to all other treatments at t = 48 h (one-way ANOVA followed by Holm-Sidak tests, $p < 0.001$). ⁘⁘ in the ammonium concentration panel reflects a significant difference between incubations with *I. basta* and taurine and only *I. basta* (⁘) (two-tailed t-test, $p < 0.05$). ⸸⸸ in the nitrite concentration panel reflects a significant difference between time points for the incubations with *I. basta* and taurine (one-way ANOVA followed by Tukey pairwise multiple comparison tests, $p < 0.05$). (**B**) Ammonium and nitrite accumulation in *I. basta* holobiont batch incubations (48 h) in natural filtered seawater (FSW) performed either with added taurine (filled symbols: t = 0, +1 mM taurine; t = 36 h, +0.6 mM taurine) or without added taurine (9.45 ± 2.53 SD g sponge wet wt; range: 5.5–12.7 g sponge wet wt). Taurine was added as an equimolar mixture of ^{13}C- and ^{15}N-labeled taurine to obtain sponge material for subsequent EA-IRMS and NanoSIMS analyses (by sacrificially sampling sponges at the end of the experiment) as well as to measure ^{15}N in NH_4^+ and NO_2^-. Incubations were performed in biological triplicates and water samples from each incubation were taken at each time point. The data from the corresponding control experiments with FSW without *I. basta* holobiont are displayed as white circles and diamonds as in A). Pink symbols represent the measured concentrations of ^{15}N in ammonium and nitrite that were a product of the added ^{15}N-labeled taurine (t = 0, +0.5 mM ^{15}N-labeled taurine; t = 36 h, +0.3 mM ^{15}N-labeled taurine). The number of ⁜ symbols in the ammonium concentration panel in B) reflects significant differences in $^{15}NH_4^+$ concentrations measured in incubations with *I. basta* and isotopically labeled taurine and in incubations without *I. basta* between different time points and across treatments (one-way ANOVA followed by Student-Newman-Keuls pairwise multiple comparison tests, $p < 0.05$). Error bars refer to +/- 1SE.

Furthermore, we observed NH_4^+ and NO_2^- production in *I. basta* incubations, with and without added unlabeled taurine (SFASW + *I. basta* + taurine; SFASW + *I. basta*), where NH_4^+ production at the end of the experiment with added taurine, drastically exceeded those of *I. basta* incubations without added taurine, consistent with the production of ammonium from taurine (Fig. 5a; $p < 0.05$, two-tailed t-test). Furthermore, a clear increase in NO_2^- production occurred within the first 12 h and was found to be significant for incubations with added taurine (Fig. 5a, SFASW + *I. basta* + taurine: one-way ANOVA followed by Tukey pairwise multiple comparison tests, $p < 0.05$). NH_4^+ and NO_2^- production in *I. basta* has previously been observed [10] and the NO_2^- production was attributed to the only ammonia-oxidizing symbiont, '*Ca*. Nitrosospongia ianthellae'. At the end of the experiment, NO_2^- concentrations decreased in the incubations with added taurine, consistent with the previously demonstrated inhibition of ammonia oxidation at concentrations ≥100 µM NH_4^+ in the *I.* basta holobiont and/or with the concomitant increased hypoxia facilitating denitrification activities [10], a capability that is encoded in '*Ca*. Taurinisymbion ianthellae'. Additional experiments with refined temporal resolution, displayed the same general trends (Fig. S7).

In addition, FSW incubations of the *I. basta* holobiont with ^{15}N-labeled taurine were performed. In these experiments, we were able to detect $^{15}NH_4^+$ and $^{15}NO_2^-$ formation by *I. basta* sponge explants (Fig. 5b), demonstrating taurine metabolism and oxidation of taurine-derived ammonia to nitrite by the sponge holobiont. Specifically, this indicates that the dominant thaumarchaeal symbiont hosted by *I. basta*, '*Ca*. Nitrosospongia ianthellae', which was previously shown to be the only resident ammonia-oxidizer in this

holobiont [10], oxidized the taurine-derived ammonia to nitrite. Importantly, the amount of $^{15}NH_4^+$ and $^{15}NO_2^-$ derived from ^{15}N-labeled taurine in incubations with *I. basta* was always higher than the amount detected in FSW control incubations without *I. basta* at t = 12 and t = 48 h (Figure 5b; one-way ANOVA followed by Student-Newman-Keuls pairwise multiple comparison tests, $p < 0.05$; $^{15}NO_2^-$ in FSW control incubations were undetectable in all samples). This demonstrates that these successive conversions were catalyzed by holobiont members and that this process starts as early as 12 h, suggesting that the similarly observed increase in SO_4^{2-} was also derived from taurine. In *I. basta* explant incubations without externally added taurine (FSW + *I. basta*), NH_4^+ and NO_2^- concentrations still greatly exceeded those incubations without sponge explants and added taurine (FSW + taurine), providing additional support for the production of these compounds by *I. basta* holobiont members. Finally, HPLC analysis of taurine concentrations for these FSW incubations confirmed uptake of taurine by the *I. basta* holobiont, since incubations with *I. basta* explants and added taurine, exhibited concentrations (FSW + *I. basta* + taurine: 1.20 ± 0.05) significantly lower than in control incubations without *I. basta* explants and added taurine (FSW + taurine: 1.73 ± 0.02 mM taurine; Kruskal-Wallis, $p < 0.05$; Fig. S8).

These incubation experiments demonstrate that the *I. basta* holobiont explants actively processed and incorporated taurine-derived carbon and nitrogen, while then dissimilating the resultant ammonia moiety to nitrite and oxidizing the sulfonate moiety to sulfate, from the onset through the course of the incubations. Taken together with the metaproteogenomic dataset, we conclude that '*Ca*. Taurinisymbion ianthellae' is the holobiont member most likely responsible for this taurine incorporation and dissimilation, including the oxidation of the resultant sulfite to sulfate. Although the observed rates are potential net rates under specific incubation conditions, which will undoubtedly differ from *in situ* rates, as in all *ex vivo* experiments, the prompt incorporation and dissimilation of added exogenous taurine, suggests a N-limited sponge holobiont that is metabolically ready to access and transform this compound for anabolic as well as catabolic purposes. Recently, several laboratory and field-based studies have documented the exchange of sulfonates between marine bacteria and eukaryotic phytoplankton, with taurine (or its derivatives), always being one of the main sulfonates involved [32–34]. In co-culture, the coastal diatom *Pseudo-nitzschia multiseries* apparently provides a *Sulfitobacter* strain (SA11) with taurine in exchange for the plant hormone indole-3-acetic acid as well as ammonium [38]. Similarly, cross-feeding between nitrite oxidizers and ammonia oxidizers has been

demonstrated wherein urea or cyanate is first processed by nitrite oxidizers [117, 118], thereby providing the cleaved ammonia for oxidation to nitrite. While marine Thaumarchaeota have been shown to directly utilize urea and cyanate [119] as well as taurine exuded by mesozooplankton [120], the mechanisms for the direct utilization of these dissolved organic compounds has yet to be elucidated and indirect utilization remains a confounding factor.

A proposed model of an auxotrophy network between symbionts and the sponge host

In the *I. basta* metagenome we found no indications for taurine biosynthesis by its microbial symbionts. Consequently, we believe it to be most likely that taurine detected in the *I. basta* holobiont is synthesized by the sponge and after release, via export or sponge cell lysis, is utilized by '*Ca*. Taurinisymbion ianthellae'. In fact, the 5.5 µmol g^{-1} sponge wet weight taurine we measured in *I. basta* would correspond to ~6.6 mM concentrations within the sponge explants used for the incubations, (under the assumption that 1.2 g sponge wet wt = 1 cm^{-3}; [96]) which is on the high end of measured cytosolic concentrations in eukaryotic phytoplankton [34], and this still does not take into account further potential concentration in specialized cells, cell compartments, or in the mesohyl. We thus propose that endogenous taurine is either translocated by *I. basta* to '*Ca*. Taurinisymbion ianthellae', as has been shown recently for DOM-derived carbon and nitrogen in a high microbial abundance (HMA) sponge [15], or is derived from host cell lysis, as a result of the rapid cell turnover characteristic of marine sponges [121, 122]. Nevertheless, it is possible that marine sponges themselves process taurine from the environment, as has been shown for bulk DOM mixtures [15, 17], instead of synthesizing it, and there remains the possibility that in our incubation experiments *I. basta* itself took up taurine and processed it. Even though our incubation experiments involved the addition of exogenous taurine, the almost immediate and continued production of the dissimilated and oxidized end products, ammonia and sulfate, two processes for which, so far, only microbially mediated pathways have been demonstrated, suggests that the added taurine was processed primarily by the dominant gammaproteobacterial symbiont, '*Ca*. Taurinisymbion ianthellae', and that it is metabolically 'primed' to process this compound. Alongside the incubation experimental results and the measured taurine concentrations in *I. basta*, we therefore assume endogenous taurine supplementation by *I. basta* to be the most parsimonious route given the low taurine seawater concentrations [34, 120], competition with other

microbes in seawater where taurine turnover rates are high [120], the ability of the sponge to be able to gain the taurine precursors cysteine and methionine from multiple sources (exogenous DOM and POM, as well as from its microbial symbionts), and the expression of the taurine degradation pathway by '*Ca*. Taurinisymbion ianthellae' *in vivo*.

Taurine biosynthesis is typically considered to be a metazoan feature but has not yet been demonstrated experimentally in sponges nor has this capability been suggested in sequenced sponges. More recent discoveries have demonstrated bacterial and microalgal capabilities in synthesizing taurine [123, 124], and yet taurine biosynthesis in microalgae and bacteria is not universal and quite sporadic in distribution [34, 125]. Taurine biosynthesis pathways in mammalian and invertebrate tissues use cysteine as a precursor [45] and references therein] and cysteine synthesis in metazoans requires methionine [126]. The methionine synthesis pathway, along with the biosynthesis pathways for several other essential amino acids, are generally lacking in metazoans [127–129], and thus these compounds are typically acquired via heterotrophy. While at least some sponges seem to possess the genomic repertoire for cysteine and methionine synthesis [130, 131], so far, no studies have confirmed their biosynthesis in radiolabeled amino acid precursor or feeding studies, whereas scleratinian corals can produce low amounts of methionine and perhaps even cysteine [131, 132]. '*Ca*. Taurinisymbion ianthellae' encodes pathways for the synthesis of methionine and cysteine (Fig. S9), which are connected with the degradation of taurine and DMSP (Fig. 3), and also encodes several copies of EamA domain-containing proteins (IPR000620) and two copies of LeuE-type export proteins (IPR001123) (Fig. 3 & Table S5), with both families containing homologous representatives demonstrated to export cysteine [133] and methionine [134], respectively. Although further studies are needed to confirm whether marine sponges can indeed synthesize cysteine and methionine (as well as taurine), it is tempting to speculate that '*Ca*. Taurinisymbion ianthellae' synthesizes and provides methionine and cysteine, either via active transport or symbiont cell lysis, to *I. basta*, and that the latter is used as a precursor for taurine biosynthesis. Cysteine or methionine auxotrophy in *I. basta* is not a precondition for this metabolic exchange to occur, as highly regulated kinetic switches may exist within the *I. basta* holobiont that is dependent on dynamic shifts in resource limitation and environmental stressors.

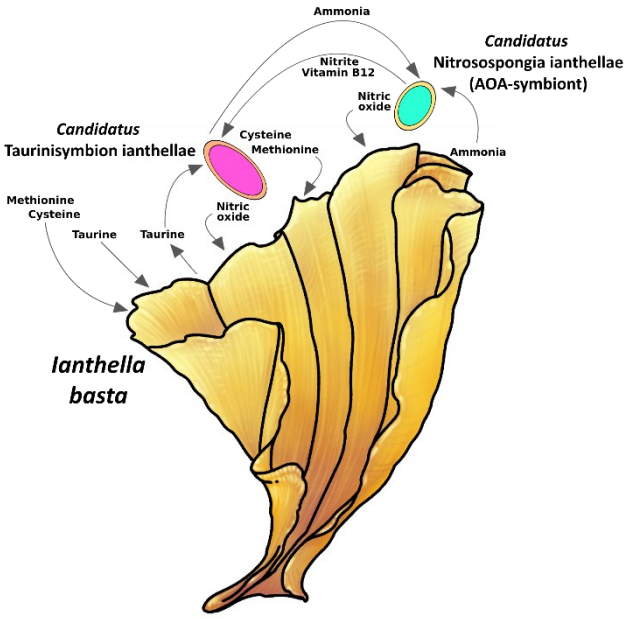


Figure 6. Conceptual scheme of tripartite metabolic interactions within the *Ianthella basta* holobiont.

Although the methionine synthase (MetH)-containing pathway for methionine biosynthesis in '*Ca*. Taurinisymbion ianthellae' is vitamin B_{12} dependent (this organism also encodes a vitamin B_{12}-independent pathway using DMSP as precursor as well as the cobalamin-independent methionine synthase, MetE; Fig 3, Fig S9), its MAG does not contain the genes required for synthesis of this vitamin, while it does encode a putative vitamin B_{12} transporter BtuCDF (Fig. 3 & Table S5). As the thaumarchaeal symbiont of *I. basta* '*Ca.* Nitrosospongia ianthellae' has the genomic repertoire for cobalamin (vitamin B_{12}) biosynthesis [10] it is conceivable that it provides vitamin B_{12} for methionine biosynthesis in '*Ca*. Taurinisymbion ianthellae' if DMSP is not available, while receiving ammonia produced by the γ-symbiont from taurine degradation. Nitrite formed by ammonia oxidation by '*Ca.* Nitrosospongia ianthellae' can then serve under low oxygen concentrations as electron acceptor in '*Ca*. Taurinisymbion ianthellae'. Furthermore, since nitric oxide has been shown to have an immediate contractile effect on sponges [135] and seems to be an ancient and key regulator in the marine sponge life cycle [136], the fact that '*Ca*. Taurinisymbion ianthellae' encodes two denitrification genes — a nitrite reductase, *nirK* and a single subunit quinol-dependent nitric oxide reductase, *qNOR* [137] — that produces and consumes NO, suggests that this symbiont may be able to modulate host behavior. The frequent presence of an incomplete denitrification pathway along with the quinol-dependent *qNOR* — a nitrite reductase commonly found in pathogenic species [138] — in other sponge metagenomes [28], further points at the possibility that this is a widespread feature of microbial symbionts residing in

marine sponges. Taken together, our data indicate a tightly connected metabolic network in the *I. basta* holobiont (Fig. 6), albeit in contrast to selected other sponges, no data are yet available for *I. basta* on the uptake of amino acids directly from the environment [16, 17] and on the acquisition of amino acids from ingested free-living and symbiotic prokaryotes [139].

Conclusions

In this study, we showed through a combination of metaproteogenomic analyses and incubation experiments combined with isotope labeling, that the dominant *I. basta* gammaproteobacterial symbiont, '*Ca*. Taurinisymbion ianthellae', most likely uses endogenously occurring taurine, that is potentially produced by the sponge holobiont, for assimilation and energy conservation. We experimentally demonstrated that the taurine dissimilation products, sulfate and ammonia, are secreted and that the latter compound is used for ammonia oxidation by the dominant thaumarchaeal symbiont — '*Ca.* Nitrosospongia ianthellae'. Taurine seems to be an important metabolite, not only in *I. basta*, but in many sponge holobiont systems, as indicated by the seeming ubiquity of taurine and taurine dissimilating genes within sponge microbiomes [25, 26, 35–37], and the appearance of a specific taurine-dissimilation/sulfite-oxidation pathway restricted mostly to symbiotic microorganisms [25, 26, 108]. Taurine also appears to be a central metabolite in several other metazoan holobiont systems such as in cnidarians, various bivalve species [44, 45, 140, 141], and in mammalian gut systems [39, 40] . Future research is needed to further disentangle the mechanisms of taurine synthesis via symbiont-derived or environmentally supplied precursors in marine sponge holobionts, to better understand the role of exogenous and endogenous taurine uptake, and the numerous roles taurine may play for the functioning of these symbioses.

Acknowledgements

We thank Emmanuelle Botte for assisting with experimental incubations, Heidi Luter for assisting with field collections and Bettina Glasl and Pam Engelberts for helpful discussions. Florian Moeller and Maria Mooshammer were supported by the ERC Advanced Grant Nitricare (294343 to MW). We thank Bernhard Schink for valuable advice regarding the provisional naming of the gammaproteobacterial symbiont.

Compliance with ethical standards

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Bell JJ. The functional roles of marine sponges. *Estuarine, Coastal and Shelf Science* 2008; **79**: 341–353.
2. de Goeij J, van den Berg H, van Oostveen M, Epping E, van Duyl F. Major bulk dissolved organic carbon (DOC) removal by encrusting coral reef cavity sponges. *Mar Ecol Prog Ser* 2008; **357**: 139–151.
3. de Goeij JM, van Oevelen D, Vermeij MJA, Osinga R, Middelburg JJ, de Goeij AFPM, et al. Surviving in a Marine Desert: The Sponge Loop Retains Resources Within Coral Reefs. *Science* 2013; **342**: 108–110.
4. Rix L, de Goeij JM, Mueller CE, Struck U, Middelburg JJ, van Duyl FC, et al. Coral mucus fuels the sponge loop in warm- and cold-water coral reef ecosystems. *Scientific Reports* 2016; **6**: 18715.
5. Webster NS, Thomas T. The Sponge Hologenome. *mBio* 2016; **7**.
6. Zhang F, Jonas L, Lin H, Hill RT. Microbially mediated nutrient cycles in marine sponges. *FEMS Microbiol Ecol* 2019; **95**.
7. Wilkinson CR. Net Primary Productivity in Coral Reef Sponges. *Science* 1983; **219**: 410–412.
8. Wilkinson CR, Fay P. Nitrogen fixation in coral reef sponges with symbiotic cyanobacteria. *Nature* 1979; **279**: 527–529.
9. Bayer K, Schmitt S, Hentschel U. Physiology, phylogeny and *in situ* evidence for bacterial and archaeal nitrifiers in the marine sponge *Aplysina aerophoba*. *Environmental Microbiology* 2008; **10**: 2942–2955.
10. Moeller FU, Webster NS, Herbold CW, Behnam F, Domman D, Albertsen M, et al. Characterization of a thaumarchaeal symbiont that drives incomplete nitrification in the tropical sponge Ianthella basta. *Environmental Microbiology* 2019; **21**: 3831–3854.
11. Hoffmann F, Larsen O, Thiel V, Rapp HT, Pape T, Michaelis W, et al. An Anaerobic World in Sponges. *Geomicrobiology Journal* 2005; **22**: 1–10.
12. Hoffmann F, Radax R, Woebken D, Holtappels M, Lavik G, Rapp HT, et al. Complex nitrogen cycling in the sponge *Geodia barretti*. *Environmental Microbiology* 2009; **11**: 2228–2243.
13. Freeman CJ, Thacker RW. Complex interactions between marine sponges and their symbiotic microbial communities. *Limnology and Oceanography* 2011; **56**: 1577–1586.
14. Freeman CJ, Thacker RW, Baker DM, Fogel ML. Quality or quantity: is nutrient transfer driven more by symbiont identity and productivity than by symbiont abundance? *The ISME Journal* 2013; **7**: 1116–1125.
15. Hudspith M, Rix L, Achlatis M, Bougoure J, Guagliardo P, Clode PL, et al. Subcellular view of host–microbiome nutrient exchange in sponges: insights into the ecological success of an early metazoan–microbe symbiosis. *Microbiome* 2021; **9**: 44.
16. Achlatis M, Pernice M, Green K, de Goeij JM, Guagliardo P, Kilburn MR, et al. Single-cell visualization indicates direct role of sponge host in uptake of dissolved organic matter. *Proceedings of the Royal Society B: Biological Sciences* 2019; **286**: 20192153.
17. Rix L, Ribes M, Coma R, Jahn MT, de Goeij JM, van Oevelen D, et al. Heterotrophy in the earliest gut: a single-cell view of heterotrophic carbon and nitrogen assimilation in sponge-microbe symbioses. *ISME J* 2020; **14**: 2554–2567.
18. Taylor MW, Radax R, Steger D, Wagner M. Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol Mol Biol Rev* 2007; **71**: 295–347.
19. Webster NS, Taylor MW. Marine sponges and their microbial symbionts: love and other relationships: Marine sponge microbiology. *Environmental Microbiology* 2012; **14**: 335–346.

20. Thomas T, Moitinho-Silva L, Lurgi M, Björk JR, Easson C, Astudillo-García C, et al. Diversity, structure and convergent evolution of the global sponge microbiome. *Nature Communications* 2016; **7**.

21. Siegl A, Kamke J, Hochmuth T, Piel J, Richter M, Liang C, et al. Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges. *The ISME Journal* 2011; **5**: 61–70.

22. Moitinho-Silva L, Díez-Vives C, Batani G, Esteves AI, Jahn MT, Thomas T. Integrated metabolism in sponge–microbe symbiosis revealed by genome-centered metatranscriptomics. *The ISME Journal* 2017; **11**: 1651–1666.

23. Fiore CL, Labrie M, Jarett JK, Lesser MP. Transcriptional activity of the giant barrel sponge, Xestospongia muta Holobiont: molecular evidence for metabolic interchange. *Frontiers in Microbiology* 2015; **6**.

24. Moreno-Pino M, Cristi A, Gillooly JF, Trefault N. Characterizing the microbiomes of Antarctic sponges: a functional metagenomic approach. *Scientific Reports* 2020; **10**: 645.

25. Tian R-M, Wang Y, Bougouffa S, Gao Z-M, Cai L, Bajic V, et al. Genomic analysis reveals versatile heterotrophic capacity of a potentially symbiotic sulfur-oxidizing bacterium in sponge: A versatile SOB as a potential sponge symbiont. *Environmental Microbiology* 2014; **16**: 3548–3561.

26. Gauthier M-EA, Watson JR, Degnan SM. Draft Genomes Shed Light on the Dual Bacterial Symbiosis that Dominates the Microbiome of the Coral Reef Sponge Amphimedon queenslandica. *Frontiers in Marine Science* 2016; **3**.

27. Hallam SJ, Konstantinidis KT, Putnam N, Schleper C, Watanabe Y -i., Sugahara J, et al. Genomic analysis of the uncultivated marine crenarchaeote Cenarchaeum symbiosum. *Proceedings of the National Academy of Sciences* 2006; **103**: 18296–18301.

28. Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, Webster NS, et al. Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proceedings of the National Academy of Sciences* 2012; **109**: E1878–E1887.

29. Hentschel U, Piel J, Degnan SM, Taylor MW. Genomic insights into the marine sponge microbiome. *Nature Reviews Microbiology* 2012; **10**: 641–654.

30. Liu M, Fan L, Zhong L, Kjelleberg S, Thomas T. Metaproteogenomic analysis of a community of sponge symbionts. *The ISME Journal* 2012; **6**: 1515–1525.

31. Hooper JNA, Capon RJ, Keenan CP, Parry DL, Smit N. Chemotaxonomy of marine sponges: Families Microcionidae, Raspailiidae and Axinellidae, and their relationships with other families in the orders Poecilosclerida and Axinellida (Porifera : Demospongiae). *Invert Systematics* 1992; **6**: 261–301.

32. Durham BP, Sharma S, Luo H, Smith CB, Amin SA, Bender SJ, et al. Cryptic carbon and sulfur cycling between surface ocean plankton. *Proc Natl Acad Sci USA* 2015; **112**: 453–457.

33. Landa M, Burns AS, Roth SJ, Moran MA. Bacterial transcriptome remodeling during sequential co-culture with a marine dinoflagellate and diatom. *ISME J* 2017; **11**: 2677–2690.

34. Durham BP, Boysen AK, Carlson LT, Groussman RD, Heal KR, Cain KR, et al. Sulfonate-based networks between eukaryotic phytoplankton and heterotrophic bacteria in the surface ocean. *Nat Microbiol* 2019; **4**: 1706–1715.

35. Karimi E, Keller-Costa T, Slaby BM, Cox CJ, da Rocha UN, Hentschel U, et al. Genomic blueprints of sponge-prokaryote symbiosis are shared by low abundant and cultivatable Alphaproteobacteria. *Scientific Reports* 2019; **9**: 1999.

36. Botté ES, Nielsen S, Abdul Wahab MA, Webster J, Robbins S, Thomas T, et al. Changes in the metabolic potential of the sponge microbiome under ocean acidification. *Nature Communications* 2019; **10**: 4134.

37. Engelberts JP, Robbins SJ, de Goeij JM, Aranda M, Bell SC, Webster NS. Characterization of a sponge microbiome using an integrative genome-centric approach. *The ISME Journal* 2020; **14**: 1100–1110.

38. Amin SA, Hmelo LR, van Tol HM, Durham BP, Carlson LT, Heal KR, et al. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* 2015; **522**: 98–101.

39. Stacy A, Andrade-Oliveira V, McCulloch JA, Hild B, Oh JH, Perez-Chaparro PJ, et al. Infection trains the host for microbiota-enhanced resistance to pathogens. *Cell* 2021; **184**: 615-627.e17.

40. Ye H, Borusak S, Eberl C, Hanson BT, Zwirzitz B, Herbold CW, et al. A novel taurine-respiring murine gut bacterium contributes to colonization resistance against enteropathogens. 2022. Microbiology.

41. Atlas M, Bahl JJ, Roeske W, Bressler R. In vitro osmoregulation of taurine in fetal mouse hearts. *Journal of Molecular and Cellular Cardiology* 1984; **16**: 311–320.

42. Schaffer S, Takahashi K, Azuma J. Role of osmoregulation in the actions of taurine. *Amino Acids* 2000; **19**: 527–546.

43. Huxtable RJ. Physiological actions of taurine. *Physiological Reviews* 1992; **72**: 101–163.

44. Lee R, Childress J, Desaulniers N. The effects of exposure to ammonia on ammonia and taurine pools of the symbiotic clam. *Journal of Experimental Biology* 1997; **200**: 2797–2805.

45. Joyner JL, Peyer SM, Lee RW. Possible Roles of Sulfur-Containing Amino Acids in a Chemoautotrophic Bacterium-Mollusc Symbiosis. *The Biological Bulletin* 2003; **205**: 331–338.

46. Alberic P, Boulegue J. Unusual amino compounds in the tissues of Calyptogena phaseoliformis (Japan Trench): Possible link to symbiosis. *Progress in Oceanography* 1990; **24**: 89–101.

47. Conway NM, Howes BL, McDowell Capuzzo JE, Turner RD, Cavanaugh CM. Characterization and site description of Solemya borealis (Bivalvia; Solemyidae), another bivalve-bacteria symbiosis. *Marine Biology* 1992; **112**: 601–613.

48. Conway NM, McDowell Capuzzo JE. High taurine levels in the Solemya velum symbiosis. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* 1992; **102**: 175–185.

49. Pranal V, Fiala-Médioni A, Colomines J. Amino acid and related compound composition in two symbiotic mytilid species from hydrothermal vents. *Mar Ecol Prog Ser* 1995; **119**: 155–166.

50. Pruski AM, Fiala-Médioni, A., Boulegue, J., Colomines, J.C. Sulphur metabolism in chemosynthetic symbiont-containing invertebrates from hydrothermal vents. *Cahiers de Biologie Marine* 1997; **2**.

51. Pallas PS. Elenchus zoophytorum sistens generum adumbrationes generaliores et specierum cognitarum succintas descriptiones, cum selectis auctorum synonymis / P. S. Pallas. 1766. Apud Petrum van Cleef, Hagae-Comitum :

52. Bergquist PR, Kelly-Borges M. Systematics and Biogeography of the Genus Ianthella (Demospongiae: Verongida: Ianthellidae) in the south-west Pacific. *Beagle: Records of the Museums and Art Galleries of the Northern Territory, The* 1995; **12**: 151.

53. Hentschel U, Usher KM, Taylor MW. Marine sponges as microbial fermenters: Marine sponges as microbial fermenters. *FEMS Microbiology Ecology* 2006; **55**: 167–177.

54. Webster NS, Taylor MW, Behnam F, Lücker S, Rattei T, Whalan S, et al. Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environmental Microbiology* 2009.

55. Luter HM, Whalan S, Webster NS. Exploring the Role of Microorganisms in the Disease-Like Syndrome Affecting the Sponge Ianthella basta. *Applied and Environmental Microbiology* 2010; **76**: 5736–5744.

56. Freckelton ML, Luter HM, Andreakis N, Webster NS, Motti CA. Qualitative variation in colour morphotypes of Ianthella basta (Porifera: Verongida). *Hydrobiologia* 2012; **687**: 191–203.

57. Luter HM, Whalan S, Webster NS. Thermal and Sedimentation Stress Are Unlikely Causes of Brown Spot Syndrome in the Coral Reef Sponge, Ianthella basta. *PLoS ONE* 2012; **7**: e39779.

58. Williams TJ, Long E, Evans F, DeMaere MZ, Lauro FM, Raftery MJ, et al. A metaproteomic assessment of winter and summer bacterioplankton from Antarctic Peninsula coastal surface waters. *The ISME Journal* 2012; **6**: 1883–1900.

59. Strehlow BW, Pineda M-C, Duckworth A, Kendrick GA, Renton M, Wahab MAA, et al. Sediment tolerance mechanisms identified in sponges using advanced imaging techniques. *PeerJ* 2017; **5**: e3904.

60. Lu X, Zou L, Clevinger C, Liu Q, Hollibaugh JT, Mou X. Temporal dynamics and depth variations of dissolved free amino acids and polyamines in coastal seawater determined by high-performance liquid chromatography. *Marine Chemistry* 2014; **163**: 36–44.

61. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* 2012; **19**: 455–477.

62. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 2015; **3**: e1165.

63. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal* 2017; **11**: 2864–2868.

64. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* 2015; **25**: 1043–1055.

65. Vallenet D, Engelen S, Mornico D, Cruveiller S, Fleury L, Lajus A, et al. MicroScope: a platform for microbial genome annotation and comparative genomics. *Database* 2009; **2009**.

66. Zhang S, Song W, Wemheuer B, Reveillaud J, Webster N, Thomas T. Comparative Genomics Reveals Ecological and Evolutionary Insights into Sponge-Associated *Thaumarchaeota*. *mSystems* 2019; **4**: e00288-19.

67. Stoecker K, Dorninger C, Daims H, Wagner M. Double Labeling of Oligonucleotide Probes for Fluorescence In Situ Hybridization (DOPE-FISH) Improves Signal Intensity and Increases rRNA Accessibility. *Applied and Environmental Microbiology* 2010; **76**: 922–926.

68. Daims H, Brühl A, Amann R, Schleifer K-H, Wagner M. The Domain-specific Probe EUB338 is Insufficient for the Detection of all Bacteria: Development and Evaluation of a more Comprehensive Probe Set. *Systematic and Applied Microbiology* 1999; **22**: 434–444.

69. Wallner G, Amann R, Beisker W. Optimizing fluorescent in situ hybridization with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms. *Cytometry* 1993; **14**: 136–143.

70. Simister RL, Deines P, Botté ES, Webster NS, Taylor MW. Sponge-specific clusters revisited: a comprehensive phylogeny of sponge-associated microorganisms: Sponge-associated microorganisms. *Environmental Microbiology* 2012; **14**: 517–524.

71. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* 2015; **32**: 268–274.

72. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 2017; **14**: 587–589.

73. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* 2018; **35**: 518–522.

74. Lagkouvardos I, Joseph D, Kapfhammer M, Giritli S, Horn M, Haller D, et al. IMNGS: A comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. *Scientific Reports* 2016; **6**.

75. Berger SA, Krompass D, Stamatakis A. Performance, Accuracy, and Web Server for Evolutionary Placement of Short Sequence Reads under Maximum Likelihood. *Systematic Biology* 2011; **60**: 291–302.

76. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014; **30**: 1312–1313.

77. Otto A, Bernhardt J, Meyer H, Schaffer M, Herbst F-A, Siebourg J, et al. Systems-wide temporal proteomic profiling in glucose-starved Bacillus subtilis. *Nature Communications* 2010; **1**: 137.

78. Verberkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M, Halfvarson J, et al. Shotgun metaproteomics of the human distal gut microbiota. *The ISME Journal* 2009; **3**: 179–189.

79. Heinz E, Williams TA, Nakjang S, Noël CJ, Swan DC, Goldberg AV, et al. The Genome of the Obligate Intracellular Parasite Trachipleistophora hominis: New Insights into Microsporidian Genome Dynamics and Reductive Evolution. *PLoS Pathogens* 2012; **8**: e1002979.

80. Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP. Evaluation of Multidimensional Chromatography Coupled with Tandem Mass Spectrometry (LC/LC−MS/MS) for Large-Scale Protein Analysis: The Yeast Proteome. *Journal of Proteome Research* 2003; **2**: 43–50.

81. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods* 2007; **4**: 207–214.

82. Florens L, Carozza M, Swanson S, Fournier M, Coleman M, Workman J, et al. Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods* 2006; **40**: 303–311.

83. Tapiolas DM, Raina J-B, Lutz A, Willis BL, Motti CA. Direct measurement of dimethylsulfoniopropionate (DMSP) in reef-building corals using quantitative nuclear magnetic resonance (qNMR) spectroscopy. *Journal of Experimental Marine Biology and Ecology* 2013; **443**: 85–89.

84. Angel R, Claus P, Conrad R. Methanogenic archaea are globally ubiquitous in aerated soils and become active under wet anoxic conditions. *ISME J* 2012; **6**: 847–862.

85. Fry B. Stable isotope ecology. 2006. Springer.

86. Sørensen P, Jensen ES. Sequential diffusion of ammonium and nitrate from soil extracts to a polytetrafluoroethylene trap for 15N determination. *Analytica Chimica Acta* 1991; **252**: 201–203.

87. Lachouani P, Frank AH, Wanek W. A suite of sensitive chemical methods to determine the δ15N of ammonium, nitrate and total dissolved N in soil extracts. *Rapid Communications in Mass Spectrometry* 2010; **24**: 3615–3623.

88. Tian R-M, Sun J, Cai L, Zhang W-P, Zhou G-W, Qiu J-W, et al. The deep-sea glass sponge *L ophophysema eversa* harbours potential symbionts responsible for the nutrient conversions of carbon, nitrogen and sulfur: Potential symbionts in a glass sponge. *Environmental Microbiology* 2016; **18**: 2481–2494.

89. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology* 2018.

90. Reveillaud J, Maignien L, Eren AM, Huber JA, Apprill A, Sogin ML, et al. Host-specificity among abundant and rare taxa in the sponge microbiome. *The ISME Journal* 2014; **8**: 1198–1209.

91. Rust M, Helfrich EJN, Freeman MF, Nanudorn P, Field CM, Rückert C, et al. A multiproducer microbiome generates chemical diversity in the marine sponge *Mycale hentscheli*. *Proc Natl Acad Sci USA* 2020; **117**: 9508–9518.

92. Storey MA, Andreassend SK, Bracegirdle J, Brown A, Keyzers RA, Ackerley DF, et al. Metagenomic Exploration of the Marine Sponge *Mycale hentscheli* Uncovers Multiple Polyketide-Producing Bacterial Symbionts. *mBio* 2020; **11**: e02997-19.

93. Glasl B, Robbins S, Frade PR, Marangon E, Laffy PW, Bourne DG, et al. Comparative genome-centric analysis reveals seasonal variation in the function of coral reef microbiomes. *ISME J* 2020; **14**: 1435–1450.

94. Luo C, Rodriguez-R LM, Konstantinidis KT. MyTaxa: an advanced taxonomic classifier for genomic and metagenomic sequences. *Nucleic Acids Research* 2014; **42**: e73–e73.

95. Konstantinidis KT, Rosselló-Móra R, Amann R. Uncultivated microbes in need of their own taxonomy. *The ISME Journal* 2017; **11**: 2399–2406.

96. Schläppy M-L, Schöttner SI, Lavik G, Kuypers MMM, de Beer D, Hoffmann F. Evidence of nitrification and denitrification in high and low microbial abundance sponges. *Marine Biology* 2010; **157**: 593–602.

97. Cook AM, Denger K. Metabolism of Taurine in Microorganisms. In: Oja SS, Saransaari P (eds). *Taurine 6*. 2006. Springer US, pp 3–13.

98. Grein F, Ramos AR, Venceslau SS, Pereira IAC. Unifying concepts in anaerobic respiration: Insights from dissimilatory sulfur metabolism. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 2013; **1827**: 145–160.

99. Rein U, Gueta R, Denger K, Ruff J, Hollemeyer K, Cook AM. Dissimilation of cysteate via 3-sulfolactate sulfo-lyase and a sulfate exporter in Paracoccus pantotrophus NKNCYSA. *Microbiology,* 2005; **151**: 737–747.

100. Kertesz MA. Riding the sulfur cycle – metabolism of sulfonates and sulfate esters in Gram-negative bacteria. *FEMS Microbiol Rev* 2000; **24**: 135–175.

101. Elser JJ, Bracken MES, Cleland EE, Gruner DS, Harpole WS, Hillebrand H, et al. Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. *Ecol Letters* 2007; **10**: 1135–1142.

102. Reese AT, Pereira FC, Schintlmeister A, Berry D, Wagner M, Hale LP, et al. Microbial nitrogen limitation in the mammalian large intestine. *Nat Microbiol* 2018; **3**: 1441–1450.

103. Kleiner M, Wentrup C, Lott C, Teeling H, Wetzel S, Young J, et al. Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *PNAS* 2012; **109**: E1173–E1182.

104. Lackner G, Peters EE, Helfrich EJN, Piel J. Insights into the lifestyle of uncultured bacterial natural product factories associated with marine sponges. *PNAS* 2017; **114**: E347–E356.

105. Engelberts JP, Robbins SJ, Herbold CW, Moeller FU, Jehmlich N, Laffy PW, et al. Metabolic reconstruction of the near complete microbiome of the model sponge *IANTHELLA BASTA* . *Environmental Microbiology* 2022; 1462-2920.16302.

106. Baldock MI, Denger K, Smits THM, Cook AM. *Roseovarius* sp. strain 217: aerobic taurine dissimilation via acetate kinase and acetate-CoA ligase. *FEMS Microbiology Letters* 2007; **271**: 202–206.

107. Lehmann S. Sulfite dehydrogenases in organotrophic bacteria: enzymes, genes and regulation. 2013. University of Konstanz.

108. Woyke T, Teeling H, Ivanova NN, Huntemann M, Richter M, Gloeckner FO, et al. Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* 2006; **443**: 950–955.

109. Schwalbach MS, Tripp HJ, Steindler L, Smith DP, Giovannoni SJ. The presence of the glycolysis operon in SAR11 genomes is positively correlated with ocean productivity. *Environmental Microbiology* 2010; **12**: 490–500.

110. Steindler L, Schwalbach MS, Smith DP, Chan F, Giovannoni SJ. Energy Starved Candidatus Pelagibacter Ubique Substitutes Light-Mediated ATP Production for Endogenous Carbon Respiration. *PLOS ONE* 2011; **6**: e19725.

111. Gamagedara S, Shi H, Ma Y. Quantitative determination of taurine and related biomarkers in urine by liquid chromatography–tandem mass spectrometry. *Anal Bioanal Chem* 2012; **402**: 763–770.

112. Shinagawa A, Suzuki T, Konosu S. The Role of Free Amino Acids and Betaines in Intracellular Osmoregulation of Marine Sponges. *Nippon Suisan Gakkaishi* 1992; **58**: 1717–1722.

113. Engelberts JP, Abdul Wahab MA, Maldonado M, Rix L, Marangon E, Robbins SJ, et al. Microbes from Mum: symbiont transmission in the tropical reef sponge Ianthella basta. *ISME COMMUN* 2022; **2**: 90.

114. Brüggemann C, Denger K, Cook AM, Ruff J. Enzymes and genes of taurine and isethionate dissimilation in Paracoccus denitrificans. *Microbiology* 2004; **150**: 805–816.

115. Weinitschke S, von Rekowski KS, Denger K, Cook AM. Sulfoacetaldehyde is excreted quantitatively by Acinetobacter calcoaceticus SW1 during growth with taurine as sole source of nitrogen. *Microbiology* 2005; **151**: 1285–1290.

116. Denger K, Smits THM, Cook AM. Genome-enabled analysis of the utilization of taurine as sole source of carbon or of nitrogen by Rhodobacter sphaeroides 2.4.1. *Microbiology* 2006; **152**: 3197–3206.

117. Palatinszky M, Herbold C, Jehmlich N, Pogoda M, Han P, von Bergen M, et al. Cyanate as an energy source for nitrifiers. *Nature* 2015; **524**: 105–108.

118. Koch H, Lücker S, Albertsen M, Kitzinger K, Herbold C, Spieck E, et al. Expanded metabolic versatility of ubiquitous nitrite-oxidizing bacteria from the genus *Nitrospira*. *Proc Natl Acad Sci USA* 2015; **112**: 11371–11376.

119. Kitzinger K, Padilla CC, Marchant HK, Hach PF, Herbold CW, Kidane AT, et al. Cyanate and urea are substrates for nitrification by Thaumarchaeota in the marine environment. *Nature Microbiology* 2019; **4**: 234–243.

120. Clifford EL, De Corte D, Amano C, Paliaga P, Ivančić I, Ortiz V, et al. Mesozooplankton taurine production and prokaryotic uptake in the northern Adriatic Sea. *Limnol Oceanogr* 2020; **65**: 2730–2747.

121. De Goeij JM, De Kluijver A, Van Duyl FC, Vacelet J, Wijffels RH, De Goeij AFPM, et al. Cell kinetics of the marine sponge Halisarca caerulea reveal rapid cell turnover and shedding. *Journal of Experimental Biology* 2009; **212**: 3892–3900.

122. Alexander BE, Liebrand K, Osinga R, van der Geest HG, Admiraal W, Cleutjens JPM, et al. Cell Turnover and Detritus Production in Marine Sponges from Tropical and Temperate Benthic Ecosystems. *PLoS ONE* 2014; **9**: e109486.

123. Agnello G, Chang LL, Lamb CM, Georgiou G, Stone EM. Discovery of a Substrate Selectivity Motif in Amino Acid Decarboxylases Unveils a Taurine Biosynthesis Pathway in Prokaryotes. *ACS Chem Biol* 2013; **8**: 2264–2271.

124. Tevatia R, Allen J, Rudrappa D, White D, Clemente TE, Cerutti H, et al. The taurine biosynthetic pathway of microalgae. *Algal Research* 2015; **9**: 21–26.

125. Moran MA, Durham BP. Sulfur metabolites in the pelagic ocean. *Nat Rev Microbiol* 2019; **17**: 665–678.

126. Wybouw N, Dermauw W, Tirry L, Stevens C, Grbić M, Feyereisen R, et al. A gene horizontally transferred from bacteria protects arthropods from host plant cyanide poisoning. *eLife* 2014; **3**: e02365.

127. Guedes R, Prosdocimi F, Fernandes G, Moura L, Ribeiro H, Ortega J. Amino acids biosynthesis and nitrogen assimilation pathways: a great genomic deletion during eukaryotes evolution. *BMC Genomics* 2011; **12**: S2.

128. Costa I, Thompson J, Ortega J, Prosdocimi F. Metazoan Remaining Genes for Essential Amino Acid Biosynthesis: Sequence Conservation and Evolutionary Analyses. *Nutrients* 2014; **7**: 1–16.

129. Richter DJ, Fozouni P, Eisen MB, King N. Gene family innovation, conservation and loss on the animal stem lineage. *eLife* 2018; **7**: e34226.

130. Munroe S, Sandoval K, Martens DE, Sipkema D, Pomponi SA. Genetic algorithm as an optimization tool for the development of sponge cell culture media. *In Vitro CellDevBiol-Animal* 2019; **55**: 149–158.

131. Salazar OR, N. Arun P, Cui G, Bay LK, van Oppen MJH, Webster NS, et al. The coral *Acropora loripes* genome reveals an alternative pathway for cysteine biosynthesis in animals. *Sci Adv* 2022; **8**: eabq0304.

132. Fitzgerald LM, Szmant AM. Biosynthesis of 'essential' amino acids by scleractinian corals. *Biochemical Journal* 1997; **322**: 213–221.

133. Dassler T, Maier T, Winterhalter C, Bock A. Identification of a major facilitator protein from Escherichia coli involved in efflux of metabolites of the cysteine pathway. *Mol Microbiol* 2000; **36**: 1101–1112.

134. Kutukova EA, Livshits VA, Altman IP, Ptitsyn LR, Zyiatdinov MH, Tokmakova IL, et al. The *yeaS* (*leuE*) gene of *Escherichia coli* encodes an exporter of leucine, and the Lrp protein regulates its expression. *FEBS Letters* 2005; **579**: 4629–4634.

135. Elliott GRD, Leys SP. Evidence for glutamate, GABA and NO in coordinating behaviour in the sponge, Ephydatia muelleri (Demospongiae, Spongillidae). *Journal of Experimental Biology* 2010; **213**: 2310–2321.

136. Ueda N, Richards GS, Degnan BM, Kranz A, Adamska M, Croll RP, et al. An ancient role for nitric oxide in regulating the animal pelagobenthic life cycle: evidence from a marine sponge. *Scientific Reports* 2016; **6**.

137. Cramm R, Pohlmann A, Friedrich B. Purification and characterization of the single-component nitric oxide reductase from *Ralstonia eutropha* H16. *FEBS Letters* 1999; **460**: 6–10.

138. Hendriks J, Oubrie A, Castresana J, Urbani A, Gemeinhardt S, Saraste M. Nitric oxide reductases in bacteria. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 2000; **1459**: 266–273.

139. Leys SP, Kahn AS, Fang JKH, Kutti T, Bannister RJ. Phagocytosis of microbial symbionts balances the carbon and nitrogen budget for the deep-water boreal sponge *Geodia barretti*: Metabolism of *Geodia barretti*. *Limnol Oceanogr* 2018; **63**: 187–202.

140. Deane EM, O'Brien RW. Uptake of sulphate, taurine, cysteine and methionine by symbiotic and free-living dinoflagellates. *Arch Microbiol* 1981; **128**: 311–319.

141. Wang JT, Douglas AE. Nutrients, Signals, and Photosynthate Release by Symbiotic Algae (The Impact of Taurine on the Dinoflagellate Alga Symbiodinium from the Sea Anemone Aiptasia pulchella). *Plant Physiology* 1997; **114**: 631–636.

Supplementary Information

Taurine as a key intermediate for host-symbiont interaction in the tropical sponge *Ianthella basta*

Florian U. Moeller, Craig W. Herbold, Arno Schintlmeister, Maria Mooshammer, Cherie Motti, Faris Behnam, Margarete Watzka, Thomas Schweder, Mads Albertsen, Andreas Richter, Nicole S. Webster, and Michael Wagner

Table of contents:

Supplementary Materials and Methods

***NanoSIMS analysis of* Ianthella basta**. NanoSIMS measurements were performed on a NanoSIMS 50L (Cameca, Gennevilliers, France) on two of the three *I. basta* explant samples obtained from the $^{13}C^{15}N$-taurine incubations at the Large-Instrument Facility for Environmental and Isotope Mass Spectrometry at the University of Vienna. Specimens were selected for maximum isotope enrichment, based on results from EA-IRMS bulk analysis (0.62 and 0.55 at% ^{15}N; 1.19 and 1.15 at% ^{13}C). Prior to data acquisition, analysis areas were pre-sputtered utilizing a high-intensity, slightly defocused Cs^+ ion beam (400 pA beam current, ~1 μm spot size) until a Cs^+ fluence of 3.2E15 ions cm^{-2} was obtained. Data were

subsequently acquired as multilayer image stacks by sequential scanning of a finely focused Cs^+ primary ion beam (*ca*. 80 nm probe size at 2 pA beam current) over areas between 10.3 × 15 and 29 × 29 μm^2 at 176 × 256 and 512 × 512 pixel image resolution and a primary ion beam dwell time of 10 to 40 msec (pixel*cycle)$^{-1}$, corresponding to a per-cycle Cs^+ fluence in the range from 8.0E13 to 1.5E16 ions cm^{-2} per cycle. $^{12}C^-$, $^{13}C^-$, $^{12}C^{12}C^-$, $^{12}C^{13}C^-$, $^{12}C^{14}N^-$, $^{12}C^{15}N^-$, $^{31}P^-$ or $^{32}S^-$, as well as secondary electrons, were detected simultaneously. Signal intensities were corrected for detector dead time and quasi-simultaneous arrival (QSA) effects using QSA sensitivity factors (i.e., "beta" values) of 1.1, 1.1, 1.06, 1.06, 1.05, 1.05, 1.1, 1.1 for $^{12}C^-$, $^{13}C^-$, $^{12}C^{12}C^-$, $^{12}C^{13}C^-$, $^{12}C^{14}N^-$, $^{12}C^{15}N^-$, $^{31}P^-$, and ^{32}S, respectively.

Images based on NanoSIMS measurement data were generated using the Open MIMS plugin [1] in the FIJI package based on ImageJ [2]. C isotope composition images displaying the $^{13}C/(^{12}C+^{13}C)$ isotope fraction, designated as atom% ^{13}C, were inferred from the C_2^- secondary ion signal intensity distribution images via per-pixel calculation of $^{13}C^{12}C^-/(2\cdot{}^{12}C^{12}C^-+{}^{13}C^{12}C^-)$ intensity ratios, while N isotope composition images displaying the $^{15}N/(^{14}N+^{15}N)$ isotope fraction, designated as atom% ^{15}N, were inferred from the $^{12}CN^-$ secondary ion signal intensity maps via per-pixel calculation of $^{12}C^{15}N^-/(^{12}C^{15}N^-+{}^{12}C^{14}N^-)$ intensity ratios. Secondary electron images were utilized for representation of the sample morphology.

Region of interest (ROI) specific numerical data evaluation was also conducted using the Open MIMS plugin. ROIs were defined according to the isotope enrichment patterns since it was not possible to unambiguously identify characteristic cellular features. This was likely because analysis areas were accessed via depth-profiling, which is less suited for visualization of cellular (ultra)structure than e.g., cross-sectional analysis of resin embedded samples. C and N isotope compositions were calculated from the accumulated intensities of $^{12}C^{12}C^-$ and $^{12}C^{13}C^-$ and $^{12}C^{14}N^-$ and $^{12}C^{15}N^-$ secondary ion signals detected within each ROI. The isotopic composition values shown in Fig. S6 were determined via stepwise inspection of the isotopic enrichment patterns displayed in each individual image of the acquired multilayer stacks (Fig. S7). As such, the displayed values may rather be considered as conservative estimates of the isotopic enrichment due to carry-over (e.g., through re-deposition and/or atomic mixing) of atoms and molecules from less enriched regions within the sponge extracellular matrix during the measurement process (including pre-sputtering). The analytical uncertainty of the isotope composition values (indicated by the error bars in Fig. S6), emerging from the random error in single ion counting, was estimated on the basis of Poisson statistics and calculated from the signal intensities (given in total counts within individual ROIs) via:

atom% ^{13}C inferred from C_2^- signal intensities:

$\sigma_{Pois,\ at\%13C}=100/(2\cdot {}^{12}C^{12}C^- + {}^{12}C^{13}C^-)^2 \cdot \mathrm{sqrt}(({}^{12}C^{12}C^-)^2 \cdot {}^{12}C^{13}C^- + ({}^{12}C^{13}C^-)^2 \cdot {}^{12}C^{12}C^-)$

atom% ^{15}N inferred from CN^- signal intensities:

$\sigma_{Pois,\ at\%15N}=100/({}^{12}C^{14}N^- + {}^{12}C^{15}N^-)^2 \cdot \mathrm{sqrt}(({}^{12}C^{14}N^-)^2 \cdot {}^{12}C^{15}N^- + ({}^{12}C^{15}N^-)^2 \cdot {}^{12}C^{14}N^-)$

Supplementary Results and Discussion

Dimethylsulfoniopropionate (DMSP) catabolism and other notable metabolic features. The main sources of DMSP are micro and macro-algae [3], for which it serves as an organic osmolyte. As DMSP is ubiquitous in marine surface waters and is detected in concentration from less than 1 nM in the open oceans to several micromolar in phytoplankton blooms, this compound will also be available for filter-feeding sponge symbionts. DMSP is a source of reduced sulfur and carbon for sponge symbionts microbes and the precursor for the microbially-mediated formation of the climatically active gas dimethylsulfide (DMS). Within the '*Candidatus* Taurinisymbion ianthellae' MAG, we detected an ABC transporter (*OpuABC*) and a betaine-choline-carnitine-transporter (BCCT; *RbtB*/*SbtA*) that have both been shown to import DMSP [4, 5] along with the key genes for the DMSP cleavage (*dddP*) and demethylation (*dmdA*) pathways (Figure 3). The encoded enzymes DddP and DmdA catalyze conversion of DMSP to DMS and acrylate and to 5-methyltetrahydrofolic acid (5-Methyl-THF) and methylmercapto-propionate (MMPA), respectively. Both enzymes are abundant in marine microorganisms [6]. Although the DMSP cleavage pathway has been found in Gammaproteobacteria [7, 8], in general this pathway is more frequently found in Alphaproteobacteria [7, 9]. In addition to the DMSP cleavage and demethylation genes, we were able to identify and detect the expression of many of the respective downstream enzymes that are used to assimilate or catabolize the initial intermediates of both pathways (Table S4). For the DMSP demethylation pathway, we found homologous genes necessary for MMPA degradation to methanethiol (*dmdB*, *dmdC*, and *acuH*) [9, 10] as well as for the removal of the methyl group from DMSP. Although we did not detect DmdA or DddP expressed in the metaproteomic dataset, we were able to observe the expression of the OpuC transporter subunit, and for the DMSP cleavage pathway, the expression of most of the enzymes necessary for the degradation of acrylate to the coenzyme A (CoA) derivatives propionyl-CoA (PrpE, AcuI) and acetyl-CoA (AcuH, DddC; Figure 3) [9] with the exception of the DmdD enzyme. The AcuH enzyme has been shown to exhibit activity on

acrylate/acryloyl-CoA as well as on the DmdD substrate methylthioacrylyl-CoA (MTA-CoA) [9] (and references therein), thereby allowing it to function in both the DMSP cleavage and demethylation pathways. The hydration of MTA-CoA leads to the production of methanethiol, acetaldehyde, CO_2, and free CoA [11], and our detection of three genes in the cystathionine beta-lyases/cystathionine gamma-synthases family (*metB*, *metZ*, *metY*), two of which have been shown to convert methanethiol to methionine [12, 13], suggests the assimilation of sulfur derived from DMSP into the γ-symbiont as methionine (Figure 3 and S9). These family of enzymes are pyridoxal phosphate (PLP) dependent, and '*Candidatus* Taurinisymbion ianthellae' has the genetic capacity to produce PLP, thereby providing it with another vitamin B_{12}-independent pathway for methionine biosynthesis if DMSP is available (Figure 3 and S9). Finally, '*Candidatus* Taurinisymbion ianthellae' has the metabolic potential to use the 5-methyl-THF produced by DMSP demethylation (i) as a methyl donor for methionine, (ii) as a carbon donor in the serine assimilation pathway after oxidation into 5,10-methylene-THF by an encoded *metF*, or (iii) a as source of ATP and reductant when 5,10-methylene-THF is oxidized to formate and then to CO_2. In the metaproteome, we detected enzymes for C-1 assimilation (GlyA, Hpr, Mdh, and Mcl) as well as for 5,10-methylene-THF oxidation to formate (Fhs) (Figure 3; Tables S3 and S4).

DMSP catabolizing genes have been intermittently found in microbiome members of marine sponges [14–16]. While no data for *Ianthella basta* are available, DMSP concentrations in some other marine sponges were found to be relatively low or undetectable (3 of 5 sponges: 0.2–3 μmol/g wet wt.) [17] while zooxanthellae-harboring invertebrates such as corals and anemones are known to contain much higher concentrations of this compound [17, 18]. Similarly, phytoplankton, marine plants and more recently marine bacteria have been shown to synthesize DMSP (reviewed in [19]), but we did not detect either of the two key genes (*dysB* and MSMT) necessary for bacterial DMSP production in the *I. basta*-associated microbial MAGs. Recent single-cell measurements of the model Alphaproteobacterium *Ruegeria pomeroyi* DSS-3 under variable DMSP concentrations [20] demonstrated that at least some marine bacteria typically residing in low mean seawater concentrations of DMSP (16.91 ± 22.17 nM) [21] most likely express both DMSP degradation pathways at baseline levels, upregulate both pathways at 1 μM DMSP, and are thus primed for DMSP maxima either at the spatial microscale or during periodic hotspot events. As marine sponges have been found to be important sinks for DOM in coral reef environments [22] and coral mucus can be assimilated into marine sponge biomass [23], it is conceivable that DMSP from the surrounding ocean

environment may be processed by advantageously located marine sponge microbiomes and that benefit from cyclical phytoplankton bloom events and coral mucus release (suggested to follow lunar periodicity) [24]. In this context, it is interesting to note that a cultured representative of the ubiquitous open ocean SAR11 Alphaproteobacteria, also simultaneously express the cleavage and demethylation pathways, and that a kinetic switch dependent on DMSP concentrations may regulate these metabolic processes [11].

Since DMS is a by-product of the first step of the DMSP cleavage pathway, we checked for the presence of a potential DMS-oxidizing protein within the *I. basta*-associated microbial MAGs and found that '*Candidatus* Taurinisymbion ianthellae' encodes a putative trimethylamine (TMA) monooxygenase/flavin-containing monooxygenase with very high similarities (63% AA ID) to two trimethylamine monooxygenases (Tmm) previously demonstrated to co-oxidize DMS at similarly high affinities as TMAs (Table S4) [25, 26]. Although, the oxidation of DMS to dimethylsulfoxide (DMSO) by Tmm, does not result in any net gain of electrons or reducing equivalents, co-oxidation of DMS during periods of TMA depletion may prevent the formation of free radicals and hydrogen peroxide with the subsequent DMSO accumulation potentially supporting further antioxidant activities ([26] and references therein).

In addition to pathways for metabolizing taurine and DMSP, we also identified numerous other genes in the γ-symbiont MAG putatively involved in organosulfur compound degradation, including the dissimilation of sulfopropanediol (*hpsN*), sulfoacetate (*sauS*), sulfolactate (*suyAB*), and cysteate (*cuyA*). Nevertheless, '*Candidatus* Taurinisymbion ianthellae' has the capability to assimilate sulfate via sulfate adenylyltransferase (*sat*), adenylyl sulfate kinase (*cysC*), PAPS reductase (*cysH*), and an assimilatory sulfite reductase (*sir*).

Notably, the '*Candidatus* Taurinisymbion ianthellae' MAG encodes for a wide range of transporters representing different families, but primarily characterized by the preponderance of ATP-binding cassette (ABC) transporters, many of which were found to be expressed in the metaproteomic dataset (Table S3 & S5). Within this class of transporters, most are predicted to be involved in the transport of nitrogen containing organic compounds such as amino acids, oligopeptides, polyamines (spermidine and putrescine), compatible solutes (glycine, betaine, and proline), as well as the vitamins thiamin and cobalamin (B1 and B12).

Fractions of taurine-derived carbon and nitrogen determined within the holobiont. The C:N ratio of the assimilated taurine by the *I. basta* holobiont was 1.86 ± 0.09 (SD) and 2.37 ± 0.03 (SD), for the whole sponge and nucleic acids-based analyses, respectively which were lower than the corresponding total C:N ratios of 4.6 ± 0.07 (SD) and 3.21 ± 0.2 (SD) determined in the same sponges. These values were also lower than those found for sponge clones that were not incubated with isotope-labeled taurine (4.6 ± 0.15 and 3.40 ± 0.03). This indicates that the holobiont assimilated the taurine-derived nitrogen more efficiently than the corresponding carbon, which is consistent with the inherent C:N stoichiometry of taurine (i.e., 2.0), and suggests that the *I. basta* was nitrogen limited during the course of the incubations. *Alcaligenes defragans* NKNTAU, when grown in pure culture with taurine as a sole carbon and nitrogen source, respires 50% of taurine-supplied carbon and converts the other 50% into biomass, while it incorporates 20% of taurine-derived nitrogen into biomass. The remaining taurine-derived nitrogen is accumulated in the medium as NH_4^+, with SO_4^{2-} also produced in a near 1:1 ratio to the taurine supplied [27]. *A. defragans* NKNTAU was also found to express Tpa, Xsc and a Pta when supplied with taurine, thereby converting taurine into acetyl-CoA [28]. In our taurine incubation experiments, we determined a comparable value of 21.0 ± 9.1 % (SE) of fixed taurine nitrogen in the *I. basta* holobiont biomass, obtained from normalization of the amount assimilated into the holobiont biomass to the total amount of metabolized taurine (comprising the fraction of assimilated nitrogen and the released NH_4^+ — see Material and Methods for details). The mean amounts of assimilated nitrogen and released ammonia were 20.74 ± 7.7 (SE) µmol N assimilated and 104.1 ± 6.0 (SE) µmol N holobiont tissue per g holobiont tissue wet wt, respectively. However, these values do not take into account the potential loss of dissolved inorganic nitrogen via denitrification and dissimilatory nitrate reduction to ammonium (DNRA). Similarly, rapid sponge cell turnover may also result in the loss of fixed N biomass.

By assuming that the amount of ^{15}N released as ammonium and nitrite or fixed into sponge biomass is indicative of the total taurine processed, the relative amount of ^{13}C-carbon fixed would represent 40.1 ± 19 % (SE; range of 77.3–20.0 %) of the supplied taurine ^{13}C-carbon in the incubation experiments. In concordance with the above interpretation that the *I. basta* holobiont is N-limited, this would additionally indicate that the amount of taurine supplied was in excess of the carbon biomass requirements of the *I. basta* holobiont, and that the rapid choanocyte turnover characteristic of marine sponges [29, 30] may have

contributed to the additional loss of supplied carbon. The incubations with isotopically labeled taurine were conducted in natural seawater and due to its high SO_4^{2-} concentration no reasonable SO_4^{2-} flux estimations for inferring taurine consumption could be made. In incubations conducted in SFASW with unlabeled taurine, we observed near stoichiometric production of SO_4^{2-} from the added 100 µM taurine after 48 h, while net SO_4^{2-} recovery in incubations with 1 mM and 1.6 mM taurine conducted in SFASW were 350 ± 18 and 286 ± 36 (SE) µM SO_4^{2-} after 48 h, respectively. Using the net flux of SO_4^{2-} derived from the 1.6 mM taurine incubations in SFASW as a measure of the total amount of taurine metabolized by the sponge holobiont would then yield inferred carbon and nitrogen incorporation values from taurine of 15.5 ± 8.0 % and 8.1 ± 0.4 % (SE), respectively. Since we demonstrated that 100 µM NH_4^+ does not affect SO_4^{2-} flux in the same manner as it does with ammonia oxidation within the *I. basta* holobiont [31], this would indicate that either sulfite oxidation is decoupled from carbon and nitrogen incorporation into biomass, or that the potential physiological stress caused by the high NH_4^+ concentrations caused a further increase in the turnover of marine sponge choanocytes.

It is tempting to speculate that the microbial-sized cells exhibiting ^{13}C- and ^{15}N- enrichments in the NanoSIMS analyses primarily comprise the '*Candidatus* Taurinisymbion ianthellae' population residing in *I. basta* based on (i) the detection of the key genes for the taurine dissimilation–sulfite oxidation pathway in the '*Candidatus* Taurinisymbion ianthellae' MAG, (ii) the detection of some of the respective proteins in the metaproteome, and (iii) the marked differences in the accumulation of sulfate in incubations with added taurine, which indicate the intracellular oxidation of sulfite, a process for which no other of the dominant microbial members in the *I. basta* holobiont have the apparent capability to mediate. These results are therefore consistent with the recent NanoSIMS-based analyses of marine sponge DOM uptake [32, 33], which demonstrated that high-microbial abundance (HMA) and low-microbial abundance (LMA) sponges have the capability to take up significant proportions of supplied DOM and furthermore, that within one HMA sponge, the symbiotic bacteria are primarily responsible for the processing of externally sourced DOM [33]. Although we added taurine in our incubations as an exogenous source, we also detected sulfate accumulation in the incubation media of all incubations containing *I. basta* without added taurine. In this context, it is important to note that the '*Candidatus* Taurinisymbion ianthellae' MAG, also contains genes for the catabolism of the C_3-sulfonates sulfopropanediol (*hpsN*), sulfoacetate (*sauS*), sulfolactate (*suyAB*), and cysteate (*cuyA*), which all can lead to the production of sulfate [34–36] yet these compounds are all

primarily derived from eukaryotic phytoplankton and cyanobacteria in the marine environment [36]. In addition to the fact that all incubations were conducted in the dark, cyanobacterial or eukaryotic photosymbionts do not reside in *I. basta* and the C_3-sulfonate catabolism genes were not found as expressed as proteins in the metaproteomic dataset. The sulfate concentration enhancement ranging from 109 to 10 μM SO_4^{2-} in *I. basta* incubations without added taurine and conducted in SFASW therefore most likely reflect the processing of endogenously produced taurine.

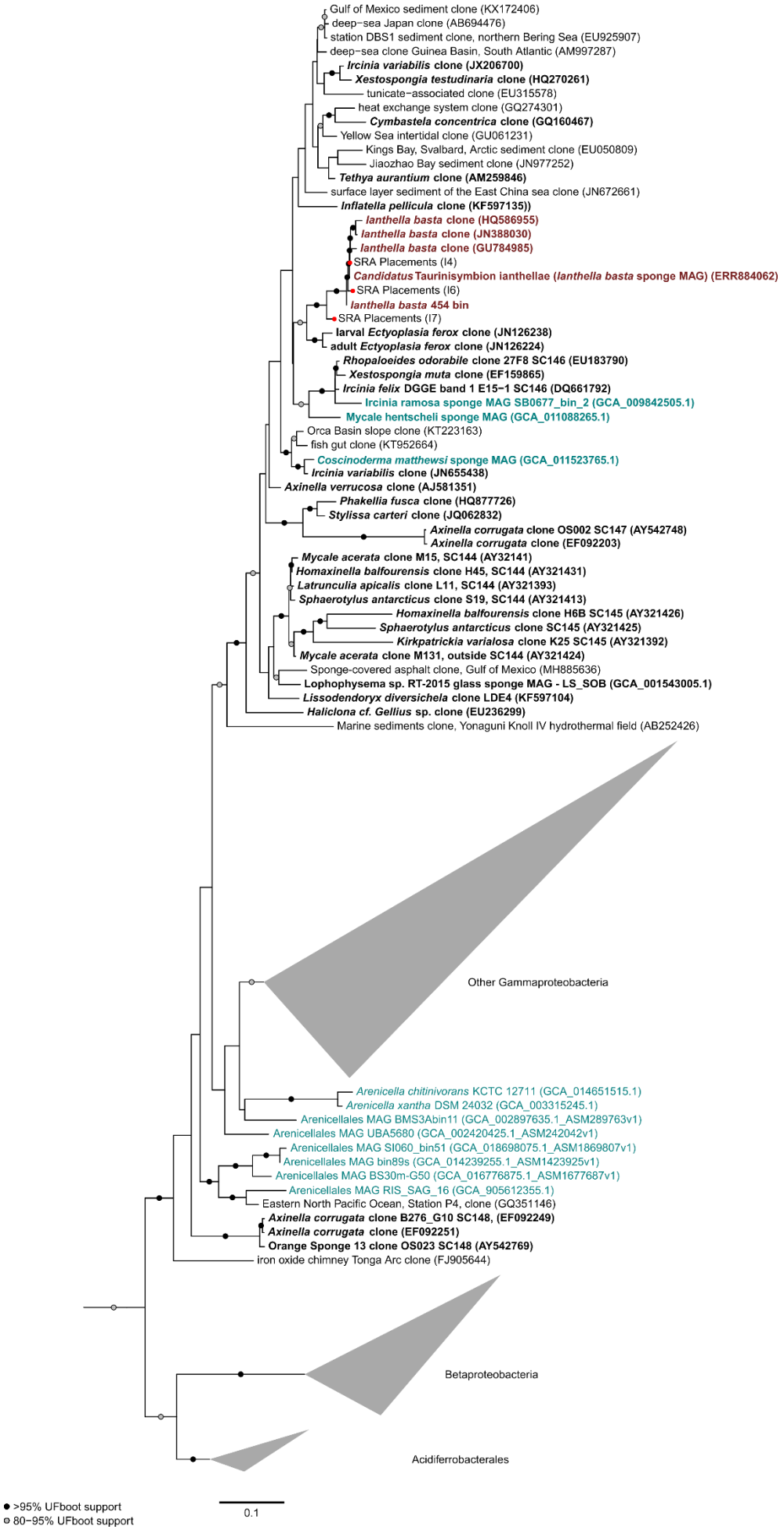

Gulf of Mexico sediment clone (KX172406)
deep−sea Japan clone (AB694476)
station DBS1 sediment clone, northern Bering Sea (EU925907)
deep−sea clone Guinea Basin, South Atlantic (AM997287)
Ircinia variabilis clone (JX206700)
Xestospongia testudinaria clone (HQ270261)
tunicate−associated clone (EU315578)
heat exchange system clone (GQ274301)
Cymbastela concentrica clone (GQ160467)
Yellow Sea intertidal clone (GU061231)
Kings Bay, Svalbard, Arctic sediment clone (EU050809)
Jiaozhao Bay sediment clone (JN977252)
Tethya aurantium clone (AM259846)
surface layer sediment of the East China sea clone (JN672661)
Inflatella pellicula clone (KF597135))
Ianthella basta clone (HQ586955)
Ianthella basta clone (JN388030)
Ianthella basta clone (GU784985)
SRA Placements (I4)
Candidatus Taurinisymbion ianthellae (Ianthella basta sponge MAG) (ERR884062)
SRA Placements (I6)
Ianthella basta 454 bin
SRA Placements (I7)
larval Ectyoplasia ferox clone (JN126238)
adult Ectyoplasia ferox clone (JN126224)
Rhopaloeides odorabile clone 27F8 SC146 (EU183790)
Xestospongia muta clone (EF159865)
Ircinia felix DGGE band 1 E15−1 SC146 (DQ661792)
Ircinia ramosa sponge MAG SB0677_bin_2 (GCA_009842505.1)
Mycale hentscheli sponge MAG (GCA_011088265.1)
Orca Basin slope clone (KT223163)
fish gut clone (KT952664)
Coscinoderma matthewsi sponge MAG (GCA_011523765.1)
Ircinia variabilis clone (JN655438)
Axinella verrucosa clone (AJ581351)
Phakellia fusca clone (HQ877726)
Stylissa carteri clone (JQ062832)
Axinella corrugata clone OS002 SC147 (AY542748)
Axinella corrugata clone (EF092203)
Mycale acerata clone M15, SC144 (AY32141)
Homaxinella balfourensis clone H45, SC144 (AY321431)
Latrunculia apicalis clone L11, SC144 (AY321393)
Sphaerotylus antarcticus clone S19, SC144 (AY321413)
Homaxinella balfourensis clone H6B SC145 (AY321426)
Sphaerotylus antarcticus clone SC145 (AY321425)
Kirkpatrickia varialosa clone K25 SC145 (AY321392)
Mycale acerata clone M131, outside SC144 (AY321424)
Sponge-covered asphalt clone, Gulf of Mexico (MH885636)
Lophophysema sp. RT-2015 glass sponge MAG - LS_SOB (GCA_001543005.1)
Lissodendoryx diversichela clone LDE4 (KF597104)
Haliclona cf. Gellius sp. clone (EU236299)
Marine sediments clone, Yonaguni Knoll IV hydrothermal field (AB252426)
Other Gammaproteobacteria
Arenicella chitinivorans KCTC 12711 (GCA_014651515.1)
Arenicella xantha DSM 24032 (GCA_003315245.1)
Arenicellales MAG BMS3Abin11 (GCA_002897635.1_ASM289763v1)
Arenicellales MAG UBA5680 (GCA_002420425.1_ASM242042v1)
Arenicellales MAG SI060_bin51 (GCA_018698075.1_ASM1869807v1)
Arenicellales MAG bin89s (GCA_014239255.1_ASM1423925v1)
Arenicellales MAG BS30m-G50 (GCA_016776875.1_ASM1677687v1)
Arenicellales MAG RIS_SAG_16 (GCA_905612355.1)
Eastern North Pacific Ocean, Station P4, clone (GQ351146)
Axinella corrugata clone B276_G10 SC148, (EF092249)
Axinella corrugata clone (EF092251)
Orange Sponge 13 clone OS023 SC148 (AY542769)
iron oxide chimney Tonga Arc clone (FJ905644)
Betaproteobacteria
Acidiferrobacterales
>95% UFboot support
80−95% UFboot support
0.1

Figure S1. Maximum-likelihood phylogenetic tree of the 16S rRNA gene (after automatic model selection with IQ-Tree) of 'Candidatus Taurisymbion ianthellae'(red and bold) along with other selected sponge (bold font). Entries highlighted in teal represent 16S rRNA sequences extracted from MAGs represented in Figure 1. 'Candidatus Taurisymbion ianthellae' 16S rRNA gene sequences were queried against the SRA database, and the hits were placed into the reference 16S rRNA gene tree (represented on the red nodes) using the Evolutionary Placement Algorithm [37]. Members of the sponge specific clusters (SC) [38] 144-148 are indicated in the tree.

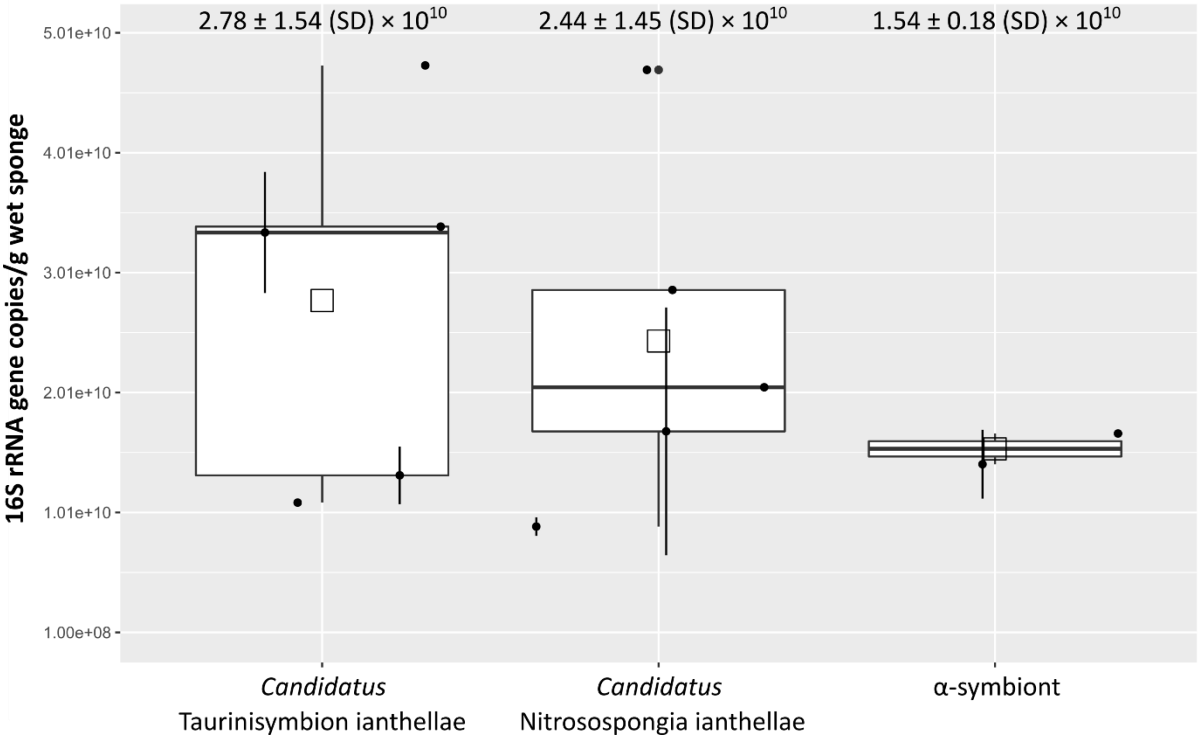


Figure S2. Absolute abundance of the three dominant microbial symbionts residing in *Ianthella basta* as quantified using qPCR from five sponge individuals with the exception of the α-symbiont, including those used for metagenome sequencing. Boxplots illustrate summary statistics, indicating the median (horizontal bars), first and third quartile (vertical bars: Q1/Q3 + 1.5 × IQR) as well as the mean (open square). The error bars at the individual data points refer to standard deviation values where samples were measured multiple times. The mean values are also displayed at the top for each microbial symbiont. Based on previously acquired FISH and TEM images from *I. basta* ([39, 40]) we conclude that '*Ca*. Taurinisymbion ianthellae' are most likely rod-shaped cells that are 0.8–1.1 µm long and 0.1–0.28 µm in width. From these assumptions, we can make a calculation of the volume of cells that the estimated abundance of '*Ca*. Taurinisymbion ianthellae' at 2.78 ± 1.5 (SD) x 10^{10} 16S rRNA gene copies per g wet sponge, would roughly occupy per cm^3. The putative '*Ca*. Taurinisymbion ianthellae' cells would thus occupy ~0.02–0.19% of 1 cm^3. Keeping in mind that HMA sponges are considered to have microbial densities of 10^8–10^{10} per gram of sponge wet weight [41] the abundance numbers of all of the three dominant microbial symbionts within *I. basta* would still add up to 10^{10} per gram of sponge wet weight.

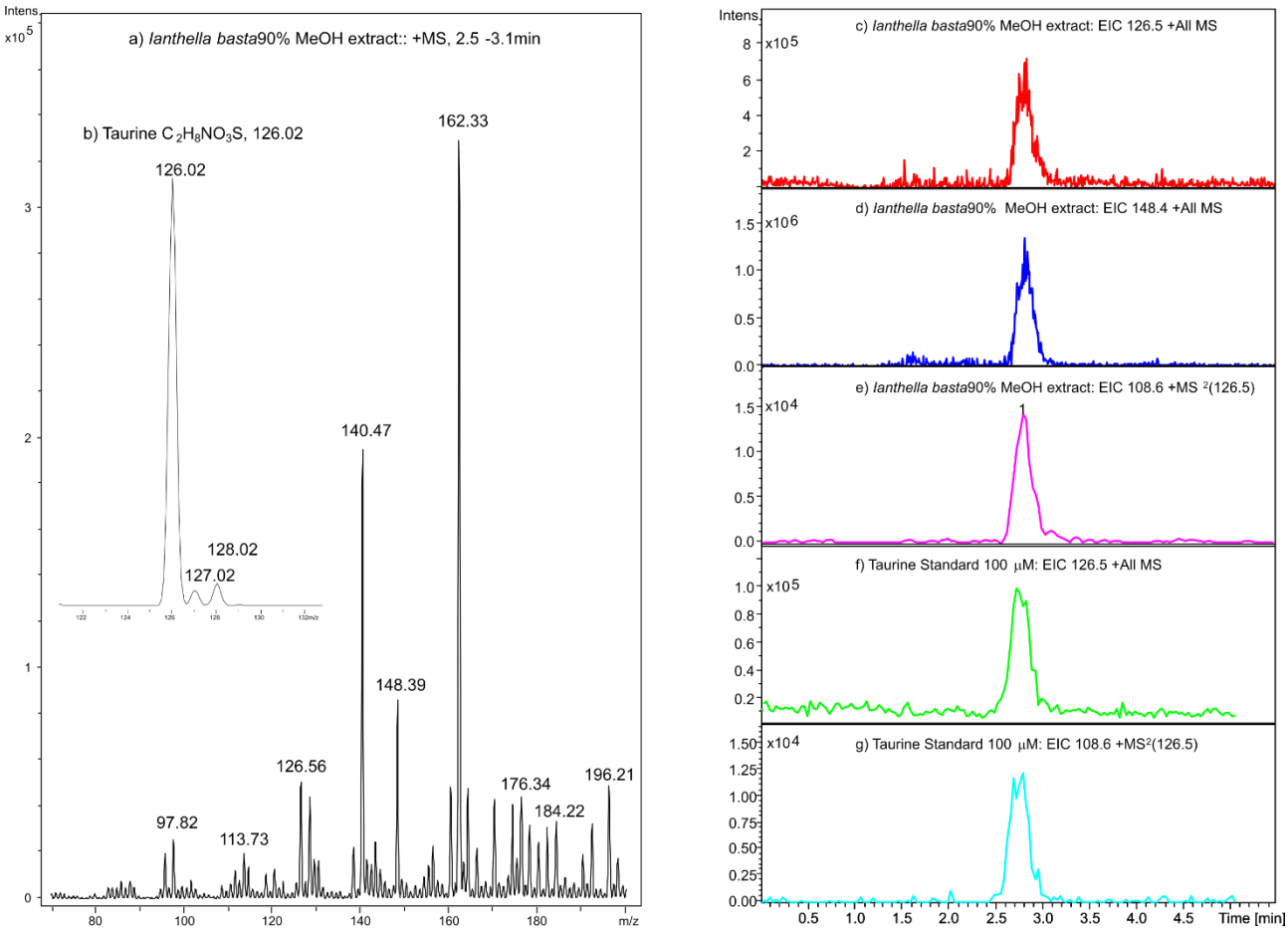


Figure S3. LCMS analysis of taurine in *Ianthella basta*. a) Mass spectrum of peak at retention time 2.5–3.1 min for the 90% methanol extract of *Ianthella basta*. Signal at 126.5 *m/z* matches that expected for taurine $[M+H]^+$ ion ($C_2H_7NO_3S$); b) inset shows the expected isotope distribution. Retention time and extracted ion chromatogram (EIC) for c) $[M+H]^+$ (126.5 *m/z*) and d) $[M+Na]^+$ (148.4 *m/z*) of the 90% methanol extract of *I. basta* and e) the expected $[M+H-H_2O]^+$ fragmentation ion at 108.6 *m/z*, as confirmed by comparison with f) the $[M+H]^+$ and g) MS^2 EIC of the taurine standard.

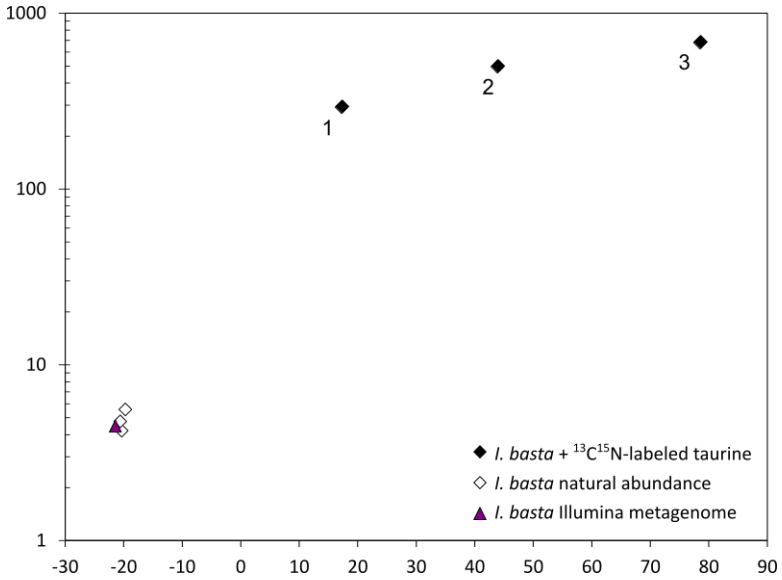


Figure S4. $\delta^{13}C$ and $\delta^{15}N$ (‰) values determined in *Ianthella basta* whole sponge tissue homogenate. Three biological replicates (numbered) were analyzed after incubation with isotopically labeled taurine and unlabeled taurine (natural abundance control), respectively. In addition, total nucleic acids (Fig. 4) were extracted from fractions of the samples, for which the results for whole sponge tissue homogenate are depicted. In the analysis, the sponge sample that was used for Illumina-sequencing based metagenomic analysis was also included as an additional natural abundance control.

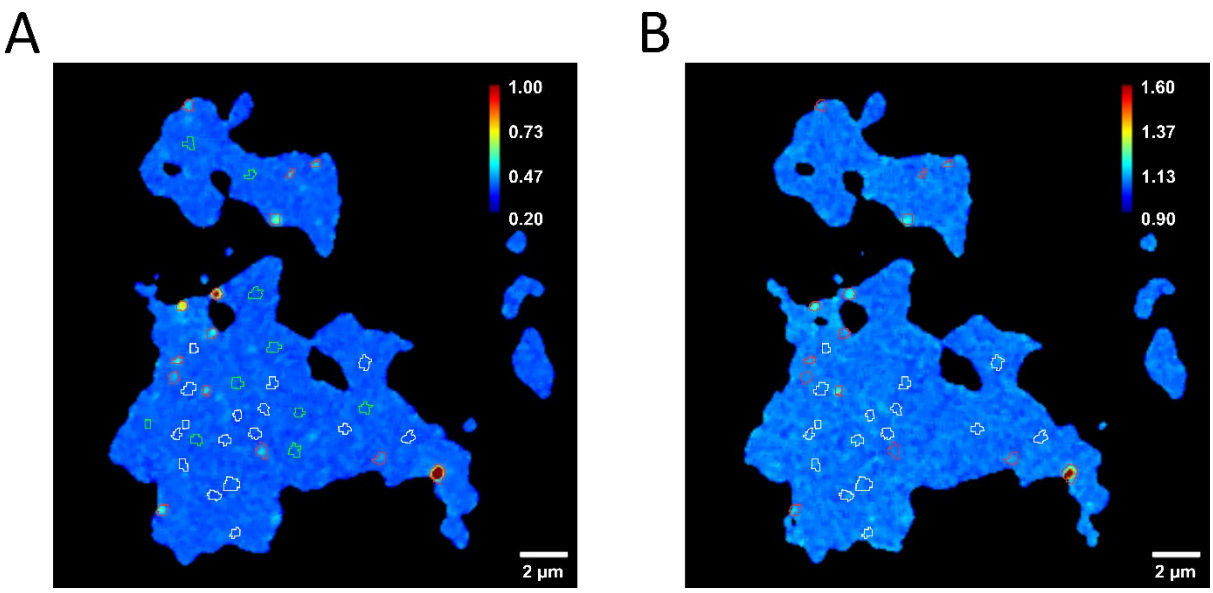


Fig. S5. Illustration of the definition and classification of regions of interest (ROI) for numerical evaluation of NanoSIMS measurement data in a representative image from a particular acquisition cycle from *I. basta* extruded extracellular material with respect to the local ^{15}N (A) and ^{13}C (B) contents (c.f. Fig. S6). The displayed images were acquired at an erosion depth below the surface of the sample, visualizing the presence of isotopically enriched compounds within the sponge extracellular matrix. Red circles refer to ROIs selected for isotope enrichment observed in a particular image cycle as exemplified by (A) and (B). White circles refer to ROIs, in which an isotope enrichment was observed above or below the displayed image layer. Green circles refer to regions exhibiting no recognizable isotope enrichment within any of the acquired image cycles. This representative image depicts the incorporation of taurine-derived ^{15}N and ^{13}C into microbial-sized cells as detected by NanoSIMS analysis of extruded extracellular matrix from *I. basta* holobiont samples after incubation with $^{15}N/^{13}C$ labeled taurine over a period of 48 h. Isotope label contents are displayed as isotope fractions on a rainbow color scale, ranging from 0.2 to 1.0 at% ^{15}N and 0.9 to 1.6 at% ^{13}C. The isotope label content of these cells ranged from 0.42–1.15 at% 15N and 1.09–1.47 at% 13C (Fig. S6), which significantly exceeded the values determined within regions of the sponge extracellular matrix (Mann–Whitney U-test, $p < 0.001$), consistent with (a) population(s) of microorganisms specialized in using taurine. Unfortunately, the sponge samples (three explants from the same sponge individual, of which two were used for NanoSIMS analyses) from the taurine incubation experiments could not be analyzed by combining fluorescence in situ hybridization (FISH) with symbiont specific probes and NanoSIMS as this sponge individual exhibited very high autofluorescence at each wavelength typically exploited for the detection of FISH-probe derived signals.

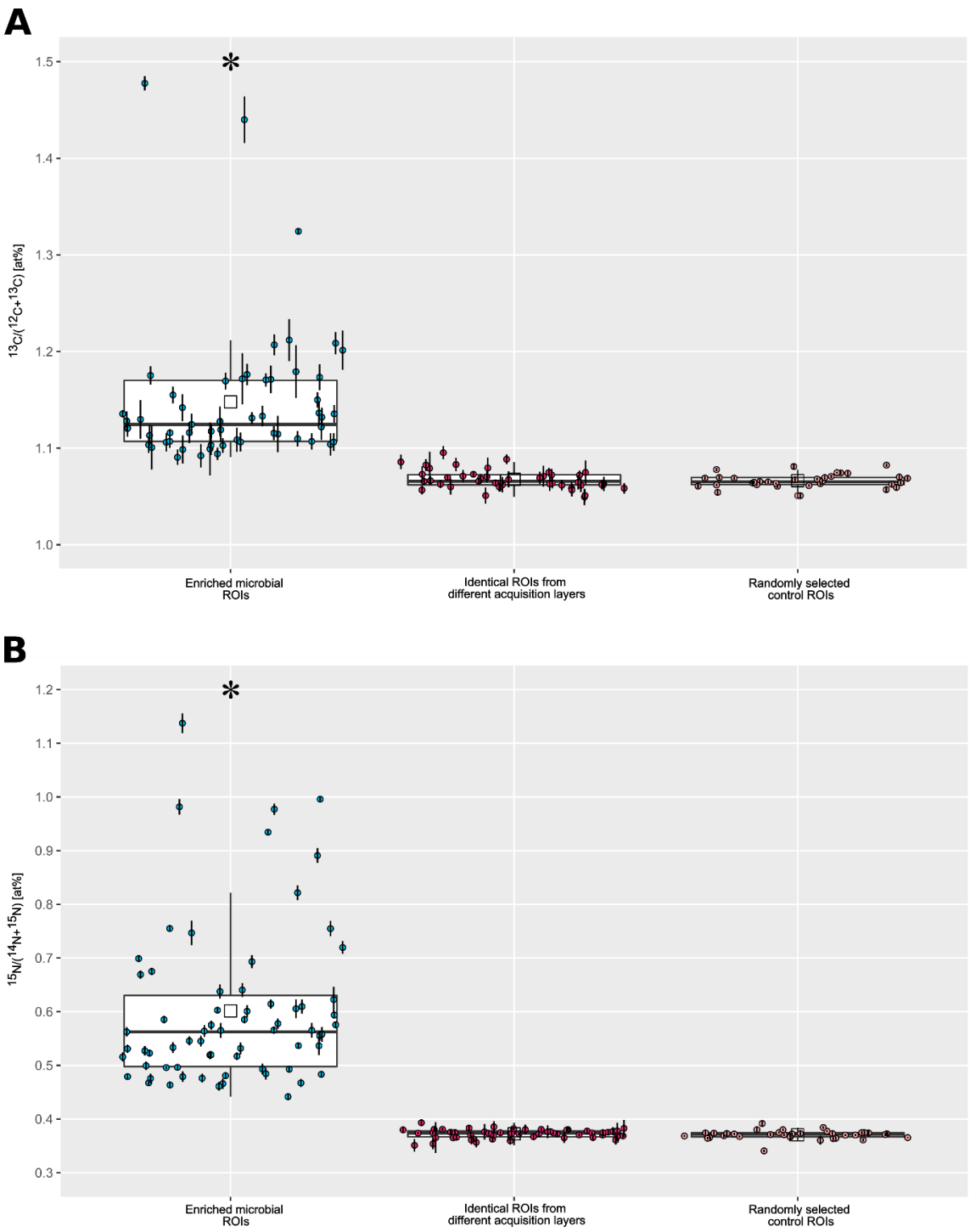


Figure S6. Results from region of interest (ROI) specific evaluation of NanoSIMS measurement data from *I. basta* extruded extracellular material. Data acquisition was conducted as multilayer imaging, yielding depth profiles through continuous sputter-erosion by the Cs^+ primary ion beam. Each data point refers to the local isotope label content of ^{13}C (A) and ^{15}N (B) detected within microbial-sized regions. The values displayed on the left refer to ROIs selected for isotope enrichment observed in individual image cycles. The values in the center refer to the values determined in the identical ROIs above or below the image layers indicating isotope enrichment. The values shown on the right correspond to randomly selected microbial-sized ROIs exhibiting no recognizable isotope enrichment within any of the acquired image cycles. The error bars at the individual data points refer to the estimated analytical uncertainty (1σ, Poisson) due to counting statistics (see section Materials and Methods). The definition and classification of ROIs is illustrated in Fig. S5. Boxplots illustrate summary statistics, indicating the median (horizontal bars), first and third quartile (vertical bars: Q1/Q3 + 1.5 × IQR) as well as the mean (open square). Significant differences in isotope enrichment (one-way ANOVA followed by Tukey's test, both $p < 0.001$) are indicated with an asterisk (*).

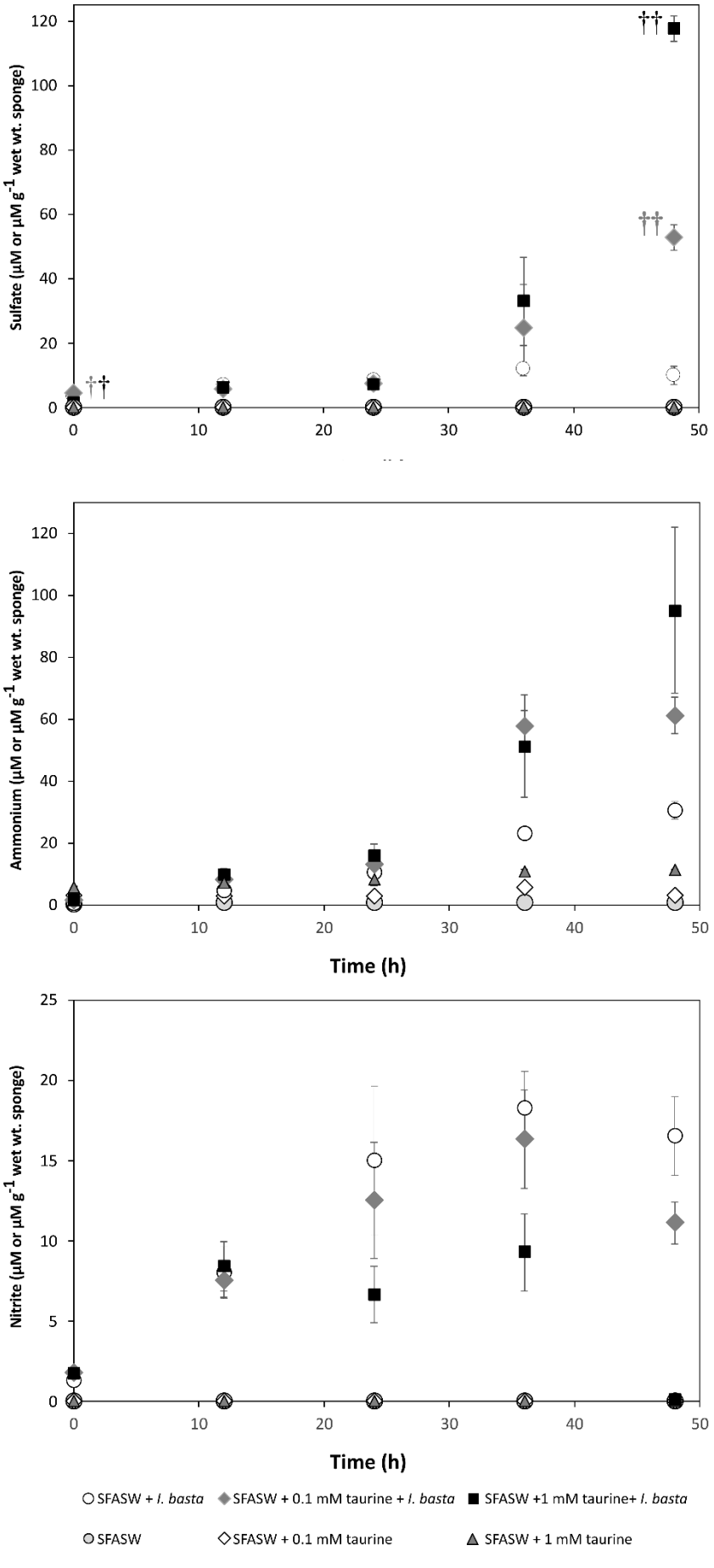

Sulfate (µM or µM g-1 wet wt. sponge)
Ammonium (µM or µM g-1 wet wt. sponge)
Nitrite (µM or µM g-1 wet wt. sponge)
Time (h)
Time (h)
SFASW + I. basta
SFASW + 0.1 mM taurine + I. basta
SFASW +1 mM taurine+ I. basta
SFASW
SFASW + 0.1 mM taurine
SFASW + 1 mM taurine

Figure S7. Time series of the concentrations of dissolved sulfate, ammonium, and nitrite in the incubation media of *I. basta* holobiont batch incubations (48 h) in sulfate-free artificial seawater (SFASW) (sampling times at t = 0, 12, 24, 36 and 48 h) performed either with added unlabeled taurine (filled black square: t = 0, +1 mM taurine; filled grey diamond: t = 0, +0.1 mM taurine) or without added taurine (white circle). This resulted in six distinct incubation treatments performed in biological triplicates. The displayed concentrations refer to the measurement values normalized to the wet weight of the respective *I. basta* explants (2.98 ± 0.83 SD g sponge wet wt; range: 1.83–5.28 g sponge wet wt) used in the incubations. The data from the corresponding control incubations with SFASW, but without *I. basta* explants are shown as grey circles for SFASW-only treatments, while SFASW with 0.1 mM taurine and 1 mM taurine treatments are represented by white diamonds and grey triangles respectively. The data for these control incubations are plotted on the same axes as µM. †† symbols in the sulfate concentration panel indicate significant differences for each treatment containing *I. basta* and taurine at t = 48 h (††) relative to at t = 0 (†) (one-way ANOVA followed by Tukey pairwise multiple comparison tests, $p < 0.05$). Nitrite concentrations were found below the limit of detection in all of the incubations conducted in the absence of the *I. basta* holobiont. Error bars refer to +/- 1SE.

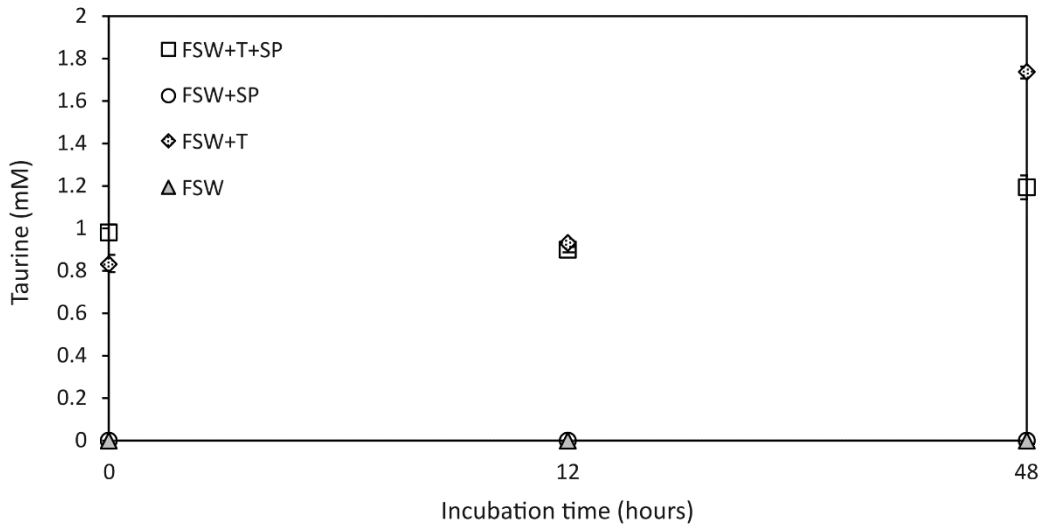


Figure S8. Changes in taurine concentrations in seawater after incubation of *Ianthella basta* with 1.6 mM taurine. FSW = natural filtered seawater, FSW+SP = natural filtered seawater conditioned with sponge, FSW+T = natural filtered seawater incubated with taurine (1 mM added at 0 h and 0.6 mM at 36 h) and FSW+T+SP = natural filtered seawater with sponge incubated with taurine (1 mM added at 0 h and 0.6 mM at 36 h). Kruskal-Wallis was used to establish significance. The lower concentration of taurine in FSW+T+SP at 48 h as compared to FSW+T confirms taurine uptake by *I. basta* ($p < 0.05$).

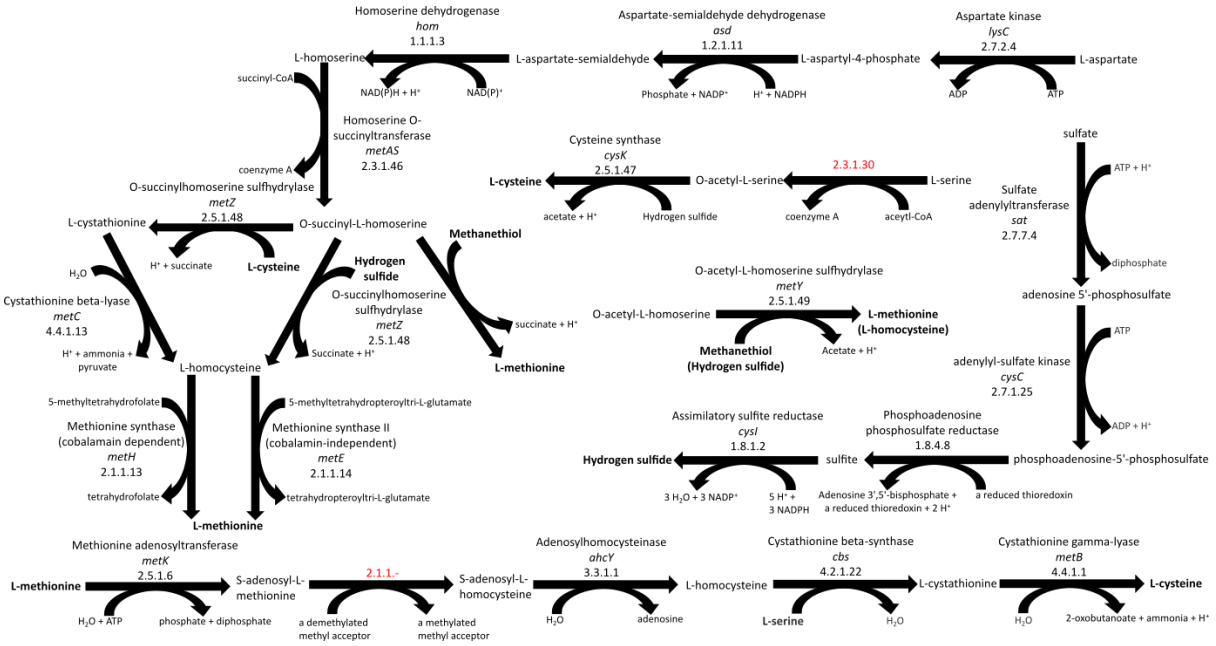


Figure S9. Schematic depictions of pathways identified in '*Ca.* Taurinisymbion ianthellae' involved in homocysteine, methionine, and cysteine synthesis. Homoserine synthesis is followed by homocysteine synthesis and culminating in methionine synthesis via either a cobalamin dependent (*metH*), or a cobalamin independent (*metE*) pathway. Alternatively, methionine could be synthesized via methanethiol if supplied as part of the DMSP demethylation pathway as described in Figure 3, which is also cobalamin independent. O-succinylhomoserine sulfhydrylase is a key enzyme (MetZ) that can convert different starting agents either via producing cystathionine when supplied with cysteine, homocysteine when supplied with hydrogen sulfide (from sulfate assimilation — also depicted), or methionine when supplied with methanethiol [12]v(https://enzyme.expasy.org/EC/2.5.1.48). Similarly, MetY, if supplied with O-acetyl-L-homoserine, produces methionine and homocysteine from methanethiol and hydrogen sulfide, respectively [13] (https://enzyme.expasy.org/EC/2.5.1.49). Enzymes labeled in red were not found encoded in the "*Ca.* T. ianthellae" MAG.

Table S1. Overview of genome binning parameters, statistics and associated metadata for the IBGammaO2 genome bin of 'Candidatus Taurisymbion ianthellae'.

Bin ID	**IBGammaO2**
Analysis project type	metagenome-assembled genome (MAG)
Taxa_id	16S rRNA and multi-marker phylogenetics
Assembly software	Spades (--careful)
Annotation	MaGe
Genome Quality	High Quality Draft
Completeness (%)	94.82
Contamination (%)	1.87
Completeness/Contamination Software	CheckM
Number of contigs	97
16S rRNA gene recovered	yes
16S rRNA gene recovery software	rnammer / MaGe
Number of standard tRNAs extracted	20
tRNA extraction software	tRNA-scan / MaGe
Binning software	metabat2
Binning parameters	kmer
Genome size (Mbp)	2.35
N50 (bp)	41,025
L50 (bp)	20
Longest contig (bp)	113,413
Average contig length (bp)	24,142.92
GC content (%)	63.49
Protein coding sequences	2072
rRNAs	3
tRNAs	40

Supplementary Note

Recipes for calcium- and magnesium-free artificial seawater (CMF-ASW) and sulfate-free artificial seawater (SFASW)

Ingredients	CMF-ASW mM	SFASW mM
NaCl	461.85	474.33
KCl	10.73	9.00
$CaCl_2$	-	9.27
$MgCl_2$	-	97.88
Na_2SO_4	7.04	-
*$NaHCO_3$	2.14	2.14
H_2O	to 1L	to 1L

* Add $NaHCO_3$ last.

Adjust to pH 7.8, autoclave, and store at 4°C

References

1. Poczatek C, Kaufman Z, Lechene C. OpenMIMS ImageJ Plugin Guide Opening MIMS images. 2009. Harvard Med. Sch.
2. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods* 2012; **9**: 676–682.
3. Yoch DC. Dimethylsulfoniopropionate: Its Sources, Role in the Marine Food Web, and Biological Degradation to Dimethylsulfide. *Appl Environ Microbiol* 2002; **68**: 5804–5815.
4. Holtmann G, Bremer E. Thermoprotection of Bacillus subtilis by Exogenously Provided Glycine Betaine and Structurally Related Compatible Solutes: Involvement of Opu Transporters. *J Bacteriol* 2004; **186**: 1683–1693.
5. Sun L, Curson ARJ, Todd JD, Johnston AWB. Diversity of DMSP transport in marine bacteria, revealed by genetic analyses. *Biogeochemistry* 2012; **110**: 121–130.
6. Moran MA, Reisch CR, Kiene RP, Whitman WB. Genomic Insights into Bacterial DMSP Transformations. *Annu Rev Mar Sci* 2012; **4**: 523–542.
7. Zheng Y, Wang J, Zhou S, Zhang Y, Liu J, Xue C-X, et al. Bacteria are important dimethylsulfoniopropionate producers in marine aphotic and high-pressure environments. *Nat Commun* 2020; **11**: 4658.
8. Zhang S, Cao H-Y, Zhang N, Teng Z-J, Yu Y, Wang Z-B, et al. Novel Insights into Dimethylsulfoniopropionate Catabolism by Cultivable Bacteria in the Arctic Kongsfjorden. *Appl Environ Microbiol* 2022; **88**: e01806-21.
9. Bullock HA, Luo H, Whitman WB. Evolution of Dimethylsulfoniopropionate Metabolism in Marine Phytoplankton and Bacteria. *Front Microbiol* 2017; **8**.
10. Reisch CR, Stoudemayer MJ, Varaljay VA, Amster IJ, Moran MA, Whitman WB. Novel pathway for assimilation of dimethylsulphoniopropionate widespread in marine bacteria. *Nature* 2011; **473**: 208–211.
11. Sun J, Todd JD, Thrash JC, Qian Y, Qian MC, Temperton B, et al. The abundant marine bacterium Pelagibacter simultaneously catabolizes dimethylsulfoniopropionate to the gases dimethyl sulfide and methanethiol. *Nat Microbiol* 2016; **1**: 1–5.
12. Kiene RP, Linn LJ, González J, Moran MA, Bruton JA. Dimethylsulfoniopropionate and Methanethiol Are Important Precursors of Methionine and Protein-Sulfur in Marine Bacterioplankton. *Appl Environ Microbiol* 1999; **65**: 4549–4558.
13. Ferla MP, Patrick WM. Bacterial methionine biosynthesis. *Microbiology* 2014; **160**: 1571–1584.
14. Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, Webster NS, et al. Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proc Natl Acad Sci* 2012; **109**: E1878–E1887.
15. Tian R-M, Wang Y, Bougouffa S, Gao Z-M, Cai L, Bajic V, et al. Genomic analysis reveals versatile heterotrophic capacity of a potentially symbiotic sulfur-oxidizing bacterium in sponge: A versatile SOB as a potential sponge symbiont. *Environ Microbiol* 2014; **16**: 3548–3561.
16. Moreno-Pino M, Cristi A, Gillooly JF, Trefault N. Characterizing the microbiomes of Antarctic sponges: a functional metagenomic approach. *Sci Rep* 2020; **10**: 645.
17. Van Alstyne KL, Schupp P, Slattery M. The distribution of dimethylsulfoniopropionate in tropical Pacific coral reef invertebrates. *Coral Reefs* 2006; **25**: 321–327.
18. Raina J-B, Dinsdale EA, Willis BL, Bourne DG. Do the organic sulfur compounds DMSP and DMS drive coral microbial associations? *Trends Microbiol* 2010; **18**: 101–108.
19. Shaw DK, Sekar J, Ramalingam PV. Recent insights into oceanic dimethylsulfoniopropionate biosynthesis and catabolism. *Environ Microbiol* 2022; 1462-2920.16045.

20. Gao C, Fernandez VI, Lee KS, Fenizia S, Pohnert G, Seymour JR, et al. Single-cell bacterial transcription measurements reveal the importance of dimethylsulfoniopropionate (DMSP) hotspots in ocean sulfur cycling. *Nat Commun* 2020; **11**: 1942.

21. Kettle AJ, Andreae MO, Amouroux D, Andreae TW, Bates TS, Berresheim H, et al. A global database of sea surface dimethylsulfide (DMS) measurements and a procedure to predict sea surface DMS as a function of latitude, longitude, and month. *Glob Biogeochem Cycles* 1999; **13**: 399–444.

22. de Goeij JM, van Oevelen D, Vermeij MJA, Osinga R, Middelburg JJ, de Goeij AFPM, et al. Surviving in a Marine Desert: The Sponge Loop Retains Resources Within Coral Reefs. *Science* 2013; **342**: 108–110.

23. Rix L, de Goeij JM, Mueller CE, Struck U, Middelburg JJ, van Duyl FC, et al. Coral mucus fuels the sponge loop in warm- and cold-water coral reef ecosystems. *Sci Rep* 2016; **6**: 18715.

24. Coffroth MA. Cyclical mucous sheet formation on poritid corals in the San Blas Islands, Panama. *Mar Biol* 1991; **109**: 35–40.

25. Chen Y, Patel NA, Crombie A, Scrivens JH, Murrell JC. Bacterial flavin-containing monooxygenase is trimethylamine monooxygenase. *Proc Natl Acad Sci U S A* 2011; **108**: 17791–17796.

26. Lidbury I, Murrell JC, Chen Y. Trimethylamine *N* -oxide metabolism by abundant marine heterotrophic bacteria. *Proc Natl Acad Sci* 2014; **111**: 2710–2715.

27. Denger K, Laue H, Cook AM. Anaerobic taurine oxidation: a novel reaction by a nitrate-reducing Alcaligenes sp. *Microbiology* 1997; **143**: 1919–1924.

28. Ruff J, Denger K, Cook AM. Sulphoacetaldehyde acetyltransferase yields acetyl phosphate: purification from Alcaligenes defragrans and gene clusters in taurine degradation. *Biochem J* 2003; **369**: 275–285.

29. De Goeij JM, De Kluijver A, Van Duyl FC, Vacelet J, Wijffels RH, De Goeij AFPM, et al. Cell kinetics of the marine sponge Halisarca caerulea reveal rapid cell turnover and shedding. *J Exp Biol* 2009; **212**: 3892–3900.

30. Alexander BE, Liebrand K, Osinga R, van der Geest HG, Admiraal W, Cleutjens JPM, et al. Cell Turnover and Detritus Production in Marine Sponges from Tropical and Temperate Benthic Ecosystems. *PLoS ONE* 2014; **9**: e109486.

31. Moeller FU, Webster NS, Herbold CW, Behnam F, Domman D, Albertsen M, et al. Characterization of a thaumarchaeal symbiont that drives incomplete nitrification in the tropical sponge Ianthella basta. *Environ Microbiol* 2019; **21**: 3831–3854.

32. Hudspith M, Rix L, Achlatis M, Bougoure J, Guagliardo P, Clode PL, et al. Subcellular view of host–microbiome nutrient exchange in sponges: insights into the ecological success of an early metazoan–microbe symbiosis. *Microbiome* 2021; **9**: 44.

33. Rix L, Ribes M, Coma R, Jahn MT, de Goeij JM, van Oevelen D, et al. Heterotrophy in the earliest gut: a single-cell view of heterotrophic carbon and nitrogen assimilation in sponge-microbe symbioses. *ISME J* 2020; **14**: 2554–2567.

34. Rein U, Gueta R, Denger K, Ruff J, Hollemeyer K, Cook AM. Dissimilation of cysteate via 3-sulfolactate sulfo-lyase and a sulfate exporter in Paracoccus pantotrophus NKNCYSA. *Microbiology,* 2005; **151**: 737–747.

35. Weinitschke S, Hollemeyer K, Kusian B, Bowien B, Smits THM, Cook AM. Sulfoacetate Is Degraded via a Novel Pathway Involving Sulfoacetyl-CoA and Sulfoacetaldehyde in *Cupriavidus necator* H16. *J Biol Chem* 2010; **285**: 35249–35254.

36. Durham BP, Boysen AK, Carlson LT, Groussman RD, Heal KR, Cain KR, et al. Sulfonate-based networks between eukaryotic phytoplankton and heterotrophic bacteria in the surface ocean. *Nat Microbiol* 2019; **4**: 1706–1715.

37. Berger SA, Krompass D, Stamatakis A. Performance, Accuracy, and Web Server for Evolutionary Placement of Short Sequence Reads under Maximum Likelihood. *Syst Biol* 2011; **60**: 291–302.

38. Simister RL, Deines P, Botté ES, Webster NS, Taylor MW. Sponge-specific clusters revisited: a comprehensive phylogeny of sponge-associated microorganisms: Sponge-associated microorganisms. *Environ Microbiol* 2012; **14**: 517–524.

39. Luter HM, Whalan S, Webster NS. Exploring the Role of Microorganisms in the Disease-Like Syndrome Affecting the Sponge Ianthella basta. *Appl Environ Microbiol* 2010; **76**: 5736–5744.

40. Engelberts JP, Abdul Wahab MA, Maldonado M, Rix L, Marangon E, Robbins SJ, et al. Microbes from Mum: symbiont transmission in the tropical reef sponge Ianthella basta. *ISME Commun* 2022; **2**: 90.

41. Hentschel U, Usher KM, Taylor MW. Marine sponges as microbial fermenters: Marine sponges as microbial fermenters. *FEMS Microbiol Ecol* 2006; **55**: 167–177.

CHAPTER V

Metabolic reconstruction of the near complete microbiome of the model sponge *Ianthella basta*

Authors:

Joan Pamela Engelberts, Steven J. Robbins, Craig W. Herbold, **Florian U. Moeller**, Nico Jehmlich, Patrick W. Laffy, Michael Wagner, Nicole S. Webster

Published as an Open Access publication in *Environmental Microbiology* (2022; https://doi.org/10.1111/1462-2920.16302).

Received: 18 October 2022 | Accepted: 5 December 2022

DOI: 10.1111/1462-2920.16302

RESEARCH ARTICLE

ENVIRONMENTAL MICROBIOLOGY Applied Microbiology International

Metabolic reconstruction of the near complete microbiome of the model sponge *Ianthella basta*

Joan Pamela Engelberts[1] | **Steven J. Robbins**[1] | **Craig W. Herbold**[2] | **Florian U. Moeller**[2] | **Nico Jehmlich**[3] | **Patrick W. Laffy**[4] | **Michael Wagner**[2,5] | **Nicole S. Webster**[1,4,6]

[1]Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Queensland, Australia

[2]Centre for Microbiology and Environmental Systems Science, Division of Microbial Ecology, University of Vienna, Austria

[3]Department of Molecular Systems Biology, Helmholtz-Centre for Environmental Research – UFZ, Leipzig, Germany

[4]Australian Institute of Marine Science, Townsville, Queensland, Australia

[5]Center for Microbial Communities, Department of Chemistry and Bioscience, Aalborg University, Aalborg, Denmark

[6]Australian Antarctic Division, Kingston, Tasmania, Australia

Correspondence
Joan Pamela Engelberts, Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD, Australia.
Email: j.engelberts@uqconnect.edu.au

Funding information
European Research Council, Grant/Award Number: 294343; Austrian Science Fund, Grant/Award Number: Z-383-B

Abstract

Many marine sponges host highly diverse microbiomes that contribute to various aspects of host health. Although the putative function of individual groups of sponge symbionts has been increasingly described, the extreme diversity has generally precluded in-depth characterization of entire microbiomes, including identification of syntrophic partnerships. The Indo-Pacific sponge *Ianthella basta* is emerging as a model organism for symbiosis research, hosting only three dominant symbionts: a Thaumarchaeotum, a Gammaproteobacterium, and an Alphaproteobacterium and a range of other low abundance or transitory taxa. Here, we retrieved metagenome assembled genomes (MAGs) representing >90% of *I. basta*'s microbial community, facilitating the metabolic reconstruction of the sponge's near complete microbiome. Through this analysis, we identified metabolic complementarity between microbes, including vitamin sharing, described the importance of low abundance symbionts, and characterized a novel microbe–host attachment mechanism in the Alphaproteobacterium. We further identified putative viral sequences, highlighting the role viruses can play in maintaining symbioses in *I. basta* through the horizontal transfer of eukaryotic-like proteins, and complemented this data with metaproteomics to identify active metabolic pathways in bacteria, archaea, and viruses. This data provide the framework to adopt *I. basta* as a model organism for studying host–microbe interactions and provide a basis for in-depth physiological experiments.

INTRODUCTION

Sponges are ecologically important components of aquatic ecosystems around the world, providing habitat and mediating biochemical cycles by filtering thousands of litres of water each day (Bell, 2008; de Goeij et al., 2013; Weisz et al., 2008). On coral reefs, for example, sponges recycle and retain nutrients through the 'sponge-loop', in which sponges take up and transform dissolved organic matter (DOM) into biomass, which in turn is consumed by reef-dwelling detritivores to complete the loop (de Goeij et al., 2013). Marine sponges host diverse and stable communities of microorganisms that are distinct from the surrounding seawater (Schmitt et al., 2012; Taylor et al., 2013) and contribute to holobiont health and overall function. For example, sponge symbionts are actively involved in the assimilation of DOM, recycle host waste products, metabolize toxic ammonia, and are enriched in restriction-modification systems that could aid in defence against mobile genetic elements, such as viruses (Hudspith et al., 2021; Moeller et al., 2019; Pita et al., 2018; Robbins et al., 2021; Slaby et al., 2017).

Despite the importance of the sponge microbiome, its complexity hampers the in-depth characterization of each microbial taxa, with many sponge species hosting up to 3000 distinct microbial species (Webster et al., 2010). Consequently, few sponge symbionts have been described in detail and hypotheses on microbial syntrophy have remained confined to select members of the community (Bayer et al., 2020; Moitinho-Silva et al., 2017). Moreover, sponge species for which the function of all dominant microbial members has been inferred from omics data frequently lack data on symbionts present in low abundance (Gauthier et al., 2016). However, low abundance microbes can drive fundamentally different processes than high abundance microbes (Rivett & Bell, 2018). To systematically understand how the sponge holobiont functions and to allow for manipulative experiments to further unravel the mechanisms of establishment and maintenance of symbiosis, fully characterized model organisms are needed. Such models will also be extremely valuable for studying more complex animal–microbiota symbiotic relationships.

Ianthella basta is an abundant sponge throughout the Indo-Pacific and hosts only three dominant symbionts: a Thaumarchaeotum, a Gammaproteobacterium, and an Alphaproteobacterium (Luter et al., 2010; Moeller et al., 2019), and a range of other low abundance taxa, making this sponge an ideal candidate for establishing a model for sponge symbiosis. *Ianthella basta*'s microbiome is stable across different environments (Luter et al., 2010), with the dominant symbionts, which are vertically transmitted (Engelberts et al., 2022), typically comprising over 90% of the community. Previous metagenomic and physiological studies of *I. basta* have characterized the dominant Thaumarchaeotum and Gammaproteobacterium as key players within important pathways, such as ammonia oxidation and taurine metabolism, respectively (Moeller et al., 2019, 2022). Here, we complete the microbiome by recovering MAGs for the dominant Alphaproteobacterium and several low abundant microbes that are persistent across samples, which allowed us to not only describe the full genomically predictable metabolic potential of *I. basta*'s microbiome, but also to reconstruct metabolic complementarity between these microbes. We further highlight the contribution of viruses to the metabolic capability of the sponge and complement all data with metaproteomic sequencing. This study provides an unprecedented framework within which to use *I. basta* in future manipulative experiments (Engelberts et al., 2021).

EXPERIMENTAL PROCEDURES

Sample collection

The sponge *Ianthella basta* presents in two colour morphotypes, purple and yellow (Freckelton et al., 2012), of which the tissue can be designated as either thick or thin. For this study, three yellow *Ianthella basta* individuals with thin tissue were collected in November 2019 from Orpheus Island, central Great Barrier Reef (18°37′ S, 146°30′ E) between 11 and 15 m depth using scuba. Sponges were kept in aquaria for 12 h, after which they were fragmented, snap frozen, and sent to the Australian Centre for Ecogenomics (ACE), Brisbane, for metagenomic sequencing and to the Centre for Environmental Research, Helmholtz (Leipzig, Germany), for metaproteomic analysis.

To determine whether the metagenomes recovered in this study were representative of the dominant symbionts found in different morphotypes of *I. basta*, we blasted previously recovered 16S rRNA gene amplicon sequences of the three dominant symbionts of a purple, thick morphotype (Engelberts et al., 2022) to our genomes using Blast v2.9.0+ (Camacho et al., 2009). A further three yellow, thin and three yellow, thick individuals were collected from Davies Reef (S 18°49.354′, E 147°38.253′) in February 2021, using SCUBA at depths between 12 and 15 m depth. For these individuals, 16S rRNA gene amplicon sequences of the three dominant symbionts were obtained following the methods described in Engelberts et al. (2022) and again blasted to the genomes from this study (see Table S2 for 16S rRNA gene amplicon sequences).

Microbial enrichment, DNA extraction, and metagenomic sequencing

Microbial cells were separated from the host tissue for all three *I. basta* individuals (~1 to 3 g sponge wet weight per individual) according to the methods described in Botté et al. (2019), with the exception that samples were ground on ice in 1X Calcium/Magnesium-Free Sea Water (CMFSW) for 5 min to homogenize the sponge tissue. Samples were treated with DNase I (RNase-free, New England Biolabs, M0303) to remove any remaining host and extracellular DNA, after which the microbial DNA was extracted using the phenol-chloroform protocol described in Robbins et al. (2019), omitting all freeze–thaw cycles which could shear the DNA. We note that it is possible that the removal of host cells could remove intracellular symbionts, such as Chlamydiae (Engelberts et al., 2022). Quantity and quality of the extracted DNA were checked using Qubit assays and 16S/18S rRNA PCR. Libraries were prepared with the Illumina Nextera Flex library prep kit at Microba Life Sciences (Brisbane, Australia) and sequenced on the Illumina NovaSeq 6000 (2×250 bp). All samples ($n = 3$) were first sequenced at an average depth of 5.5 Gbp for differential coverage binning (Albertsen et al., 2013), followed by sequencing at a depth of 20 Gbp to increase the coverage of microorganisms present at low abundance.

Metagenome assembly, binning, and taxonomic assignment

Paired-end DNA sequencing data were demultiplexed and adapters trimmed using Illumina BaseSpace Bcl2fastq2 v2.20, accepting one mismatch in index sequences. All reads were assembled using metaSPAdes v3.14.0 (Nurk et al., 2017). To supplement this dataset, microbial metagenomic reads previously retrieved from six sequencing runs on one purple, thick *I. basta* individual (sampled in October 2011; Moeller et al., 2019) were reassembled using metaSPAdes v3.14.0 (Nurk et al., 2017) and analysed together with data from this study. Binning of all data (separately for each sponge individual) was performed with the 'recover' pipeline of Aviary (https://github.com/rhysnewell/aviary), which uses the binning algorithms MetaBAT v1 (Kang et al., 2015), MetaBAT v2 (Kang et al., 2019), MaxBin v2 (Wu et al., 2015), CONCOCT (Alneberg et al., 2014), VAMB (Nissen et al., 2021), and Rosella (https://github.com/rhysnewell/rosella). DASTool v1.1.2 (Sieber et al., 2018) was used to select for the best representative metagenome assembled genome (MAG) for each sponge individual produced from the different binning algorithms. Completeness and contamination of each MAG were determined using CheckM v1.1.3 (Parks et al., 2015) and only MAGs that were over 50% complete with less than 5% contamination were retained for downstream analyses.

GTDB-Tk v1.5.0 (Chaumeil et al., 2019), which is based on the Genome Taxonomy Database (GTDB, http://gtdb.ecogenomic.org) taxonomy Release 202, was used to assign taxonomy to each MAG. GTDB-Tk uses FASTANI (Jain et al., 2018) and pplacer (Matsen et al., 2010) to classify MAGs based on their placement in a reference tree, which is inferred using a set of 120 bacterial and 122 archaeal markers.

MAGs were dereplicated at 95% identity with Drep v2.6.2 using otherwise default parameters (Olm et al., 2017) to remove duplicate MAGs recovered from *I. basta* individuals from our study and Moeller et al. (2019). To confirm that the recovered MAGs represented the majority of *I. basta*'s microbial community, single-copy marker genes were identified in all raw reads and MAGs and the fraction of marker genes captured in the MAGs was subsequently assessed using the 'pipe' and 'appraise' functions of SingleM v0.13.2 (https://github.com/wwood/singlem). To calculate the relative abundance of each MAG in all four *I. basta* individuals (three from this study, one study from Moeller et al., 2019), MAGs were first dereplicated at 95% identity with Drep v2.6.2 (Olm et al., 2017) to avoid arbitrary mapping between representatives of highly similar genomes, and reads were mapped to these MAGs with CoverM v0.2.0 (https://github.com/wwood/CoverM, min-covered-fraction set to 0.1). The final heatmap was visualized with R v4.0.5 (R Core Team, 2013), showing reads from the deep sequencing runs on the three *I. basta* individuals from this study, reads from one *I. basta* individual from Moeller et al. (2019), and MAGs that were consistently present over the four *I. basta* individuals (i.e. present in at least three out of the four individuals; Figure 1). These MAGs were used for the metabolic reconstruction of *I. basta*'s microbiome. We further analysed a dataset originating from a 454-pyrosequencing run from Moeller et al. (2019). From this dataset, a Thaumarchaeotal and Gammaproteobacterial MAG were retrieved with ANI values of 98.90% and 97.78% compared to the genomes recovered in this study. However, due to the low sequencing depth, this data was not included in Figure 1.

Tree building

Archaeal and bacterial phylogenomic trees were inferred for each of *I. basta*'s three dominant symbionts, using the 'de_novo_wf' workflow of GTDB-Tk v1.5.0 (Chaumeil et al., 2019). This workflow calculates de novo trees using FastTree v2.1.9 (Price et al., 2010) using the WAG + GAMMA model. Each tree included all publicly available genomes from GTDB that belonged to the same phylum or class as the respective *I. basta* symbiont - i.e. all Thaumarchaeota (a total of

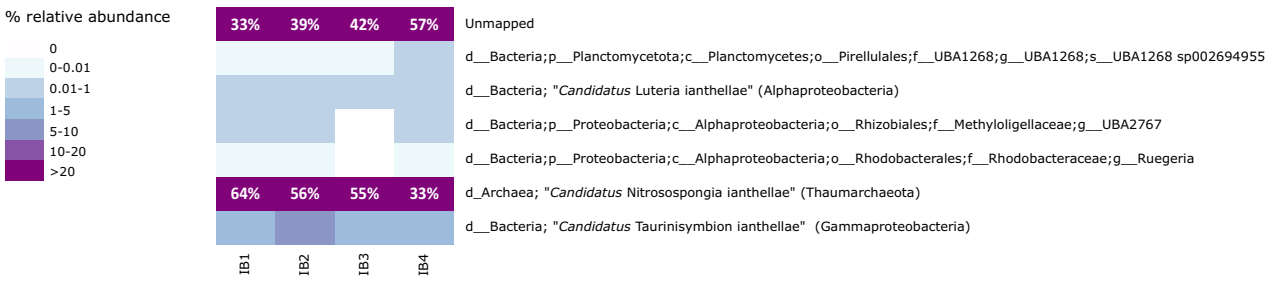


FIGURE 1 Relative abundance of MAGs present in ≥3 *I. basta* individuals after de-replication at 95% identity. For *I. basta* individual 4 the average relative abundance is shown, as this individual was sequenced six times in the study Moeller et al. (2019). 'Unmapped' represents reads that could not be mapped to any of the six MAGs. IB: *Ianthella basta* followed by the replicate number. The proposed species name (dominant symbionts; Moeller et al., 2019, 2022) or the GTDB-Tk taxonomy (low abundant symbionts) is displayed.

564 MAGs), Gammaproteobacteria (8980 MAGs), or Alphaproteobacteria (7361 MAGs). Tree files were further visualized and refined with the Interactive Tree Of Life (iTOL) v6.4.1 (Letunic & Bork, 2021) and Inkscape v0.92.4 (https://inkscape.org/) for Figures S1–S3.

The closest relatives to the *I. basta* Thaumarchaeotal MAG and Gammaproteobacterial 16S rRNA gene amplicon sequence (see Moeller et al., 2022) came from the sponge *Hexadella detritifera*, which like *I. basta*, belongs to the *Ianthellidae* family. To elucidate whether the dominant Alphaproteobacterial and Gammaproteobacterial MAGs from *I. basta* were also most closely related to microbial symbionts found in *H. detritifera*, we assembled and binned previously published raw reads from *H. detritifera* (Zhang et al., 2019). The resulting Alpha- and Gammaproteobacterial MAGs were added to the phylogenomic trees (see Supplementary Note 1).

Functional annotation of MAGs

The EnrichM v0.6.3 (https://github.com/geronimp/enrichM) 'annotate' function was used to functionally annotate MAGs with the Kyoto Encyclopaedia of Genes and Genomes (KEGG) Orthologies (KOs) and PFAM databases (Kanehisa et al., 2015; Mistry et al., 2020), as well as with Hidden Markov Models (HMMs) from the database for Automated carbohydrate-Active Enzyme Annotation (dbCAN) (Yin et al., 2012) to identify carbohydrate-active enzymes (CAZYs). For carbon fixation, vitamin production, and amino acid biosynthesis pathways, EnrichM v0.6.3 'classify' was used to identify pathways that were ≥60% complete and consequently assigned to a MAG. Notably, the KEGG module for cobalamin biosynthesis does not include the full pathway, thus the cobalamin module defined by Engelberts et al. (2020) was used. The presence of amino acid biosynthesis pathways was further validated using GapMind (Price et al., 2020). For ABC transporters to be positively assigned to a MAG, genes for both the substrate-specific component and the integral membrane protein had to be present. Secondary metabolite clusters were predicted in each MAG using AntiSMASH v5.0 (Blin et al., 2019). Proteins with predicted signal peptides for the translocation across the cytoplasmic and outer membrane were identified using SignalP v5.0 (Almagro Armenteros et al., 2019).

Identification, cross-validation, and annotation of viral sequences

Putative viral sequences were identified in all metagenomic assemblies using VirSorter v1.1 (Roux et al., 2015). To assign taxonomy, the putative viral sequences were aligned to the viral component of the RefSeq Nonredundant (Nr) database (https://www.ncbi.nlm.nih.gov/refseq/), with an E-value cut-off of 1e-05. Sequences were also blasted against previously recovered viral sequences from *I. basta* (Laffy et al., 2018) using BlastP v2.9.0 (Camacho et al., 2009) for cross-validation and to further confirm their taxonomic assignment (Table S6). Viral contigs were functionally annotated by calling open reading frames (ORFs) in the metagenomic assemblies using Prodigal v2.6.3 (Hyatt et al., 2010). ORFs were subsequently annotated with the KO and PFAM databases using EnrichM v0.6.3 (https://github.com/geronimp/enrichM).

Protein extraction and proteome analyses

Tissue from the same three *I. basta* individuals metagenomically sequenced in this study was sent to the Centre for Environmental Research, Helmholtz, for metaproteomic analysis. Sponge tissue was disrupted by bead beating (FastPrep-24, MP Biomedicals, Sanra Ana, CA, USA; 5.5 ms, 1 min, 3 cycles) followed by ultra-sonication (UP50H, Hielscher, Teltow, Germany; cycle 0.5, amplitude 60%) and centrifugation (10,000 × *g*, 10 min). The protein lysate was loaded on an SDS-gel and run for 10 min. The gel piece was cut, washed, and incubated with 25 mM 1,4 dithiothreitol (in 20 mM ammonium bicarbonate) for 1 h and 100 mM iodoacetamide (in 20 mM ammonium bicarbonate) for 30 min, and destained, dehydrated, and proteolytically cleaved overnight at 37°C with trypsin (Promega). Digested peptides were extracted and desalted using ZipTip μC18 tips (Merck Millipore, Darmstadt, Germany). Peptide lysates were re-suspended in 15 μl 0.1% formic acid and analysed by nanoliquid chromatography mass spectrometry (UltiMate 3000 RSLCnano, Dionex, Thermo Fisher Scientific). Mass spectrometric analyses of eluted peptide lysates were performed on a Q Exactive HF mass spectrometer (Thermo Fisher Scientific) coupled with a TriVersa NanoMate (Advion, Ltd., Harlow, UK). Peptide lysates were injected on a trapping column (Acclaim PepMap 100 C18, 3 μm, nanoViper, 75 μm × 2 cm, Thermo Fisher Scientific) with 5 μl/min by using 98% water/2% ACN 0.5% trifluoroacetic acid and separated on an analytical column (Acclaim PepMap 100 C18, 3 μm, nanoViper, 75 μm × 25 cm, Thermo Fisher Scientific) with a flow rate of 300 nl/min. Mobile phase was 0.1% formic acid in water (A) and 80% ACN/0.08% formic acid in water (B). Full MS spectra (350–1550 m/z) were acquired in the Orbitrap at a resolution of 120,000 with automatic gain control (AGC) target value of 3 × 106 ions.

Data resulting from LC–MS/MS measurements were analysed with the Proteome Discoverer v2.4 (Thermo Fischer Scientific) using SEQUEST

HT. Protein identification was performed using a custom protein database, which consisted of proteins identified by Prodigal v2.6.3 (Hyatt et al., 2010) in all metagenomic assemblies and MAGs from the three *I. basta* individuals from this study and the one from Moeller et al. (2019). All protein sequences were functionally annotated with the KO and PFAM databases using EnrichM v0.6.3 (https://github.com/geronimp/enrichM). To avoid arbitrary mapping, only non-redundant protein sequences were used for protein identification (clustered at an identity threshold of 100%). Searches were conducted with the following parameters: Trypsin as enzyme specificity and two missed cleavages allowed. A peptide ion tolerance of 10 ppm and an MS/MS tolerance of 0.02 Da were used. As modifications, oxidation (methionine) and carbamidomethylation (cysteine) were selected. Peptides that scored a q-value >1% based on a decoy database and with a peptide rank of 1, were considered identified. Redundant proteins were grouped in protein groups by applying the strict parsimony principle. Only the protein groups that explain at least one unique identified peptide were reported through the Top3 approach implemented in the Proteome Discoverer v2.4.

This protein dataset was supplemented with previously recovered metaproteomics data from the *I. basta* individual sampled in Moeller et al. (2019). To this end, all 49 raw protein spectra from Moeller et al. (2019) (https://www.ebi.ac.uk/pride/archive/projects/PXD012484) were searched against the custom *I. basta* protein database using the same parameters as described above. However, as the MS-analysis in Moeller et al. (2019) was performed on a LTQ Velos Orbitrap and MS/MS spectra were measured in 'unit resolution'-mode, protein reports were further filtered so that each listed protein had at least two peptides to prevent the identification of a protein based on one peptide.

RESULTS AND DISCUSSION

General MAG statistics and selection

In total, 43 microbial MAGs were recovered from four *I. basta* individuals, of which 23 originated from the three *I. basta* individuals collected in this study and 20 from an individual sampled in 2011 (Moeller et al., 2019). The recovered MAGs had an average completeness of 86% ± 13%, contamination of 1% ± 1%, and GC content of 61% ± 3% (Table S1). In contrast to the low GC percentage typically observed in seawater microbes (Giovannoni et al., 2005), a relatively high GC content (i.e. >50%) is commonly observed in sponge symbiont metagenomes (Horn et al., 2016; Moeller et al., 2019). This could be driven by a high environmental availability of nitrogen (Luo et al., 2015) derived from the concentration of nutrients (DOM and DON) and the release of taurine by the sponge host (Moeller et al., 2022; Rix et al., 2020; Robbins et al., 2021). De-replication of the MAGs at 95% identity to remove duplicate MAGs between *I. basta* individuals reduced the total number to 10, including one archaeal MAG from the phylum Thaumarchaeota representing the previously characterized '*Candidatus* Nitrosospongia ianthellae' (Moeller et al., 2019) (Thermoproteota as per GTDB Release 202) and nine bacterial MAGs, belonging to the phyla Proteobacteria (classes Alpha- and Gammaproteobacteria; three MAGs and one MAG, respectively, the gammaproteobacterial MAG represents the previously analysed '*Candidatus* Taurinisymbion ianthellae'; Moeller et al., 2022), Planctomycetota (three MAGs), Cyanobacteria (1), and Acidobacteriota (1). All genomes of the three dominant symbionts recovered from the four *I. basta* individuals had an ANI between 96.1% and 99.9%. Analysis of the relative abundance of the MAGs in each *I. basta* individual showed that MAGs belonging to the phyla Acidobacteriota, Cyanobacteria, and Planctomycetetota (two MAGs) were absent in half of the individuals. Thus, these MAGs were omitted from further analyses. The resulting six MAGs that were consistently (i.e. in ≥3 of the 4 tested individuals) present across *I. basta* individuals had a relative abundance between 0.2% ± 0.6% and 51% ± 11% (Figure 1). Based on the presence of single-copy marker genes identified by SingleM in both the raw reads and MAGs, these six symbionts represented 92% of the sponge's microbial community, allowing for the metabolic reconstruction of all major members of the *I. basta* microbiome (i.e. the near complete microbiome). The remaining 8% of the detected microbial community likely represent food bacteria or very low abundance symbionts.

Dominant symbionts fall within sponge-specific symbiont clades

To determine whether metagenomes recovered in this study were representative of the dominant symbionts found in all morphotypes of *I. basta*, we blasted previously (purple, thick morphotype) and newly (yellow, thin and thick morphotype) recovered 16S rRNA gene amplicon sequences of the three dominant symbionts to the genomes from this study (yellow, thin morphotype). All 16S rRNA gene sequences mapped with ≥99.2% identity to their respective genome, showing that highly similar symbiont communities are found in geographically and morphologically different *I. basta* individuals (Table S2). To then determine the phylogenetic placement and closest relatives of *I. basta*'s dominant symbionts, three phylogenomic trees were inferred for the Thaumarchaeotum '*Candidatus* Nitrosospongia ianthellae' (Figure S1), the

Gammaproteobacterium '*Candidatus* Taurinisymbion ianthellae' (Figure S2), and the abundant Alphaproteobacterium (Figure S3). All three dominant symbionts clustered with previously recovered microbial MAGs from sponges, allowing for the extrapolation of our findings to other sponge species and supporting the wider applicability of *I. basta* as a model species for symbiosis research. The Thaumarchaeotum ('*Ca*. Nitrosospongia ianthellae'; Moeller et al., 2019), which was the most abundant symbiont in *I. basta* (Figure 1), was most closely related to MAGs from the genus Nitrosopumilus, associated with the sponge *Hexadella dedritifera* (Zhang et al., 2019) (64.00% average amino acid identity [AAI] to GB_GCA_003724325.1; Figure S1), which, like *I. basta*, is a member of the sponge family *Ianthellidae*. '*Ca*. Nitrosospongia ianthellae' also clustered with Nitrosopumilus-related MAGs from the marine demosponge *Coelocarteria singaporensis* (Botté et al., 2019). *Ianthella basta*'s dominant Gammaproteobacterium ('*Ca*. Taurinisymbion ianthellae'; Moeller et al., 2022) was most closely related to two MAGs from the LS-SOB family which were recovered from the sponge *H. dedritifera* (79.83% AAI to Hexadella_Gamma_MAG_1 and 77.48% AAI to Hexadella_Gamma_MAG_2, see Supplementary Note 1). Together, these three genomes formed a sister group to one LS-SOB MAG from the sponge *Ircinia ramosa* (Engelberts et al., 2020), which in turn formed a sister group to six other LS-SOB MAGs from the sponges *Coscinoderma matthewsi* (Glasl et al., 2020), *Lophophysema eversa* (Tian et al., 2016), *Mycale hentscheli* (Rust et al., 2020; Storey et al., 2020), and *Ircinia ramosa* (Engelberts et al., 2020) (Figure S2). The Alphaproteobacterium (o_JABSOH01) had the lowest resolved taxonomy of the three dominant symbionts and was most closely related to an Alphaproteobacterial MAG recovered from *H. dedritifera* (71.68% AAI), which together formed a sister group to six genomes found in the deep-sea sponge *Vazella pourtalesii* (Bayer et al., 2020) (Figure S3) belonging to the same order. Compared to the six Alphaproteobacterial MAGs from the deep-sea sponge *Vazella pourtalesii*, the Alphaproteobacterial MAG in *I. basta* had highest AAI of 43.87% with GCA_014238935.1 (Figure S3), which was well below an AAI of 60%–80% that is typical for MAGs grouped at the same genus (Konstantinidis et al., 2017). Thus, we propose that this symbiont represents a new species within a new genus and tentatively named it '*Ca*. Luteria ianthellae'. The species name ianthellae refers to the host of this microorganism, the sponge *Ianthella basta*. The genus is named after Dr. Heidi Luter, who performed the foundational work on *I. basta*.

All three dominant symbionts of *I. basta* were most closely related to microbial symbionts of *H. dedritifera*. Both sponge species are members of the *Ianthellidae* family, suggesting that the three main symbionts were acquired by a common ancestor of these two sponge genera. Future research should focus on characterizing the microbial communities of other members of this sponge family to better understand the evolutionary history of the *Ianthellidae* microbiome.

General overview of metaproteomics data

In total, 11,543 expressed protein sequences (8675 non-redundant) were recovered from the four *I. basta* individuals, (where three were sequenced in this study and one in the study by Moeller et al., 2019), of which 6480 could be functionally annotated, either by the (KEGG) Orthologies (KOs) or PFAM database (Table S3). A total of 1700 proteins were captured in the genomes of *I. basta*'s six symbionts of which 1134 were functionally annotated: 631 (37%) protein sequences were assigned to the Gammaproteobacterium, 496 (29%) to the dominant Alphaproteobacterium, 354 (21%) to the Thaumarchaeotum, 124 (7%) to Ruegeria (Alphaproteobacterium), 88 (5%) to UBA1268 (Planctomycetota), and 7 (0.4%) to UBA2767 (Alphaproteobacterium; Table S3).

Metaproteogenomic analysis of *Ianthella basta's* microbiome

A metaproteogenomic analysis of *I. basta*'s complete microbiome provided unique insights into the distribution of functions across the six symbionts (Figure 2, Tables S3 and S4). For example, microbial oxidation of ammonia can prevent its accumulation to toxic levels (Zhang et al., 2014), benefitting the sponge host. Functional analysis of the gene responsible for this step, ammonia monooxygenases (*amoA*), revealed that only Thaumarchaeota are capable of ammonia oxidation in sponges (Robbins et al., 2021), which was confirmed in *I. basta* with isotope-based functional assays (Moeller et al., 2019), and further corroborated here (Figure 2). Interestingly, *I. basta* harbours no nitrite oxidizer and thus performs as a holobiont incomplete nitrification of ammonia to nitrite (Moeller et al., 2019). In addition, we show that low abundance microbes have the genomic potential to reduce nitrite and nitric oxide produced by the oxidation of ammonia. For example, one of the two low abundant Alphaproteobacteria (g_Ruegeria) encodes genes for the reduction of nitrite to dinitrogen through nitric and nitrous oxide (i.e. denitrification, catalysed by *nirS*, *norBC*, and *nosZ*), which could potentially release nitrogen from the system (Figure 2, Table S4). Notably, the first two steps of denitrification can also be performed by the Gammaproteobacterium ('*Ca*. Taurinisymbion ianthellae') through NirK and qNOR, of which the latter was further encoded in '*Ca*. Luteria ianthellae'. Alternatively, nitric oxide could be

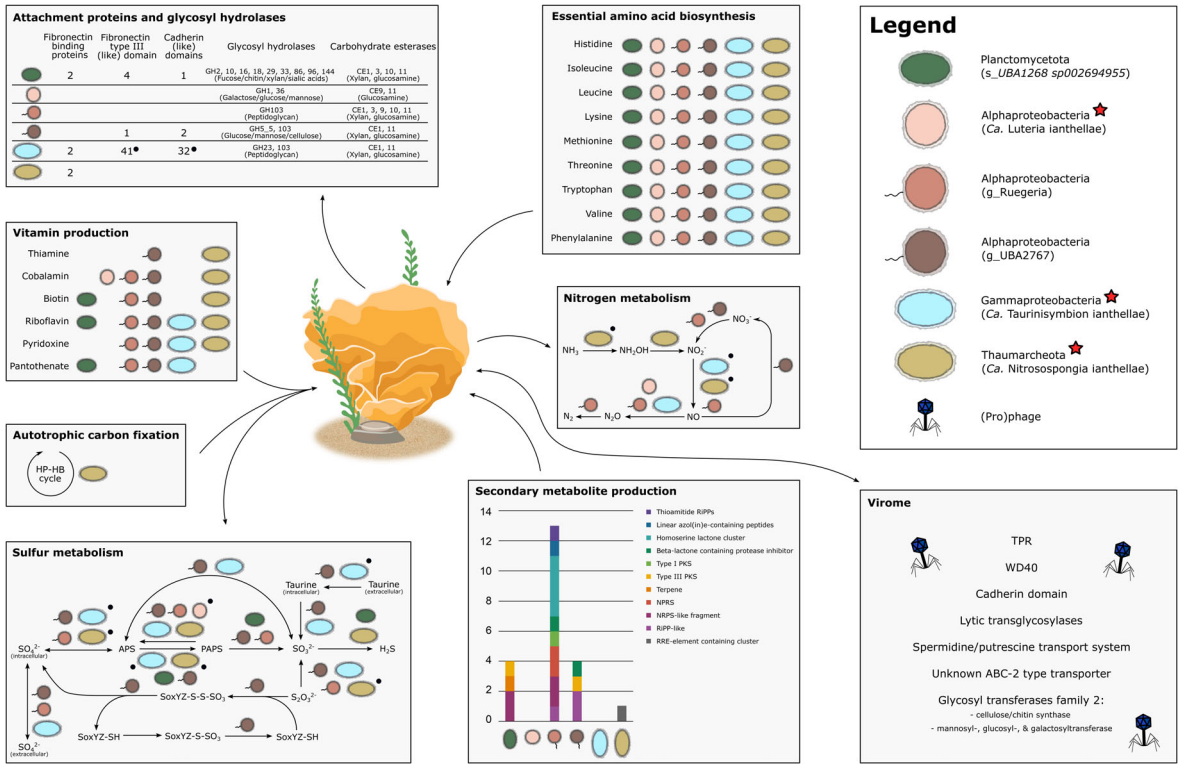


FIGURE 2 Metabolic reconstruction of *I. basta*'s microbiome and virome. Arrowheads point to the hypothesized direction of transfer. Genes expressed in the metaproteomics data are indicated by a black dot. For secondary metabolite production and vitamin and amino acid biosynthesis, proteins were either not detected (secondary metabolite production) or detected for <60% of the genes in the metabolic pathway (vitamin and amino acid biosynthesis). The legend displays the proposed species name (for the dominant symbionts, Moeller et al., 2019, 2022) or the lowest resolved taxonomy as per GTDB taxonomy Release 202, with the three dominant symbionts highlighted by a star. HP-HB cycle, hydroxypropionate-hydroxybutyrate cycle; NPRS, non-ribosomal peptide synthetase; PKS, polyketide synthase; RiPP, ribosomally synthesized and post-translationally modified peptides; RRE, RiPP recognition element; TPR, tetratricopeptide repeats

detoxified to nitrate via the other low abundance alphaproteobacterium (g_UBA2767), catalysed by nitric oxide dioxygenase (Hmp; Figure 2, Table S4). Interestingly, denitrification and nitrate production were previously suggested to occur in *I. basta* based on an isotope-based functional assays (Moeller et al., 2019), but the identity of the organisms responsible could not be confirmed. Based on our results, we suggest that the low abundance Alphaproteobacteria from the genera Ruegeria and UBA2767 are responsible for these key processes of removing and transforming nitrogen compounds from the sponge host.

Taurine is a naturally occurring compound in marine sponges, with concentrations up to 6 μmol/g wet weight measured in *I. basta* (Moeller et al., 2022). It has been shown that the gammaproteobacterium '*Candidatus* Taurinisymbion ianthellae' has the metabolic potential to use taurine for energy conservation coupled with the release of sulfate and ammonia (Moeller et al., 2022). Interestingly, not only the abundant symbiont '*Ca*. Taurinisymbion ianthellae' but also the low abundance Alphaproteobacterium (g_UBA2767) is well equipped for the use of taurine. The latter symbiont encodes the taurine transporter (*tauABC*) as well as genes for two taurine degradation pathways to convert taurine to sulfite (either via taurine dioxygenase [*tauD*] or taurine dehydrogenase [of which the large subunit *TauY* is encoded, but not the small subunit *TauX*] and sulfoacetaldehyde acetyltransferase [*xsc*]). This symbiont further encodes genes to subsequently convert sulfite into adenylyl sulfate (APS), via adenylylsulfate reductase (*aprAB*), and to reduce APS into 3′-phosphoadenylyl sulfate (PAPS) via the bifunctional enzyme CysN/CysC (*cysNC*; Figure 2, Table S4). *AprAB* can further shuttle electrons from the oxidation of sulfite to the quinone pool. PAPS is an active sulfate donor, which can be used by microorganisms to synthesize sulfated compounds. PAPS can also be converted into sulfide via sulfite and incorporated into amino acids, such as cysteine. These reactions are catalysed by phosphoadenosine phosphosulfate reductase (*cysH*) and sulfite reductase (*sir*), respectively, of which the latter was found in the genomes of the Planctomycetota (g_UBA1268) and '*Ca*. Nitrosospongia ianthellae' (Figure 2). Alternatively, APS can be converted to sulfate via sulfate adenylyltransferase (*sat*), generating ATP, thus allowing the low abundance Alphaproteobacterium (g_UBA2767) to use taurine for energy conservation.

Moreover, two of *I. basta*'s symbionts, the Alphaproteobacteria (g_UBA2767 and g_Ruegeria), are postulated to import sulfate from the surrounding seawater via sulfate permease (*sulP*). This environmentally obtained sulfate could also be reduced to PAPS for assimilation (Figure 2). However, it must be noted that sulfate permease can also export sulfate generated from taurine, which has been previously proposed to occur in '*Ca*. Taurinisymbion ianthellae' (Moeller et al., 2022). Overall, these findings suggest that sulfur metabolism in *I. basta*'s microbiome is predominantly geared towards gaining energy for the Alphaproteobacterium (g_UBA2767) and '*Ca*. Taurinisymbion ianthellae' and towards gaining cellular sulfur for the other symbionts. Notably, the Alphaproteobacterium (g_UBA2767) is predicted to also be able to obtain energy through the oxidation of thiosulfate to sulfate through sulfur-oxidizing proteins (*soxAB*, *soxXYZ*), called the SOX complex (Tian et al., 2014). This further indicates the importance of the Alphaproteobacterium (g_UBA2767) in the sulfur metabolism of *I. basta*, which was initially thought to be solely carried out by the dominant Gammaproteobacterium, which can also convert DMSP (Moeller et al., 2022).

The third low abundance symbiont, the Planctomycetota (g_UBA1268), is not predicted to be a major player in nitrogen and sulfur metabolism, but it plays an important role in catabolizing glycoconjugates and carbohydrates via glycosyl hydrolases (GHs) and carbohydrate esterases (CEs; Figure 2). Marine sponges continuously (re)cycle dissolved organic matter (DOM), a rich source of complex sugars, and a recent study showed that sponge symbionts are actively involved in utilizing this DOM (Campana et al., 2021). Similarly, free-living Planctomycetota play an important role in metabolizing complex carbohydrates (Martinez-Garcia et al., 2012; Orellana et al., 2022; Sichert et al., 2020) Here, we found that *I. basta*'s Planctomycetota (g_UBA1268) encoded nine glycosyl hydrolases, acting amongst others on fucose (GH29), chitin (GH18), xylan (GH10), and sialic acids (GH33; Figure 2), compounds often found in marine algae or the sponge itself (Hsieh & Harris, 2019; Kappelmann et al., 2019; Morya et al., 2012; Mutsenko et al., 2017). These nine GHs were absent from all other symbionts, highlighting the unique role of the Planctomycetota (g_UBA1268) in breaking down a wide range of complex carbohydrates. Interestingly, the three Alphaproteobacterial symbionts and '*Ca.* Taurinisymbion ianthellae' could import sialic acid through the sialic acid tripartite ATP-independent periplasmic transporter (*SiaPQM*). GH29 and GH33 were also expressed in the metaproteome of *I. basta*'s microbiome and have been identified previously as important hydrolase classes for many sponge symbionts and were often present in other sponge-associated Planctomycetota (Robbins et al., 2021). The carbohydrate esterases that were identified in the genomes of *I. basta*'s symbionts predominantly targeted xylan (CE1) and glycosamine (CE11), a structural component of sponges (Fernandez-Busquets & Burger, 2003; Kamke et al., 2013), while CE10, an arylesterase, was further encoded and expressed in the proteome of the Planctomycetota (g_UBA1268; Figure 2, Table S3). Arylesterases are involved in detoxifying reactive oxygen species, acting as antioxidants in humans (Asare et al., 2018), and could potentially play similar roles in marine sponges. No GHs and CEs were identified in the Thaumarchaeotum, which was consistent with previous broad-scale analyses on sponge associated Thaumarchaeota (Robbins et al., 2021). Notably, Thaumarchaeota are predominantly autotrophic and primarily obtain their carbon from fixation via the 3-hydroxypropionate/4-hydroxybutyrate cycle (Figure 2) (Burgsdorf et al., 2021; Engelberts et al., 2020; Moeller et al., 2019; Offre et al., 2013; Robbins et al., 2021; Spang et al., 2012), which likely omits the need for GHs and CEs to break down complex sugars. Furthermore, ammonia-oxidizing archaea do not have peptidoglycan cell walls, but an S-layer and as such do not require enzymes for the modification of peptidoglycans during growth.

In addition to its role in catabolizing glycoconjugates and carbohydrates, the Planctomycetota (g_UBA1268) also encoded a polyphosphate kinase (*ppk*). Polyphosphate kinase catalyses the reversible transfer of the terminal phosphate of ATP to form a long-chain polyphosphate (polyP), which can contribute to the phosphorus sequestration in *Ianthella basta* (Zhang et al., 2015). This gene was further found in the Gammaproteobacterial symbiont and the two low abundance Alphaproteobacteria (g_Ruegeria and g_UBA2767) and expressed in the metaproteome of *I. basta*.

Ianthella basta's microbiome is characterized by syntrophy

Sponge symbionts do not only interact with their host, but they also have the potential to engage in complex interactions with each other. Microbes that are unable to synthesize a particular compound themselves (auxotrophy) have to live off products excreted by other species in the community (syntrophy) in order to persist in the environment. For example, in *I. basta*, all symbionts, apart from the Alphaproteobacterium (g_UBA2767), were auxotrophic for at least one of the six essential B-vitamins, that is, thiamine (vitamin B1), riboflavin (B2), pantothenate (B5), pyridoxine (B6), biotin (B7), or cobalamin (B12). B-vitamins are essential for the central metabolism of microorganisms (Combs Jr & McClung, 2016), including the biosynthesis of branched amino acids and methionine, and microbes that cannot synthesize these vitamins themselves must acquire them from the environment. Thiamine for example, could only be biosynthesized by the

Alphaproteobacterium (g_UBA2767) and Thaumarchaeotum (Figure 2). The other symbionts could import thiamine through the thiamine transporter (*tbpA*, *thiPQ*), which was encoded and expressed in the proteomes of the three dominant symbionts and encoded in Ruegeria. All of the other essential B-vitamins (i.e. B2, B5, B6, B7, and B12) could be biosynthesized by at least four of *I. basta*'s symbionts, leaving up to two symbionts auxotrophic to the respective vitamin (Figure 2). For example, genes necessary for cobalamin production were identified in the three alphaproteobacterial symbionts and the Thaumarchaeotum (Figure 2), which was further supported by their unique ability to produce heme, a co-factor necessary for cobalamin production. In contrast, the Gammaproteobacterium and Planctomycetota (g_UBA1268), cobalamin auxotrophs, require an external source of cobalamin to persist in the environment (Figure 2).

Microbial mobility and attachment

Symbionts must be able to avoid phagocytosis by host cells to maintain stable symbioses and it has been hypothesized that eukaryotic-like proteins (ELPs), such as ankyrin (ARP), WD40, NHL, leucine-rich (LRR), tetratricopeptide (TPR), and HEAT repeat families play an important role in this process (Jahn et al., 2019; Nguyen et al., 2014; Reynolds & Thomas, 2016). For example, ankyrin-domain-containing-proteins encoded by phages could modulate the immune response of a eukaryotic host against bacteria (Jahn et al., 2019). All six symbionts in *I. basta* encoded at least one type of ELP and all ELPs were expressed in *I. basta*'s microbiome (Tables S3 and S4). However, the distribution of ELPs varied between symbionts, where the Alphaproteobacterium (g_UBA2767) encoded ARP, WD40, LRR, NHL, TPR, and HEAT repeat families (50 copies in total), displaying the most diverse array of ELPs, the dominant Alphaproteobacterium ('*Ca*. Luteria ianthellae') only encoded TPR repeat families (19 copies). The factors driving these differences and whether each ELP serves a different function in maintaining stable symbiosis remains to be elucidated.

Persistence of microbes within the sponge can also be facilitated by direct attachment to host tissue or by attachment to other microbes that are attached to the host tissue. Proteins that mediate cell–cell adhesion and biofilm formation, such as cadherin domains and fibronectin-binding proteins, were found to be enriched in sponge microbiomes (Robbins et al., 2021) and may play similarly important roles in maintaining symbiosis in *I. basta*. Cadherin domains can bind directly to cell surfaces or to cellulose, xylan, and related compounds, mediating not only attachment to the host tissue, but also the degradation of complex carbohydrates by microbes (Fraiberg et al., 2010; Fraiberg et al., 2011). Fibronectin-binding proteins are proteins that bind to fibronectin III domains, which can be found in both eukaryotes and prokaryotes (Hymes & Klaenhammer, 2016; Schwarz-Linek et al., 2003). Cadherin domains were encoded in the Planctomycetota (g_UBA1268) and Alphaproteobacterium (g_UBA2767) and encoded and expressed in the Gammaproteobacterium proteome (Figure 2, Tables S3 and S4). Interestingly, these three genomes also contained genes for GH10 and CE1 that degrade xylan, a process that could be facilitated by attachment to the carbohydrate itself through cadherin domains. Genes for fibronectin-binding proteins were identified in the Planctomycetota (g_UBA1268), the Gammaproteobacterium, and the Thaumarchaeotum. Fibronectin III domains were encoded in the Planctomycetota (g_UBA1268) and encoded and expressed in the Gammaproteobacterium, allowing for biofilm formation. No fibronectin-binding proteins were found in the two low abundance Alphaproteobacteria (g_UBA2767 and g_Ruegeria). Interestingly, these microbes encoded a flagellum and expressed the flagellar protein (fliL; Figure 2, Tables S3 and S4), suggesting these symbionts are motile. In addition to motility, flagella can support adhesion (Kimkes & Heinemann, 2020), where fliL could play an important role in sensing the presence of a surface (Lee & Belas, 2015). Thus, the two low abundance Alphaproteobacteria may use their flagella as an alternative mechanism to maintain stable symbiosis.

The dominant Alphaproteobacterium ('*Ca*. Luteria ianthellae') lacked attachment proteins, suggesting it uses alternative mechanisms to persist within the sponge tissue. This genome instead encoded and expressed secretion system III (Figure S4, Tables S3 and S4), which can be used by microorganisms to attach to their host and inject effector proteins to modulate its metabolism (Costa et al., 2015). These proteins often contain signal peptides that mark them for secretion. In the dominant Alphaproteobacterium, signal peptides were identified on TPR repeat proteins and ubiquitin-like proteins (i.e. proteins that can modify other proteins after translation). As TPRs and ubiquitin-like proteins have previously been linked to symbiont establishment within the host (Robbins et al., 2021; Thomas et al., 2010; Zhou & Zhu, 2015), injection of these proteins could aid the persistence of the Alphaproteobacterium within the sponge tissue by altering *I. basta*'s metabolism (Figure S4). Additionally, secretion system III has been found to facilitate direct attachment of pathogenic bacteria to host tissue through the insertion of the so-called virulence factor translocated intimin receptor (Tir) into the host membrane (Franzin & Sircili, 2015; Kenny et al., 1997). Once Tir is injected into the host's membrane through secretion system III, it can bind to an inverse autotransporter located in the membrane of the bacterium (Kenny et al., 1997). The dominant Alphaproteobacterium also encoded and expressed an inverse autotransporter, which would

allow the symbiont to directly anchor itself to the host tissue if Tir is translocated by secretion system III (Figure S4).

To investigate whether the ability to attach to the host tissue through Tir and the inverse autotransporter was present in other sponge symbionts, we searched all available sponge symbiont genomes for secretion system III and the inverse autotransporter. Both proteins were found in three Alphaproteobacterial MAGs (GCA_014239005.1, SAMN15855071, and SAMN15855069) (Bayer et al., 2020; Robbins et al., 2021), all of which one belonged to the same order as '*Ca*. Luteria ianthellae' (o_JABSOH01), and in two MAGs belonging to the Endozoicomonadaceae family (GCA_002238585.1 and SAMN15854979) (Robbins et al., 2021; Slaby et al., 2017), suggesting this mechanism aids attachment in a variety of sponge symbionts. The ability to anchor to host tissue would indicate a tight symbiosis and consistent with such a highly dependent lifestyle, the dominant Alphaproteobacterium in *I. basta* lacked a multitude of metabolic pathways, such as the pentose phosphate cycle, vitamin production (apart from cobalamin), fatty acid B-oxidation, aromatic ring degradation, nitrogen metabolism, and dTDP-L-rhamnose biosynthesis (Figure S4). It further encoded and expressed a wide range of transporters used to import essential nutrients, including amino acid transporters and sugar transporters (Figure S4). Lastly, '*Ca*. Luteria ianthellae' encoded several genes involved in DMSP metabolism, which can be a source of reduced carbon and sulfur for marine bacteria. DMSP metabolism was also identified in '*Ca*. Taurinisymbion ianthellae' (Moeller et al., 2022), suggesting that DMSP, in addition to taurine, may be an important substrate for *I. basta* symbionts.

Low abundance microbes are sources of secondary metabolite clusters

Marine sponges are rich sources of bioactive and economically important compounds that are frequently produced by their microbial symbionts and potentially aid in the chemical defence of the holobiont (Rust et al., 2020; Tianero et al., 2019; Wilson et al., 2014). *I. basta* is known to produce bastadins (Kazlauskas et al., 1981; Kunze et al., 2013; Pordesimo & Schmitz, 1990), a biotechnologically relevant secondary metabolite that can act as an antifouling agent and has cytotoxic activity against human tumour cells (Bayer et al., 2011; Greve et al., 2008). Here, we searched the genomes of the six symbionts of *I. basta* for the ability to produce bioactive compounds and found a total of 22 biosynthetic gene clusters (BGCs), of which 21 were encoded in genomes from the low abundance members of the community. (Figure 2, Table S5). Notably, genome size correlated with the number of encoded BGCs, with Ruegeria having the largest genome and most BGCs (4.33 kb, 13 BGCs), followed by UBA2767 (4.30 kb, 4 BGCs), UBA1267 (3.26 kb, 4 BGCs), '*Ca*. Taurinisymbion ianthellae' (2.08 kb, no BGCs), '*Ca*. Luteria ianthellae' (2.01 kb, no BGCs), and '*Ca*. Nitrosospongia ianthellae' (1.78 kb, 1 BGC). The most abundant BGCs were non-ribosomal peptide synthases (NRPS/NRPS-like, 6 clusters), hserlactones (4 clusters), and type I and III polyketide synthases (T1PKS and T3PKS, 3 clusters). Only five BGCs showed similarity to previously identified clusters, frequently with similarity scores below 30%, highlighting the novelty of the BGCs in *I. basta* (Table S5). One BGC, an NRPS-like cluster encoded in the genome of the Planctomycetota (g_UBA1267), showed 100% similarity to a previously identified BGC thought to produce 1-heptadecene. Interestingly, 1-n-heptadecane can be used by denitrifying bacteria to grow anaerobically (Spormann & Widdel, 2000), but whether this is the case for Planctomycetota or whether 1-heptadecene is used by the other symbionts as an energy source remains to be determined. Although the novelty of *I. basta*'s BGCs precluded further identification of the metabolites produced, the identification of BGCs is the first step towards the discovery of new bioactive compounds. Thus, we suggest future research focus on identifying the activity of these BGCs.

Ianthella basta's virome extends the metabolic capability of the holobiont

Marine sponges host distinct viral communities that shape their microbiome, alter host–microbe interactions (Jahn et al., 2019; Pascelli et al., 2020), and directly contribute to their host's (eukaryotic or prokaryotic) metabolism through auxiliary metabolic genes (AMGs). Viruses can further mediate horizontal gene transfer between members of the holobiont. Ankyrin repeat proteins, for example, were suggested to be horizontally transferred by viruses as these proteins are widespread in sponge symbionts from different sponge species (Reynolds & Thomas, 2016) and enriched in the sponge virome (Pascelli et al., 2020). Similarly, any of the previously described functions/proteins in *I. basta*'s microbiome can theoretically be horizontally transferred by viruses or have a viral origin.

To identify putative viral sequences in *I. basta*, VirSorter was run on all metagenomic assemblies. A total of 276 viral contigs were identified, of which 250 were phage and 25 prophage (Table S6). Of the 250 contigs that were assigned as phage, 25 were within category 1 ('most confident' predictions), 157 in Category 2 ('likely' predictions), and 68 in Category 3 ('possible' predictions). For the prophages, 4 were in Category 2 and 21 in Category 3 (Table S6). Of all prophages, nine were captured in bacterial MAGs and one in the

Thaumarchaeotal MAG, which encoded a Type IV toxin–antitoxin system. Interestingly, toxin–antitoxin systems were previously found to be enriched in sponge associated Thaumarchaeota (Haber et al., 2021; Moeller et al., 2019). Taxonomy was subsequently assigned to all viral contigs by blasting them against the viral RefSeq Nr database and their viral origin and taxonomy was further cross-validated by blasting them against previously recovered viral sequences from *I. basta*. The most diverse viral taxa in *I. basta* were dsDNA viruses from the order *Caudovirales*, including the families *Myoviridae*, *Podoviridae*, and *Siphoviridae* and ssDNA viruses from the order *Microviridae* (Table S6). Interestingly, we identified a *Lavidaviridae* in *I. basta*'s virome, which are known virophages that can coinfect eukaryotes with giant dsDNA viruses of the family *Mimiviridae* (Fischer, 2020). Although no contigs were assigned to the family *Mimiviridae* in this study, *Mimiviridae* were previously identified in *I. basta* and suggested to infect sponge amoeba-like cells (Pascelli et al., 2020).

Viral contigs were functionally annotated to investigate the role of the virome in the overall metabolic capability of *I. basta*. A total of 854 genes were identified across all viral contigs, of which 89% could be functionally annotated (Table S6). Most annotated genes were replication related genes, such as transposable elements and viral DNA replication genes (Table S6). However, putative auxiliary metabolic genes (AMGs) were also detected (Figure 2, Table S6). For example, eukaryotic-like proteins, such as TPR repeat proteins and WD40 repeat proteins, were found on seven viral contigs, of which two were captured in the MAGs of the Thaumarchaeotum (prophage, Category 3) and Alphaproteobacterium (g_Ruegeria; prophage, Category 2). Interestingly, TPR repeat proteins were previously suggested to be laterally transferred between sponge symbionts (Robbins et al., 2021), which could be mediated by viruses. Similarly, attachment proteins, including cadherins, are thought to be laterally transferred between sponge symbionts and were identified on the viral contigs in *I. basta* (Figure 2, Table S6). *I. basta*'s virome also encoded various transporters, such as the spermidine/putrescine transport system (prophage Category 2, binned in the Alphaproteobacterium from the genus Ruegeria). Spermidine and putrescine are polyamines that play essential roles in cell differentiation and proliferation. Thus, the virus-encoded spermidine/putrescine transport system could further aid in the overall transport of vital compounds in *I. basta*'s microbiome.

CONCLUSION AND OUTLOOK

To fully understand the functioning of the sponge holobiont and establish a model species for future experiments, we present a metaproteogenomic analysis of >90% of *I. basta*'s microbiome and virome, providing key insights into the unique functional role of each symbiont and their metabolic interplay. We show that low abundance microbes drive fundamentally different processes compared with the dominant symbionts (e.g. complex carbohydrate degradation and vitamin production) and are repositories of key functions involved in nitrogen cycling and secondary metabolite production. We further highlight the auxotrophic nature of sponge symbionts to vitamin production and their dependency on other microbes in the environment. *Ianthella basta*'s symbionts were also predicted to use a variety of mechanisms to avoid host phagocytosis and to attach to the host tissue. These mechanisms, such as the presence of secretion system III, could be further investigated using cryo-EM. Lastly, we characterized *I. basta*'s virome by identifying putative viral contigs in its microbiome, cross-validating each sequence against previously recovered viral sequences from the sponge. Functional characterization of *I. basta*'s virome suggested a role for viruses in the lateral gene transfer of ELPs and attachment proteins and their ability to extend the metabolic capability of the holobiont through transporters. Taken together, this metaproteogenomic data provide the metabolic framework to adopt *I. basta* as a model organism for studying host–microbe interactions, particularly the establishment, development, and maintenance of sponge symbionts.

AUTHOR CONTRIBUTIONS

Pamela Engelberts: Conceptualization (lead); data curation (lead); formal analysis (lead); investigation (lead); methodology (lead); validation (lead); visualization (lead); writing – original draft (lead); writing – review and editing (equal). **Steven Robbins:** Data curation (supporting); formal analysis (supporting); investigation (supporting); methodology (supporting); supervision (lead); validation (supporting); writing – review and editing (supporting). **Craig W Herbold:** Formal analysis (equal); investigation (equal); methodology (equal); writing – review and editing (supporting). **Florian Moeller:** Investigation (supporting); validation (supporting). **Nico Jehmlich:** Data curation (supporting); formal analysis (supporting); investigation (equal); methodology (supporting); writing – review and editing (supporting). **Patrick Laffy:** Formal analysis (supporting); investigation (supporting); writing – review and editing (supporting). **Michael Wagner:** Resources (equal); supervision (supporting); writing – review and editing (equal). **Nicole Webster:** Resources (equal); supervision (supporting); writing – review and editing (supporting).

ACKNOWLEDGEMENTS

The authors thank Dr. Kenneth Wasmund, Dr. Bettina Glasl, and Dr. Paul Evans for their assistance in data

analysis and interpretation. Michael Wagner thanks the European Research Council (ERC Advanced Grant NITRICARE 294343) and the Austrian Science Fund FWF (Wittgenstein Award Z-383-B) for financial support. Open access publishing facilitated by The University of Queensland, as part of the Wiley - The University of Queensland agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Raw reads and MAGs are available at NCBI under Bioproject ID PRJNA807825. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2022) partner repository with the dataset identifier PXD032278.

ETHICAL STATEMENT

All prevailing scientific ethical practices have been respected.

ORCID

Joan Pamela Engelberts https://orcid.org/0000-0002-1185-3741

REFERENCES

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W. & Nielsen, P.H. (2013) Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnology*, 31(6), 533–538.

Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S. et al. (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, 37(4), 420–423.

Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z. et al. (2014) Binning metagenomic contigs by coverage and composition. *Nature Methods*, 11(11), 1144–1146.

Asare, G.A., Andam, S.E., Asare-Anane, H., Ammanquah, S., Anang-Quartey, Y., Afriyie, D.K. et al. (2018) Lipid associated antioxidants: arylesterase and paraoxonase-1 in benign prostatic hyperplasia treatment-naïve patients. *Prostate International*, 6(1), 36–40.

Bayer, K., Busch, K., Kenchington, E., Beazley, L., Franzenburg, S., Michels, J. et al. (2020) Microbial strategies for survival in the glass sponge Vazella pourtalesii. *Msystems*, 5(4), e00473–e00420.

Bayer, M., Hellio, C., Maréchal, J.P., Frank, W., Lin, W., Weber, H. et al. (2011) Antifouling bastadin congeners target mussel phenoloxidase and complex copper (II) ions. *Marine Biotechnology*, 13(6), 1148–1158.

Bell, J.J. (2008) The functional roles of marine sponges. *Estuarine, Coastal and Shelf Science*, 79(3), 341–353.

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y. et al. (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research*, 47(W1), W81–W87.

Botté, E., Nielsen, S., Abdul Wahab, M.A., Webster, J., Robbins, S., Thomas, T. et al. (2019) Changes in the metabolic potential of the sponge microbiome under ocean acidification. *Nature Communications*, 10(1), 1–10.

Burgsdorf, I., Sizikov, S., Squatrito, V., Britstein, M., Slaby, B.M., Cerrano, C. et al. (2021) Lineage-specific energy and carbon metabolism of sponge symbionts and contributions to the host carbon pool. *The ISME Journal*, 16(4), 1163–1175.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, 10(1), 421.

Campana, S., Busch, K., Hentschel, U., Muyzer, G. & de Goeij, J.M. (2021) DNA-stable isotope probing (DNA-SIP) identifies marine sponge-associated bacteria actively utilizing dissolved organic matter (DOM). *Environmental Microbiology*, 23, 4489–4504.

Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P. & Parks, D.H. (2019) GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics*, 36(6), 1925–1927.

Combs, G.F., Jr. & McClung, J.P. (2016) *The vitamins: fundamental aspects in nutrition and health*, 5th edition. San Diego: Academic Press Inc.

Costa, T.R., Felisberto-Rodrigues, C., Meir, A., Prevost, M.S., Redzej, A., Trokter, M. et al. (2015) Secretion systems in gram-negative bacteria: structural and mechanistic insights. *Nature Reviews. Microbiology*, 13(6), 343–359.

de Goeij, J.M., van Oevelen, D., Vermeij, M.J.A., Osinga, R., Middelburg, J.J., de Goeij, A.F.P.M. et al. (2013) Surviving in a marine desert: the sponge loop retains resources within coral reefs. *Science*, 342(6154), 108–110.

Engelberts, J.P., Abdul, W.M.A., Manuel, M., Laura, R., Emma, M., Robbins Steven, J. et al. (2022) Microbes from mum: symbiont transmission in the tropical reef sponge *Ianthella basta*. *ISME Communications,* 2(1), 1–12.

Engelberts, J.P., Robbins, S.J., Damjanovic, K. & Webster, N.S. (2021) Integrating novel tools to elucidate the metabolic basis of microbial symbiosis in reef holobionts. *Marine Biology*, 168(12), 175.

Engelberts, J.P., Robbins, S.J., de Goeij, J.M., Aranda, M., Bell, S.C. & Webster, N.S. (2020) Characterization of a sponge microbiome using an integrative genome-centric approach. *The ISME Journal*, 15(5), 1100–1110.

Fernandez-Busquets, X. & Burger, M. (2003) Circular proteoglycans from sponges: first members of the spongican family. *Cellular and Molecular Life Sciences*, 60(1), 88–112.

Fischer, M.G. (2020) The virophage family lavidaviridae. *Current Issues in Molecular Biology*, 40(1), 1–24.

Fraiberg, M., Borovok, I., Bayer, E.A., Weiner, R.M. & Lamed, R. (2011) Cadherin domains in the polysaccharide-degrading marine bacterium *Saccharophagus degradans* 2-40 are carbohydrate-binding modules. *Journal of Bacteriology*, 193(1), 283–285.

Fraiberg, M., Borovok, I., Weiner, R.M. & Lamed, R. (2010) Discovery and characterization of cadherin domains in *Saccharophagus degradans* 2-40. *Journal of Bacteriology*, 192(4), 1066–1074.

Franzin, F.M. & Sircili, M.P. (2015) Locus of enterocyte effacement: a pathogenicity Island involved in the virulence of enteropathogenic and enterohemorragic *Escherichia coli* subjected to a complex network of gene regulation. *BioMed Research International*, 2015, 534738.

Freckelton, M.L., Luter, H.M., Andreakis, N., Webster, N.S. & Motti, C.A. (2012) Qualitative variation in colour morphotypes of *Ianthella basta* (porifera: Verongida). *Hydrobiologia*, 687(1), 191–203.

Gauthier, M.-E.A., Watson, J.R. & Degnan, S.M. (2016) Draft genomes shed light on the dual bacterial symbiosis that dominates the microbiome of the coral reef sponge *Amphimedon queenslandica*. *Frontiers in Marine Science*, 3, 196.

Giovannoni, S.J., Tripp, H.J., Givan, S., Podar, M., Vergin, K.L., Baptista, D. et al. (2005) Genome streamlining in a cosmopolitan oceanic bacterium. *Science*, 309(5738), 1242–1245.

Glasl, B., Robbins, S., Frade, P.R., Marangon, E., Laffy, P.W., Bourne, D.G. et al. (2020) Comparative genome-centric analysis reveals seasonal variation in the function of coral reef microbiomes. *The ISME Journal*, 14(6), 1435–1450.

Greve, H., Kehraus, S., Krick, A., Kelter, G., Maier, A., Fiebig, H.H. et al. (2008) Cytotoxic bastadin 24 from the Australian sponge *Ianthella quadrangulata*. *Journal of Natural Products*, 71(3), 309–312.

Haber, M., Burgsdorf, I., Handley, K.M., Rubin-Blum, M. & Steindler, L. (2021) Genomic insights into the lifestyles of Thaumarchaeota inside sponges. *Frontiers in Microbiology*, 11, 622824.

Horn, H., Slaby, B.M., Jahn, M.T., Bayer, K., Moitinho-Silva, L., Förster, F. et al. (2016) An enrichment of CRISPR and other defense-related features in marine sponge-associated microbial metagenomes. *Frontiers in Microbiology*, 7, 7.

Hsieh, Y.S.Y. & Harris, P.J. (2019) Xylans of red and green algae: what is known about their structures and how they are synthesised? *Polymers*, 11(2), 354.

Hudspith, M., Rix, L., Achlatis, M., Bougoure, J., Guagliardo, P., Clode, P.L. et al. (2021) Subcellular view of host–microbiome nutrient exchange in sponges: insights into the ecological success of an early metazoan–microbe symbiosis. *Microbiome*, 9(1), 1–15.

Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W. & Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11(1), 1–11.

Hymes, J.P. & Klaenhammer, T.R. (2016) Stuck in the middle: fibronectin-binding proteins in gram-positive bacteria. *Frontiers in Microbiology*, 7, 1504.

Jahn, M.T., Arkhipova, K., Markert, S.M., Stigloher, C., Lachnit, T., Pita, L. et al. (2019) A phage protein aids bacterial symbionts in eukaryote immune evasion. *Cell Host & Microbe*, 26(4), 542–550 e5.

Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T. & Aluru, S. (2018) High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, 9(1), 5114.

Kamke, J., Sczyrba, A., Ivanova, N., Schwientek, P., Rinke, C., Mavromatis, K. et al. (2013) Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. *The ISME Journal*, 7(12), 2287–2300.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. (2015) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44(D1), D457–D462.

Kang, D.D., Froula, J., Egan, R. & Wang, Z. (2015) MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ*, 3, e1165.

Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H. et al. (2019) MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, 7, e7359.

Kappelmann, L., Krüger, K., Hehemann, J.H., Harder, J., Markert, S., Unfried, F. et al. (2019) Polysaccharide utilization loci of North Sea Flavobacteriia as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. *The ISME Journal*, 13(1), 76–91.

Kazlauskas, R., Lidgard, R.O., Murphy, P.T., Wells, R.J. & Blount, J. F. (1981) Brominated tyrosine-derived metabolites from the sponge *Ianthella basta*. *Australian Journal of Chemistry*, 34(4), 765–786.

Kenny, B., DeVinney, R., Stein, M., Reinscheid, D.J., Frey, E.A. & Finlay, B.B. (1997) Enteropathogenic *E. coli* (EPEC) transfers its receptor for intimate adherence into mammalian cells. *Cell*, 91(4), 511–520.

Kimkes, T.E. & Heinemann, M. (2020) How bacteria recognise and respond to surface contact. *FEMS Microbiology Reviews*, 44(1), 106–122.

Konstantinidis, K.T., Rosselló-Móra, R. & Amann, R. (2017) Uncultivated microbes in need of their own taxonomy. *The ISME Journal*, 11(11), 2399–2406.

Kunze, K., Niemann, H., Ueberlein, S., Schulze, R., Ehrlich, H., Brunner, E. et al. (2013) Brominated skeletal components of the marine demosponges, Aplysina cavernicola and *Ianthella basta*: analytical and biochemical investigations. *Marine Drugs*, 11(4), 1271–1287.

Laffy, P.W., Wood-Charlson, E.M., Turaev, D., Jutz, S., Pascelli, C., Botté, E.S. et al. (2018) Reef invertebrate viromics: diversity, host specificity and functional capacity. *Environmental Microbiology*, 20(6), 2125–2141.

Lee, Y.-Y. & Belas, R. (2015) Loss of FliL alters Proteus mirabilis surface sensing and temperature-dependent swarming. *Journal of Bacteriology*, 197(1), 159–173.

Letunic, I. & Bork, P. (2021) Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, 49(W1), W293–W296.

Luo, H., Thompson, L.R., Stingl, U. & Hughes, A.L. (2015) Selection maintains low genomic GC content in marine SAR11 lineages. *Molecular Biology and Evolution*, 32(10), 2738–2748.

Luter, H.M., Whalan, S. & Webster, N.S. (2010) Exploring the role of microorganisms in the disease-like syndrome affecting the sponge *Ianthella basta*. *Applied and Environmental Microbiology*, 76(17), 5736–5744.

Martinez-Garcia, M., Brazel, D.M., Swan, B.K., Arnosti, C., Chain, P. S.G., Reitenga, K.G. et al. (2012) Capturing single cell genomes of active polysaccharide degraders: An unexpected contribution of Verrucomicrobia. *PLoS One*, 7(4), e35314.

Matsen, F.A., Kodner, R.B. & Armbrust, E.V. (2010) Pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, 11(1), 538.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L. et al. (2020) Pfam: the protein families database in 2021. *Nucleic Acids Research*, 49(D1), D412–D419.

Moeller, F.U., Herbold, C.W., Schintlmeister, A., Mooshammer, M., Motti, C., Behnam, F. et al. (2022) Taurine as a key intermediate for host-symbiont interaction in the tropical sponge *Ianthella basta*. *bioRxiv*.

Moeller, F.U., Webster, N.S., Herbold, C.W., Behnam, F., Domman, D., Albertsen, M. et al. (2019) Characterization of a thaumarchaeal symbiont that drives incomplete nitrification in the tropical sponge *Ianthella basta*. *Environmental Microbiology*, 21(10), 527234.

Moitinho-Silva, L., Díez-Vives, C., Batani, G., Esteves, A.I.S., Jahn, M.T. & Thomas, T. (2017) Integrated metabolism in sponge–microbe symbiosis revealed by genome-centered metatranscriptomics. *The ISME Journal*, 11(7), 1651–1666.

Morya, V.K., Kim, J. & Kim, E.-K. (2012) Algal fucoidan: structural and size-dependent bioactivities and their perspectives. *Applied Microbiology and Biotechnology*, 93(1), 71–82.

Mutsenko, V.V., Gryshkov, O., Lauterboeck, L., Rogulska, O., Tarusin, D.N., Bazhenov, V.V. et al. (2017) Novel chitin scaffolds derived from marine sponge Ianthella basta for tissue engineering approaches based on human mesenchymal stromal cells: biocompatibility and cryopreservation. *International Journal of Biological Macromolecules*, 104, 1955–1965.

Nguyen, M.T., Liu, M. & Thomas, T. (2014) Ankyrin-repeat proteins from sponge symbionts modulate amoebal phagocytosis. *Molecular Ecology*, 23(6), 1635–1645.

Nissen, J.N., Johansen, J., Allesøe, R.L., Sønderby, C.K., Armenteros, J.J.A., Grønbech, C.H. et al. (2021) Improved metagenome binning and assembly using deep variational autoencoders. *Nature Biotechnology*, 39(5), 555–560.

Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P.A. (2017) metaSPAdes: a new versatile metagenomic assembler. *Genome Research*, 27(5), 824–834.

Offre, P., Spang, A. & Schleper, C. (2013) Archaea in biogeochemical cycles. *Annual Review of Microbiology*, 67(1), 437–457.

Olm, M.R., Brown, C.T., Brooks, B. & Banfield, J.F. (2017) dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal*, 11(12), 2864–2868.

Orellana, L.H., Francis, T.B., Ferraro, M., Hehemann, J.H., Fuchs, B. M. & Amann, R.I. (2022) Verrucomicrobiota are specialist consumers of sulfated methyl pentoses during diatom blooms. *The ISME Journal*, 16(3), 630–641.

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(7), 1043–1055.

Pascelli, C., Laffy, P.W., Botté, E., Kupresanin, M., Rattei, T., Lurgi, M. et al. (2020) Viral ecogenomics across the Porifera. *Microbiome*, 8(1), 144.

Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S. et al. (2022) The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acid Research*, 50(D1), D543–D552.

Pita, L., Rix, L., Slaby, B.M., Franke, A. & Hentschel, U. (2018) The sponge holobiont in a changing ocean: from microbes to ecosystems. *Microbiome*, 6(1), 46.

Pordesimo, E.O. & Schmitz, F.J. (1990) New bastadins from the sponge *Ianthella basta*. *The Journal of Organic Chemistry*, 55(15), 4704–4709.

Price, M.N., Dehal, P.S. & Arkin, A.P. (2010) FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One*, 5(3), e9490.

Price, M.N., Deutschbauer, A.M. & Arkin, A.P. (2020) GapMind: automated annotation of amino acid biosynthesis. *mSystems*, 5(3), e00291–e00220.

R Core Team. (2013) *R: a language and environment for statistical computing*.

Reynolds, D. & Thomas, T. (2016) Evolution and function of eukaryotic-like proteins from sponge symbionts. *Molecular Ecology*, 25(20), 5242–5253.

Rivett, D.W. & Bell, T. (2018) Abundance determines the functional role of bacterial phylotypes in complex communities. *Nature Microbiology*, 3(7), 767–772.

Rix, L., Ribes, M., Coma, R., Jahn, M.T., de Goeij, J.M., van Oevelen, D. et al. (2020) Heterotrophy in the earliest gut: a single-cell view of heterotrophic carbon and nitrogen assimilation in sponge-microbe symbioses. *The ISME Journal*, 14(10), 1–14.

Robbins, S., Singleton, C.M., Chan, C.X., Messer, L.F., Geers, A.U., Ying, H. et al. (2019) A genomic view of the reef-building coral Porites lutea and its microbial symbionts. *Nature Microbiology*, 4(12), 2090–2100.

Robbins, S. et al. (2021) A genomic view of the microbiome of coral reef demosponges. *The ISME Journal*, 15(6), 1–14.

Roux, S., Enault, F., Hurwitz, B.L. & Sullivan, M.B. (2015) VirSorter: mining viral signal from microbial genomic data. *PeerJ*, 3, e985.

Rust, M., Helfrich, E.J.N., Freeman, M.F., Nanudorn, P., Field, C.M., Rückert, C. et al. (2020) A multiproducer microbiome generates chemical diversity in the marine sponge *Mycale hentscheli*. *Proceedings of the National Academy of Sciences*, 117(17), 9508–9518.

Schmitt, S., Tsai, P., Bell, J., Fromont, J., Ilan, M., Lindquist, N. et al. (2012) Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. *The ISME Journal*, 6(3), 564–576.

Schwarz-Linek, U., Werner, J.M., Pickford, A.R., Gurusiddappa, S., Kim, J.H., Pilka, E.S. et al. (2003) Pathogenic bacteria attach to human fibronectin through a tandem β-zipper. *Nature*, 423(6936), 177–181.

Sichert, A., Corzett, C.H., Schechter, M.S., Unfried, F., Markert, S., Becher, D. et al. (2020) Verrucomicrobia use hundreds of enzymes to digest the algal polysaccharide fucoidan. *Nature Microbiology*, 5(8), 1026–1039.

Sieber, C.M.K., Probst, A.J., Sharrar, A., Thomas, B.C., Hess, M., Tringe, S.G. et al. (2018) Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature Microbiology*, 3(7), 836–843.

Slaby, B.M., Hackl, T., Horn, H., Bayer, K. & Hentschel, U. (2017) Metagenomic binning of a marine sponge microbiome reveals unity in defense but metabolic specialization. *The ISME Journal*, 11(11), 2465–2478.

Spang, A., Poehlein, A., Offre, P., Zumbrägel, S., Haider, S., Rychlik, N. et al. (2012) The genome of the ammonia-oxidizing Candidatus Nitrososphaera gargensis: insights into metabolic versatility and environmental adaptations. *Environmental Microbiology*, 14(12), 3122–3145.

Spormann, A.M. & Widdel, F. (2000) Metabolism of alkylbenzenes, alkanes, and other hydrocarbons in anaerobic bacteria. *Biodegradation*, 11(2), 85–105.

Storey, M.A., Andreassend, S.K., Bracegirdle, J., Brown, A., Keyzers, R.A., Ackerley, D.F. et al. (2020) Metagenomic exploration of the marine sponge *Mycale hentscheli* uncovers multiple polyketide-producing bacterial symbionts. *MBio*, 11(2), e02997–e02919.

Taylor, M.W., Tsai, P., Simister, R.L., Deines, P., Botte, E., Ericson, G. et al. (2013) 'Sponge-specific' bacteria are widespread (but rare) in diverse marine environments. *The ISME Journal*, 7(2), 438–443.

Thomas, T., Rusch, D., DeMaere, M.Z., Yung, P.Y., Lewis, M., Halpern, A. et al. (2010) Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. *The ISME Journal*, 4(12), 1557–1567.

Tian, R.M., Sun, J., Cai, L., Zhang, W.P., Zhou, G.W., Qiu, J.W. et al. (2016) The deep-sea glass sponge *Lophophysema eversa* harbours potential symbionts responsible for the nutrient conversions of carbon, nitrogen and sulfur. *Environmental Microbiology*, 18(8), 2481–2494.

Tian, R.M., Wang, Y., Bougouffa, S., Gao, Z.M., Cai, L., Bajic, V. et al. (2014) Genomic analysis reveals versatile heterotrophic capacity of a potentially symbiotic sulfur-oxidizing bacterium in sponge. *Environmental Microbiology*, 16(11), 3548–3561.

Tianero, M.D., Balaich, J.N. & Donia, M.S. (2019) Localized production of defence chemicals by intracellular symbionts of Haliclona sponges. *Nature Microbiology*, 4(7), 1149–1159.

Webster, N.S., Taylor, M.W., Behnam, F., Lücker, S., Rattei, T., Whalan, S. et al. (2010) Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environmental Microbiology*, 12(8), 2070–2082.

Weisz, J.B., Lindquist, N. & Martens, C.S. (2008) Do associated microbial abundances impact marine demosponge pumping rates and tissue densities? *Oecologia*, 155(2), 367–376.

Wilson, M.C., Mori, T., Rückert, C., Uria, A.R., Helf, M.J., Takada, K. et al. (2014) An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature*, 506(7486), 58–62.

Wu, Y.-W., Simmons, B.A. & Singer, S.W. (2015) MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*, 32(4), 605–607.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F. & Xu, Y. (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research*, 40(W1), W445–W451.

Zhang, F., Blasiak, L.C., Karolin, J.O., Powell, R.J., Geddes, C.D. & Hill, R.T. (2015) Phosphorus sequestration in the form of polyphosphate by microbial symbionts in marine sponges. *Proceedings of the National Academy of Sciences*, 112(14), 4381–4386.

Zhang, F., Pita, L., Erwin, P.M., Abaid, S., López-Legentil, S. & Hill, R. T. (2014) Symbiotic archaea in marine sponges show stability and host specificity in community structure and ammonia oxidation functionality. *FEMS Microbiology Ecology*, 90(3), 699–707.

Zhang, S., Song, W., Wemheuer, B., Reveillaud, J., Webster, N. & Thomas, T. (2019) Comparative genomics reveals ecological and evolutionary insights into sponge-associated Thaumarchaeota. *Msystems*, 4(4), e00288–e00219.

Zhou, Y. & Zhu, Y. (2015) Diversity of bacterial manipulation of the host ubiquitin pathways. *Cellular Microbiology*, 17(1), 26–34.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Engelberts, J.P., Robbins, S.J., Herbold, C.W., Moeller, F.U., Jehmlich, N., Laffy, P.W. et al. (2022) Metabolic reconstruction of the near complete microbiome of the model sponge *Ianthella basta*. *Environmental Microbiology*, 1–15. Available from: https://doi.org/10.1111/1462-2920.16302

CHAPTER VI

Synthesis

Synthesis

The field of sponge microbiology and sponge science, in general, has progressed by leaps and bounds in the past few decades. As ubiquitous benthic-dwelling filter feeders , marine sponges remain a fascinating research topic in symbiosis research, due to the fact that they frequently harbor abundant, highly diverse, and unique symbiotic microbial communities that may exhibit a variety of heretofore undescribed functions, including the production of secondary metabolites that may be developed into natural therapeutics. Moreover, as the sponge-microbe symbiosis may represent the oldest extant example of microbial-metazoan symbiosis, understanding this relationship may shed light on how symbiotic microbes may have contributed to the development of animal multicellularity. With the help of cutting-edge methods and next-generation sequencing technologies, it is now possible to identify the specific microbial symbiont(s), as well as the likely genes and metabolic pathways that may be involved in mediating important biogeochemical functions as well as potential symbiotic or environmental interactions. Nevertheless, the number of studies where the identity of symbiotic microbial actors are unequivocally linked to specific functions are primarily limited to photoautotrophic or chemoautotrophic processes, and even in many of these cases, unambiguous designations are still rare. This dissertation, thus sought to more tightly link symbiotic microbial identity and function in the abundant and widely distributed marine sponge *Ianthella basta*, which maintains highly stable symbioses with three dominant microorganisms. This was achieved through a combination of methods involving targeted incubation experiments using isotope-based and traditional functional assays, along with fluorescence i*n situ* hybridization (FISH) coupled with NanoSIMS, qPCR and metaproteogenomics. The work presented in this dissertation, should therefore provide a methodological framework on how to meaningfully interrogate specific ecophysiological parameters along with identifying the responsible microbial actors in these holobionts. In addition to the specific follow-up research questions brought up by the work presented in Chapters, III, IV and V, many of these issues apply to the field of sponge microbiology in general and add to the discussion on these topics in a larger context.

Chapter III [1] focused on the ecophysiology of '*Ca.* Nitrosospongia ianthellae', the dominant thaumarchaeon residing in *I. basta*, and several results derived from this study are broadly relevant to the topic of nitrification in marine sponges and the interactions among

and between the nitrifying community within sponge holobiont systems. Chapter III, represents the first use of an AOA-specific inhibitor, PTIO, in a study on sponge nitrification, and future studies on sponge nitrification, particularly in marine sponges that harbor both bacterial and ammonia-oxidizers [2–4], should consider the use of AOA or AOB-specific nitrification inhibitors in conjunction with stable-isotope based functional assays, to disentangle the actual activities of the nitrifying communities. To date, only nitrapyrin (NP), a broad nitrification-specific inhibitor affecting both ammonia-oxidizers (AOA and AOB) as well as nitrite-oxidizers ([5] and references therein), has been applied to sponge nitrification studies [2, 6], and the use of a AOA-specific inhibitor, like PTIO, alongside a broad nitrification inhibitor such as NP or acetylene, in these systems could shed further light on the activities of specific symbiont nitrifying communities.

Furthermore, cell-specific activities calculated for nitrification guild members, as in Chapter III, could be coupled with microscopic visualization of cell sizes and calculations of cellular C or N content, which would further illuminate the various ecophysiological strategies underpinning the structure of the nitrifying symbiont community [7]. The relatively low cell-specific ammonia oxidation rates calculated for the *I. basta* AOA symbiont, suggests that not all symbiont cells are physiologically active, perhaps as a result of host digestion or host control. Visualization of host digestion coupled with calculation of symbiont mortality rates, would undoubtedly aid in this endeavor. With a better understanding of sponge nitrifier physiologies, it may be possible to eventually tease apart the ecological as well as the ultimate evolutionary reasons why certain sponges only host AOA, or only AOB, or lack a nitrite oxidizer (like in *I. basta*) or host all members.

The topic of mixotrophy among sponge AOA and AOA, in general, continues to be a topic of interest. A recent metagenomic analysis of sponge-associated AOA confirmed the consistent presence of the ABC-type branched-chain amino acid transport system of the HAAT family (3.A.1.4), found in Chapter III, among 7 additional sponge AOA MAGs [8], while their presence has also been detected in numerous deep sea sediment AOA MAGs [9]. Future research could attempt to discern whether amino acids are indeed imported and utilized by sponge AOA, whether they are autochthonous (i.e., holobiont endogenous) or allochthonous (i.e., exogenous/environmental), and once imported, whether they are degraded for organic carbon incorporation [10, 11], or only for nitrogen incorporation [12],

or used for amino acid recycling (thereby precluding the energy intensive process of de novo biosynthesis of cellular materials) [13], or for the replenishment of the intracellular ammonia pool for ammonia oxidation [14]. Furthermore, the involvement of putative exported proteases found in sponge AOA should be further investigated, to see if they are indeed involved in extracellular protease activity for the purposes of eventual import of peptides and amino acids. Other *I. basta*-specific symbionts also encode various putatively exported proteases, suggesting that this may be a common strategy among sponge symbionts. Proteases are also a topic of interest among sponge researchers, and it remains to be seen whether sponges themselves also utilize proteases, from either exogenous or endogenous sources, for energetic and biomass considerations.

In Chapter IV, *I. basta* endogenous taurine is demonstrated to be utilized by the dominant γ-proteobacterial symbiont '*Ca*. Taurinisymbion ianthellae' for energy conservation and assimilation within the sponge holobiont. Incubation experiments show that the sulfate and ammonia produced by taurine dissimilation are secreted, with the latter chemical being utilized for ammonia-oxidation by the dominant thaumarchaeal symbiont, '*Ca*. Nitrosospongia ianthellae'. Chapter IV thus documents the first study within sponge microbiome research where a specific endogenously-derived metabolic compound is tracked through a sponge holobiont and multiple holobiont community members are attributed for detailed steps in this process. Previous research employing pulse-chase isotopic tracer techniques, with some recently coupling this with NanoSIMS analyses, do not always account for phylogenetic identity and have primarily involved bulk DOM preparations hypothesized to be exogenous or non-targeted DOM compounds, such as glucose [15–18]. Given the panoply of hypothesized functions derived from recent sponge holobiont metagenomic studies, future studies have the opportunity to interrogate these functions in a similarly targeted and mechanistic fashion.

In the case of the *I. basta* holobiont, a couple of notable findings from Chapter IV, deserve further investigation employing a similar approach. Upon investigation of the genome and proteomes of the dominant γ-proteobacterial symbiont, '*Ca.* Taurinisymbion ianthellae', as well as of the dominant α-proteobacterial symbiont, '*Ca.* Luteria ianthellae' (named and further elaborated on in Chapter V), it was found that the dominant γ-symbiont preferentially encodes and expresses ABC transporters predicted to be involved in the

transport of nitrogen containing organic compounds (e.g. amino acids, oligopeptides, polyamines) while the α-symbiont does so for ABC transporters putatively transporting carbohydrates and carboxylic acids. A hypothetical future set of experiments, to demonstrate this niche partitioning, could still utilize bulk preparations of amino acids and oligopeptides versus those of carbon-only compounds, add them to *I. basta* holobiont incubations, and track incorporation into different holobiont members either via single-cell/spatially explicit methods coupled to FISH (e.g. NanoSIMS, Raman microspectroscopy) or using stable-isotope probing (SIP) methods that allow for phylogenetic resolution (DNA-, RNA-, CHIP-, and Protein-SIP) [19]. Additionally, with the use of dual-labeled stable isotope mixtures of these preparations (i.e., $^{13}C^{12}C$, $^{13}C^{13}C$, $^{13}C^{15}N$, $^{15}N^{14}N$ etc.) coupled with spatial metabolomics using mass spectrometry imaging and FISH [19, 20], it may be possible to track specific compounds as they are processed in the sponge holobiont through a variety of sponge cell types, symbiont cells, and metabolic pathways. This could also be applied to determine the specific fates of the organic carbon and nitrogen compounds including their moieties (mentioned above), that are purportedly derived from the utilization of branched-chain amino acids or proteins/oligopeptides by the *I. basta* AOA, '*Ca*. Nitrosospongia ianthellae'.

Chapter IV demonstrated the functioning of a tripartite metabolic network involving sponge-endogenous taurine, but it also suggested an elaborated version of this metabolic network involving DMSP, the sulfur-containing amino acids, cysteine and methionine, vitamin B_{12}, and nitric oxide. Since taurine and DMSP are hypothesized to result in the synthesis and export of cysteine and methionine, both of which are typically acquired via heterotrophy in metazoans and also necessary for taurine biosynthesis itself, future experiments should verify if taurine or DMSP additions to '*Ca*. Taurinisymbion ianthellae' results in the export of these amino acids by the γ-symbiont, and the uptake of them by *I. basta*. Nevertheless, confirmation that *I. basta* encodes the genomic repertoire for cysteine and methionine biosynthesis or has the capacity to synthesize these compounds *in vivo* is still needed, as it is with most sponge species.

Although it would be helpful to attain time series fluctuations of DMSP concentrations in *I. basta* habitat seawater (so as to capture periods of DMSP maxima), a more pressing need would be to determine the intracellular concentration of DMSP within

'*Ca*. Taurinisymbion ianthellae', since *Pelagibacter* cells are predicted to be able to achieve high intracellular concentrations (40 – 180 µM) from naturally measured DMSP abundances (0.1 – 100 nM) [21]. This can also be undertaken in other sponges where DMSP catabolizing genes have been found among sponge microbiome members [4, 22, 23]. Similar to *Pelagibacter*, '*Ca*. Taurinisymbion ianthellae' simultaneously expresses the DMSP cleavage and demethylation pathways, which is a hypothesized adaptation that provides these cells with a kinetically regulated system that favors the pathway to DMS formation (cleavage) when intracellular DMSP concentrations are high, while favoring the demethylation pathway (sulfur for biosynthesis, methanethiol, oxidation of 5-methyl-THF) when intracellular DMSP flux into cell is low [21]. Chapter IV, thus postulates that '*Ca*. Nitrosospongia ianthellae' provides '*Ca*. Taurinisymbion ianthellae' with vitamin B_{12} during periods of low DMSP but high taurine flux into cells, since '*Ca*. Taurinisymbion ianthellae' encodes for both a vitamin B_{12}-dependent methionine synthase (MetH) as well as a vitamin B_{12}-independent methionine synthesis pathway using DMSP as precursor. This thus provides another mechanism in '*Ca*. Taurinisymbion ianthellae' for regulating intracellular sulfur levels, using an additional organic osmolyte. The role of nitric oxide should also be explored since it can induce sponge metamorphosis [24], and can be produced by the AOA symbiont, by the γ-symbiont when it is provided nitrite from the AOA (after provisioning it with ammonia from the breakdown of taurine), as well as most likely by *I. basta* itself. This elaborated tripartite metabolic network thus lends itself to more rigorous ecophysiological experimentation in the future, which can be adapted to systems, artificial or natural, which exhibit analogous features.

Finally, in Chapter V, more *I. basta* individuals were subjected to metagenomic sequencing and metaproteomics to further illuminate the metabolic potential of this species' microbiome and virome. This was done by concentrating on the dominant α-proteobacterial symbiont, '*Ca.* Luteria ianthellae', and other low abundant microbial symbionts, as well as the genomic contributions of viruses. Thus, the information and analysis presented in this chapter allow for a more comprehensive reckoning of the *I. basta* holobiont's genetic and metabolic capabilities and provide a solid theoretical foundation from which further focused ecophysiological studies may be carried out. Several notable findings from these set of analyses provide prospective answers or add critical complexity to

some of the questions raised in Chapter X and Y, while also re-confirming its findings and those of previous sponge microbiome studies.

In the sponge AOA nitrification experiments, presented in Chapter III, nitrate formation and nitrate removal (during incubations exhibiting NH_4^+ inhibition and likely hypoxic conditions), were observed and postulated to be resultant from nitric oxide detoxification and denitrification, respectively. Analyses in Chapter IV and V thus confirmed that denitrification pathway genes are encoded within the MAGs of '*Ca*. Taurinisymbion ianthellae', '*Ca.* Luteria ianthellae', and one of the two low abundant α-symbionts (g_Ruegeria), while nitric oxide detoxification to nitrate, via nitric oxide dioxygenase (Hmp), was found to be encoded in the other low abundant α-symbiont (g_UBA2767). It would be interesting in future studies to discern which holobiont members are active in these process within the *I. basta* holobiont, as well as to elucidate to what extent, and under what conditions they occur in. Another level of complexity is added to taurine and DMSP dissimilation, as well as sulfur oxidation, since the low abundant α-symbiont (g_UBA2767), possesses the genetic capacity for taurine dissimilation into sulfate, DMSP cleavage and demethylation, as well as for the oxidation of thiosulfate to sulfate. Again, future experiments may explore how active this low abundant α-symbiont (g_UBA2767) is in these processes and under what specific conditions they are active in them.

Chapter V delves further into many of the themes explored in Chapter IV and also in previous sponge metagenomic studies, in that a convergent set of functions that are frequently observed in other sponge microbiomes, are encoded in the *I. basta* microbiome, albeit with specializations for each holobiont member [4]. All six microbial symbionts in *I. basta* encoded at least one eukaryotic-like protein (ELP) and all types were found to be expressed within the metaproteomic dataset, with the distribution varying between symbionts. The function of each specific ELP on each symbiont could be further studied by recombinantly expressing the ELPs and exposing the recombinant ELP-bearing cells to sponge cell analogues such as amoeba [25] or choanoflagellates, to sponge cell lines, or even in *I. basta* itself, where for example, cellular fate could be tracked via NanoSIMS. Like '*Ca.* Luteria ianthellae', many of the other low abundance symbionts are predicted to play an important role in the breakdown of complex carbohydrates, a feature found in many sponge microbiomes [26–28], again with each symbiont encoding a specialized repertoire

and with the low abundance planctomycete symbiont (g_UBA1268), exhibiting a richer inventory of capabilities. Notably, the *I. basta* Planctomycetota (g_UBA1268) encodes glycosyl hydrolases putatively acting on chitin (GH18) and sialic acids (GH33), with GH33 being expressed. Also, all 3 α-symbionts as well as '*Ca*. Taurinisymbion ianthellae' encode an ATP-independent periplasmic transporter for sialic acid. Since chitin makes up a constitutive element of sponge skeletons, including in *I. basta* [29], and sialic acid-linked residues have been found in the sponge mesohyl [30], these two compounds deserve further attention within broader sponge and *I. basta* holobiont research.

The fact that ''*Ca*. Nitrosospongia ianthellae', along with several other members of the Thaumarchaeota, encode putative proteins necessary for sialic acid biosynthesis (neuA, neuB, neuC: all three necessary proteins for the final step found encoded in '*Ca*. Taurinisymbion ianthellae') [9], and sponges are one of the few metazoans (which also include the mammalian skin biome and the ascidians) that Thaumarchaeota consistently associates with in symbiotic interactions, it is tempting to speculate that thaumarchaeal sialic acid surface decoration may have been one of the critical factors that enabled the early establishment of this symbiosis. In addition to being reported in more than a dozen pathogenic bacterial species that synthesize sialic acid for surface decoration or utilize it in metabolic scavenging routes [31, 32], this is also a feature that has been identified in a human gut-associated methanogenic archaeon [33]. Indeed, a recent comprehensive phylogenomic analysis of Thaumarchaeota coupled with careful molecular dating suggests colonization of the shallow ocean, by the 'shallow-water group AOA' which includes the heretofore sequenced shallow water sponge AOA [8], occurred ~635-560 Mya [34], around the same time as the emergence of the putative first metazoans, the sponges.

Furthermore, phylogenomic analysis conducted in Chapters IV and V, demonstrate that the closest relatives of both '*Ca*. Taurinisymbion ianthellae' and '*Ca*. Nitrosospongia ianthellae' both reside in two members of the same sponge family (*Ianthellidae*) as *I. basta*, and this coupled with the finding that both are vertically transmitted in *I. basta* [35], suggests an ancient evolutionary association. With the addition of more sponge genomes as well as sponge Thaumarchaeota MAGs, phylogenomic analyses looking at the congruence of host and symbiont phylogenies along with large-scale phylogenomic analyses implementing gene-tree aware ancestral state reconstructions, may eventually reveal the spectre of the

last common thaumarchaeal sponge symbiont ancestor and how this symbiosis was established. With the additional crucial role of nitric oxide as an ancient and key regulator in the marine sponge life cycle, the fact that '*Ca*. Taurinisymbion ianthellae' and '*Ca*. Nitrosospongia ianthellae' both seem to be able to produce NO, provides for the opportunity to design many types of experiments to determine if these factors may help in host metamorphosis or symbiont persistence. To tackle these questions, one can imagine experiments using analogous pure culture representatives of each of the actors involved, whether it be choanoflagellates, viable sponge cell lines, free-living cultured representatives exhibiting similar metabolic pathways, or using recombinant sialic-acid bearing cells to track analog host or symbiont behavior.

Taken in totality, the work presented in this dissertation offers a unique level of detail and experimentation on a sponge holobiont system, one which hopefully serves as model framework for interrogating crucial ecophysiological questions regarding sponge-microbe symbioses. Many of the processes and features exhibited by the *I. basta* holobiont microbiome are congruently found in many sponge holobiont systems and should thereby allow for the examination of many of the similar crucial questions raised by this dissertation. The establishment of pure cell cultures of the dominant symbionts in this system and in others, of experimentally viable sponge holobiont cell models, and genetically tractable systems, would ameliorate many of the experimental challenges inherent in conducting ecophysiological investigations in sponge holobiont systems and thus, any of these advances would, undoubtedly, push the field of sponge-microbe symbiosis into another era.

1. Moeller FU, Webster NS, Herbold CW, Behnam F, Domman D, Albertsen M, et al. Characterization of a thaumarchaeal symbiont that drives incomplete nitrification in the tropical sponge Ianthella basta. *Environmental Microbiology* 2019; **21**: 3831–3854.
2. Bayer K, Schmitt S, Hentschel U. Physiology, phylogeny and *in situ* evidence for bacterial and archaeal nitrifiers in the marine sponge *Aplysina aerophoba*. *Environmental Microbiology* 2008; **10**: 2942–2955.
3. Radax R, Hoffmann F, Rapp HT, Leininger S, Schleper C. Ammonia-oxidizing archaea as main drivers of nitrification in cold-water sponges: Archaeal nitrification in cold-water sponges. *Environmental Microbiology* 2012; **14**: 909–923.
4. Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, Webster NS, et al. Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proceedings of the National Academy of Sciences* 2012; **109**: E1878–E1887.
5. Papadopoulou ES, Bachtsevani E, Lampronikou E, Adamou E, Katsaouni A, Vasileiadis S, et al. Comparison of Novel and Established Nitrification Inhibitors Relevant to Agriculture on Soil Ammonia- and Nitrite-Oxidizing Isolates. *Front Microbiol* 2020; **11**: 581283.
6. Southwell MW, Popp BN, Martens CS. Nitrification controls on fluxes and isotopic composition of nitrate from Florida Keys sponges. *Marine Chemistry* 2008; **108**: 96–108.
7. Kitzinger K, Marchant HK, Bristow LA, Herbold CW, Padilla CC, Kidane AT, et al. Single cell analyses reveal contrasting life strategies of the two main nitrifiers in the ocean. *Nat Commun* 2020; **11**: 767.
8. Haber M, Burgsdorf I, Handley KM, Rubin-Blum M, Steindler L. Genomic Insights Into the Lifestyles of Thaumarchaeota Inside Sponges. *Front Microbiol* 2021; **11**: 622824.
9. Kerou M, Ponce-Toledo RI, Zhao R, Abby SS, Hirai M, Nomaki H, et al. Genomes of Thaumarchaeota from deep sea sediments reveal specific adaptations of three independently evolved lineages. *ISME J* 2021; **15**: 2792–2808.
10. Wang Y, Huang J, Cui G, Nunoura T, Takaki Y, Li W, et al. Genomics insights into ecotype formation of ammonia-oxidizing archaea in the deep ocean. *Environ Microbiol* 2019; **21**: 716–729.
11. Li M, Baker BJ, Anantharaman K, Jain S, Breier JA, Dick GJ. Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. *Nature Communications* 2015; **6**.
12. Dekas AE, Parada AE, Mayali X, Fuhrman JA, Wollard J, Weber PK, et al. Characterizing Chemoautotrophy and Heterotrophy in Marine Archaea and Bacteria With Single-Cell Multi-isotope NanoSIP. *Front Microbiol* 2019; **10**: 2682.
13. Li J, Mara P, Schubotz F, Sylvan JB, Burgaud G, Klein F, et al. Recycling and metabolic flexibility dictate life in the lower oceanic crust. *Nature* 2020; **579**: 250–255.
14. Burton SAQ, Prosser JI. Autotrophic Ammonia Oxidation at Low pH through Urea Hydrolysis. *Applied and Environmental Microbiology* 2001; **67**: 2952–2957.

15. Achlatis M, Pernice M, Green K, de Goeij JM, Guagliardo P, Kilburn MR, et al. Single-cell visualization indicates direct role of sponge host in uptake of dissolved organic matter. *Proceedings of the Royal Society B: Biological Sciences* 2019; **286**: 20192153.
16. Hudspith M, Rix L, Achlatis M, Bougoure J, Guagliardo P, Clode PL, et al. Subcellular view of host–microbiome nutrient exchange in sponges: insights into the ecological success of an early metazoan–microbe symbiosis. *Microbiome* 2021; **9**: 44.
17. Rix L, Ribes M, Coma R, Jahn MT, de Goeij JM, van Oevelen D, et al. Heterotrophy in the earliest gut: a single-cell view of heterotrophic carbon and nitrogen assimilation in sponge-microbe symbioses. *ISME J* 2020; **14**: 2554–2567.
18. Campana S, Busch K, Hentschel U, Muyzer G, de Goeij JM. DNA -stable isotope probing (DNA-SIP) identifies marine sponge-associated bacteria actively utilizing dissolved organic matter (DOM). *Environmental Microbiology* 2021; **23**: 4489–4504.
19. Alcolombri U, Pioli R, Stocker R, Berry D. Single-cell stable isotope probing in microbial ecology. *ISME COMMUN* 2022; **2**: 55.
20. Geier B, Sogin EM, Michellod D, Janda M, Kompauer M, Spengler B, et al. Spatial metabolomics of in situ host–microbe interactions at the micrometre scale. *Nat Microbiol* 2020; **5**: 498–510.
21. Sun J, Todd JD, Thrash JC, Qian Y, Qian MC, Temperton B, et al. The abundant marine bacterium Pelagibacter simultaneously catabolizes dimethylsulfoniopropionate to the gases dimethyl sulfide and methanethiol. *Nature Microbiology* 2016; **1**: 1–5.
22. Tian R-M, Wang Y, Bougouffa S, Gao Z-M, Cai L, Bajic V, et al. Genomic analysis reveals versatile heterotrophic capacity of a potentially symbiotic sulfur-oxidizing bacterium in sponge: A versatile SOB as a potential sponge symbiont. *Environmental Microbiology* 2014; **16**: 3548–3561.
23. Moreno-Pino M, Cristi A, Gillooly JF, Trefault N. Characterizing the microbiomes of Antarctic sponges: a functional metagenomic approach. *Scientific Reports* 2020; **10**: 645.
24. Ueda N, Richards GS, Degnan BM, Kranz A, Adamska M, Croll RP, et al. An ancient role for nitric oxide in regulating the animal pelagobenthic life cycle: evidence from a marine sponge. *Scientific Reports* 2016; **6**.
25. Reynolds D, Thomas T. Evolution and function of eukaryotic-like proteins from sponge symbionts. *Molecular Ecology* 2016; **25**: 5242–5253.
26. Kamke J, Sczyrba A, Ivanova N, Schwientek P, Rinke C, Mavromatis K, et al. Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. *ISME J* 2013; **7**: 2287–2300.
27. Siegl A, Kamke J, Hochmuth T, Piel J, Richter M, Liang C, et al. Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges. *The ISME Journal* 2011; **5**: 61–70.
28. Robbins SJ, Song W, Engelberts JP, Glasl B, Slaby BM, Boyd J, et al. A genomic view of the microbiome of coral reef demosponges. *ISME J* 2021; **15**: 1641–1654.

29. Mutsenko VV, Gryshkov O, Lauterboeck L, Rogulska O, Tarusin DN, Bazhenov VV, et al. Novel chitin scaffolds derived from marine sponge Ianthella basta for tissue engineering approaches based on human mesenchymal stromal cells: Biocompatibility and cryopreservation. *International Journal of Biological Macromolecules* 2017; **104**: 1955–1965.
30. Garrone R, Thiney Y, Pavans de Ceccatty M. Electron microscopy of a mucopolysaccharide cell coat in sponges. *Experientia* 1971; **27**: 1324–1326.
31. Vimr E, Lichtensteiger C. To sialylate, or not to sialylate: that is the question. *Trends in Microbiology* 2002; **10**: 254–257.
32. Kelm S, Schauer R. Sialic Acids in Molecular and Cellular Interactions. *International Review of Cytology*. 1997. Elsevier, pp 137–240.
33. Samuel BS, Hansen EE, Manchester JK, Coutinho PM, Henrissat B, Fulton R, et al. Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut. *Proc Natl Acad Sci USA* 2007; **104**: 10643–10648.
34. Ren M, Feng X, Huang Y, Wang H, Hu Z, Clingenpeel S, et al. Phylogenomics suggests oxygen availability as a driving force in Thaumarchaeota evolution. *ISME J* 2019; **13**: 2150–2161.
35. Engelberts JP, Abdul Wahab MA, Maldonado M, Rix L, Marangon E, Robbins SJ, et al. Microbes from Mum: symbiont transmission in the tropical reef sponge Ianthella basta. *ISME COMMUN* 2022; **2**: 90.

Appendix

Abstract

Zusammenfassung

Abstract

Since sponges may be the most ancient of the extant multicellular organisms, the sponge-microbe holobiont therefore perhaps embodies one of the oldest extant examples of microbial-metazoan symbiosis. This ancient symbiotic relationship has resulted in of the most complex and diversified holobionts in the marine environment, with >8000 globally distributed sponge species. The microbial symbionts within sponges can comprise up to 35% of total sponge biomass, and microbiome communities have been found to be species-specific and stably associated across various prokaryotic taxonomic levels. As major denizens of many marine benthic environments, sponges, through their filter feeding capabilities, are able to mediate critical biogeochemical processes, that have been convincingly demonstrated, in some cases, to be attributed to the actions of their microbial symbionts. The use of next-generation sequencing technologies, alongside state-of-the-art techniques, have provided a glut of sponge microbiome data that began to identify the specific microbial symbiont(s) and the likely genes and metabolic pathways that are active in mediating hitherto overlooked biogeochemical functions as well as potential symbiotic or environmental interactions. The list of proposed symbiotic functions from this era have undoubtedly grown quite expansive, and unsurprisingly the examples of clear-cut functional assignments to specific symbionts for these newly hypothesized functions has been rather rare. Examples of precise functional designations to specific symbionts have been quite scarce given the length of the list of potential symbiotic activities, as well as due to the difficulty in cultivating dominant sponge holobiont members and experimentally manipulating sponge holobionts in targeted, meaningful ways. The goal of the work assembled in this dissertation is thus to interrogate the ecophysiology of dominant sponge microbiome residing in the marine sponge, *Ianthella basta*, through a combination of metaproteogenomics, fluorescence *in situ* hybridization, qPCR, and hypothesis-driven laboratory incubations coupled with isotope-based functional assays. The first focus was on the ecophysiological and metaproteogenomic characterization of the dominant thaumarachaeal symbiont, '*Candidatus* Nitrospongia ianthellae', and this resulted in the demonstration that it is responsible for ammonia oxidation within the sponge holobiont and that it hosts several adaptive genomic features shared with other thaumarchaeal sponge symbionts as well as other convergent functions found in sponge microbiomes. The next focus was on the dominant gammaproteobacterial symbiont '*Candidatus* Taurinisymbion ianthellae' and using a similar approach, it was demonstrated that it incorporates carbon and nitrogen derived from sponge endogenous taurine, oxidizes the dissimilated sulfite into sulfate for export, and provides

the dissimilated ammonia to '*Candidatus* Nitrospongia ianthellae' for ammonia oxidation. Finally, with additional metagenomic sequencing and metaproteomics of the *I. basta* holobiont, the full genomically predictable metabolic potential of its microbiome is provided, enabling an encompassing framework within which to use *I. basta* in future manipulative experiments.

Zusammenfassung

Da Schwämme die ältesten der existierenden mehrzelligen Organismen sein könnten, verkörpert der Schwamm-Mikroben-Holobiont daher vielleicht eines der ältesten erhaltenen Beispiele für eine Mikroben-Metazoen-Symbiose. Diese uralte symbiotische Beziehung gehört mit über 8000 weltweit verbreiteten Schwammarten zu den komplexesten und vielfältigsten Holobionten in marinen Habitaten. Die mikrobiellen Symbionten können bis zu 35% der gesamten Biomasse von Schwämmen ausmachen, und es wurde festgestellt, dass Mikrobiomgemeinschaften artenspezifisch und über verschiedene prokaryotische taxonomische Ebenen hinweg stabil assoziiert sind. Als Hauptbewohner vieler benthischer Meeresumgebungen sind Schwämme durch ihre Filterfähigkeiten in der Lage, kritische biogeochemische Prozesse zu vermitteln, von denen in einigen Fällen überzeugend gezeigt wurde, dass sie den Aktivitäten ihrer symbiotischen Mikroben zugeschrieben werden können. Die Verwendung von Sequenzierungstechnologien der nächsten Generation hat, zusammen mit anderen modernen Techniken der Molekularbiologie, eine Fülle von Daten zum Schwammmikrobiom geliefert. Damit wurde begonnen, die spezifischen mikrobiellen Symbionten und die wahrscheinlichen Gene und Stoffwechselwege zu identifizieren, die bisher ubersehener biogeochemische Funktionen vermitteln, sowie potenzielle symbiotische oder Umweltinteraktionen zu entschlüsseln. Die Liste der vorgeschlagenen symbiotischen Funktionen ist zweifellos sehr umfangreich geworden, und es überrascht nicht, dass die Beispiele für eindeutige funktionelle Zuordnungen zu bestimmten Symbionten für diese neu angenommenen Funktionen eher selten waren. Beispiele für genaue funktionale Bezeichnungen für bestimmte Symbionten waren angesichts der Länge der Liste potenzieller symbiotischer Aktivitäten sowie aufgrund der Schwierigkeit, dominante Schwamm-Holobionten-Mitglieder zu kultivieren und Schwamm-Holobionten auf gezielte, sinnvolle Weise experimentell zu manipulieren, recht selten. Das Ziel der in dieser Dissertation zusammengetragenen Arbeit ist es daher, die Ökophysiologie des dominanten Schwammmikrobioms im Meeresschwamm Ianthella basta durch eine Kombination aus Metaproteogenomik, Fluoreszenz-in-situ-Hybridisierung, qPCR und hypothesengetriebenen Laborinkubationen in Verbindung mit Isotopen-basierten funktionellen Assays zu untersuchen. Der erste Schwerpunkt lag auf der ökophysiologischen und metaproteogenomischen Charakterisierung des dominanten Thaumarachae-Symbionten 'Candidatus Nitrospongia ianthellae', und dies führte zur Demonstration, dass er für die Ammoniakoxidation innerhalb des Schwamm-Holobionten verantwortlich ist, und dass er mehrere gemeinsame adaptive genomische Merkmale beherbergt mit anderen Thaumarchae-

Schwammsymbionten, sowie anderen konvergenten Funktionen, die in Schwammmikrobiomen gefunden werden. Der nächste Schwerpunkt lag auf dem dominanten gammaproteobakteriellen Symbionten „Candidatus Taurinisymbion ianthellae“, und mit einem ähnlichen Ansatz wurde gezeigt, dass er Kohlenstoff und Stickstoff aus dem endogenen Taurin des Schwamms einbaut, welches er dissimilierte; dabei Sulfit zu Sulfat oxidiert, und Ammoniak für 'Candidatus Nitrospongia ianthellae' für die Ammoniakoxidation bereitstellt. Schließlich wurde mit zusätzlicher metagenomischer Sequenzierung und Metaproteomik des I. basta-Holobionten das volle genomisch vorhersagbare metabolische Potenzial seines Mikrobioms bereitgestellt, was einen umfassenden Rahmen ermöglicht, in dem I. basta in zukünftigen manipulativen Experimenten verwendet werden kann.