Horizontal acquisition followed by expansion and diversification of toxin-related genes in deep-sea bivalve symbionts

Lizbeth Sayavedra[1,2#], Rebecca Ansorge[1], Maxim Rubin-Blum[1,3], Nikolaus Leisch[1], Nicole Dubilier[1,4], Jillian M. Petersen[1,5,#]

[1]Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, 28359 Bremen, Germany
[2]Quadram Institute Bioscience, Norwich Research Park, Norwich, United Kingdom
[3]Israel Limnology and Oceanography Research, Tel Shikmona, 3108000, Haifa, Israel
[4]MARUM, University of Bremen, Leobener Str. 2, 28359 Bremen, Germany
[5]Division of Microbial Ecology, Department of Microbiology and Ecosystem Science, Research Network Chemistry Meets Microbiology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria
[#]Corresponding authors

Key words

Bathymodiolus, symbiosis, toxin genes, host-symbiont interactions, sulfur-oxidizing bacteria

Abstract

Deep-sea bathymodioline mussels gain their nutrition from intracellular bacterial symbionts. Their sulfur-oxidizing (SOX) symbionts were recently shown to encode abundant toxin-related genes (TRGs) in their genomes, which may play a role in beneficial host-microbe interactions. Here, we compared TRGs in the genomes of SOX symbionts from 10 bathymodioline mussel and two sponge species to better understand their potential functions and evolutionary origins. Despite the close phylogenetic relatedness of these symbionts, the number and classes of encoded toxins varied greatly between host species. One of the TRG classes, YDs, has experienced gene expansions multiple times, suggesting that these genes are under adaptive selection. Some symbiont genomes contained secretion systems, which can play a role in host-microbe interactions. Both TRGs and secretion systems had a heterogeneous distribution, suggesting that these closely related bacteria have acquired different molecular

mechanisms for interacting with the same family of animal hosts, possibly through convergent evolution.

Introduction

Beneficial associations between animals and bacteria are virtually universal (McFall-Ngai et al. 2013). Many beneficial bacteria are acquired from the environment during host development, but the mechanisms that underpin host-symbiont recognition, invasion of host tissues or cells, and maintenance of the associations, are still not well understood (Pel and Pieterse 2013). In contrast, the molecular mechanisms pathogens use to interact with their hosts have been intensively studied (e.g. Sansonetti 2002; Di Genova and Tonelli 2016; Kaufmann and Dorhoi 2016; Kendall and Sperandio 2016). A number of pathogen-encoded proteins that interfere with host cell activity have been described and characterized as toxins (Lang et al. 2010; Aktories 2011; Huber et al. 2016). Large-scale bacterial genome sequencing has revealed toxin-related genes (TRGs) in the genomes of many beneficial bacteria with homology to characterized toxins of pathogens. This suggests that pathogens and beneficial bacteria use similar molecular mechanisms to interact with their hosts (Moya et al. 2008; Pérez-Brocal et al. 2011).

Bathymodioline mussels thrive at deep-sea hydrothermal vents and cold seeps by gaining nutrition from intracellular sulfur- and methane-oxidizing bacteria, which they harbor in their gill cells (Fisher et al. 1987; Duperron et al. 2009; Ponnudurai et al. 2016). Sayavedra et al. (2015) recently discovered diverse and abundant TRGs in the genomes of the sulfur-oxidizing (SOX) symbionts from two *Bathymodiolus* mussel species. These TRGs were hypothesized to play a role in beneficial host-microbe interactions, including host-symbiont communication and defense against parasites (Sayavedra et al., 2015). The sulfur-oxidizing (SOX) symbionts are acquired from the environment by each new host generation (Won et al. 2003; Wentrup et al. 2014), but little is known about the mechanisms the symbionts use to invade and survive within host cells.

In this study, we investigated the distribution of toxin-related genes (TRGs) in the SOX symbionts of ten *Bathymodiolus* species and in the closely-related SOX symbionts of two deep-sea sponge species. We hypothesized that TRGs encoded by all symbionts associated with a certain animal group (mussels or sponges) would be essential for interactions with their animal host such as recognition and invasion of host cells. Furthermore, given that TRGs were most likely acquired by the SOX symbionts through horizontal gene transfer, we aimed to understand how TRG acquisition has influenced genome evolution in this closely-related group of symbiotic bacteria.

Results and Discussion

Phylogenomic analyses reveal two well-supported symbiont clades

Previously, genome sequences were available from the SOX symbionts of three bathymodioline species from vents in the Pacific and Atlantic Oceans (Ikuta et al. 2015; Sayavedra et al. 2015). We sequenced and assembled the draft genomes of SOX symbionts from seven additional mussel species from vents and seeps around the world (Table 1, Supplementary Table 1). Furthermore, we assembled SOX symbiont genomes from metagenomes of three poecilosclerid sponges from the Gulf of Mexico that co-occur with two of the bathymodioline species investigated in this study (Rubin-Blum et al. 2017). The draft genomes sequenced in this study were between 90.8 to 98.5% complete, and were sequenced to depths ranging from 24x to 3600x. Their estimated genome sizes ranged from 1.41 to 2.82 Mbp. Many of these symbiont genomes may thus be larger than the only closed genome, that of the SOX symbiont of *B. septemdierum,* which is 1.47 Mbp (Ikuta et al. 2015).

We constructed a well-supported phylogenomic tree with 38 orthologous protein-coding genes from the sponge and mussel SOX symbionts and their close relatives. Consistent with previous 16S rRNA phylogenies, the SOX symbionts from mussels, sponges and clams did not form a monophyletic clade, as they were interspersed with free-living sulfur-oxidizing bacteria called 'SUP05' (Petersen et al. 2012; Sayavedra et al. 2015). The sponge-associated SOX symbionts formed a cluster together with most *Bathymodiolus* SOX symbionts, which we termed Clade 1 (Fig. 1 and Fig. 2). The symbionts of two mussel species, *B. heckerae* (BheckSOX) and *B*. sp. nov GoM (BspGoMSOX), clustered in a separate well-supported clade, together with the cultivated sulfur oxidizer *Candidatus* Thioglobus autotrophica EF1 and SUP05 bacteria from the Pacific Northwest (Clade 2, Fig. 1). The intermixing of symbiotic and free-living bacteria in our phylogenomic analysis, and in previous 16S rRNA phylogenies, suggests that either 1) free-living SOX bacteria acquired the ability to associate with bathymodioline mussels multiple times or 2) the free-living bacteria that fall within the highly supported clade of SOX symbionts from mussels, sponges and clams evolved from a symbiotic ancestor. So far, there is no evidence that these symbionts have a free-living stage that is metabolically active, although very closely-related free-living bacteria from the SUP05/*Ca*. *Thioglobus* clade are often abundant in hydrothermal vent environments (Anantharaman et al. 2012; Meier et al. 2017). In fact, the symbionts may rely on their hosts for some essential metabolites since they appear to lack two enzymes considered to be critical for anaplerotic metabolism (Ponnudurai et al. 2016). However, the isolate *Ca*. Thioglobus autotrophica also

lacks one of these central metabolic enzymes: malate dehydrogenase. Thus, a free-living existence may be possible without enzymes previously assumed to be essential. We cannot rule out either of our two explanations above, but clearly, the well-supported clustering of sponge and mussel symbionts suggests that they shared a common ancestor, possibly undergoing a host-switching event, as well as multiple lifestyle switches from free-living to symbiotic and possibly symbiotic to free-living.

Horizontal acquisition, expansion and diversification of toxin-related gene families

SOX symbiont genomes from the two mussel species described by Sayavedra et al. (2015) encoded TRGs from three toxin classes: 1) RTX, or 'repeat in toxin' proteins, 2) MARTX toxins, which are large proteins containing multiple repeat motifs and domains of diverse functions, and 3) YD repeat toxins, named for their characteristic repeat sequence. In this study, we searched for TRGs in the SOX symbiont genomes of eight additional mussel species and two sponge species, as well as their closest free-living and symbiotic relatives (Table 1) (see SI Materials and Methods).

We consistently found TRGs in the SOX symbiont genomes of mussels and sponges, and these were highly abundant in the symbionts of mussel species (Fig. 1). In contrast, none of the genome sequences from bacteria closely related to the mussel and sponge SOX symbionts, such as free-living SUP05 and the vertically-transmitted, obligate intracellular symbionts of clams, encoded TRGs (Fig. 1).

MARTX. One toxin class, MARTX, was found in all of the mussel SOX symbiont genomes, regardless of whether they belonged to Clade 1 or 2. Intriguingly, MARTX were not found in any of the sponge symbiont genomes, even though these symbionts formed a highly-supported phylogenomic cluster together with mussel SOX symbionts. MARTX-like genes are known to be enriched in the genomes of symbiotic and pathogenic bacteria that associate with eukaryotes, and often have domains involved in attachment (Satchell 2011). The presence of MARTX-like genes in all mussel SOX symbionts from two distinct clades, and their absence in closely related free-living bacteria and the symbionts of clams and sponges, is consistent with a role in specific interactions with the mussel hosts, which could include attachment and recognition during colonization and intracellular infection of host gill cells. The length, sequences, domain content and arrangement of MARTX genes were highly diverse as shown previously for symbionts of two mussel species (Sayavedra et al. 2015) (Fig. 3). Despite this variable domain architecture, the symbionts of all 10 mussel species investigated had at least

one MARTX-like gene with domains involved in attachment such as haemmagglutinin, cadherin, and integrin, indicating their role in attachment to host cells (Supplementary Table 2). If they are involved in attachment, they might also play a key role in mediating recognition and specificity. The mussel SOX symbioses are clearly highly specific: all except one of the known host species associate with only one or two 16S rRNA SOX types, which are not found in any other mussel species (see Duperron et al. 2008 for the only known exception). This host specificity is strictly maintained even when multiple mussel species co-occur, such as *B. brooksi* and *B. heckerae* at cold seeps in the Gulf of Mexico (Raggi et al. 2013). The highly divergent sequence and domain architectures of the MARTX genes in different symbiont lineages might be one of the mechanisms that determine this specificity. Although lacking MARTX genes, the SOX symbionts of sponges encoded proteins with leucine-rich repeats and cadherin domains, which have been hypothesized to play a role in recognition in shallow-water sponge symbioses (Thomas et al. 2010; Hentschel et al. 2012).

RTX and YD repeats. The second toxin class, RTX, was found in some but not all mussel and sponge symbionts from Clade 1, and not in any of the Clade 2 symbionts. The third class of genes, YDs, was found in all members of Clade 1 except the basal *B. septemdierum* symbiont (Fig. 1). Clade 2 symbionts did not contain any YD repeat genes, but these symbionts co-exist in a dual symbiosis with Clade 1 symbionts that did encode YD repeats (Fig. S1). These observations support the following hypotheses: I) RTX and YD repeats are not essential for establishing and maintaining an intracellular symbiotic association with mussels, II) RTX genes were acquired by the common ancestor of Clade 1 and lost on multiple occasions, III) YD genes were acquired by the common ancestor of Clade 1, and YD genes were subsequently lost in the *B. septemdierum* symbiont, and IV) gene duplication contributed to the expansion of the YD genes (SI Results and Discussion). Given that YD and RTX appear to not be essential for intracellular symbiosis, their main role might be to defend their hosts against possible pathogens or parasites (SI Results and Discussion).

Secretion system genes. Secretion systems (SS) are often essential for pathogens to survive inside host cells (Green and Mecsas 2016). We therefore searched for SS components in the genomes of the SOX symbionts and their free-living relatives. We found genes encoding components of almost all known SS types. Like the TRGs, these SS components were patchily distributed among the SOX symbiont genomes, and not a single SS was specific to all of the intracellular bacteria (Supplementary Table 3).

All genome bins from SOX symbionts of Clades 1 and 2 encoded VgrG, a component of the type VI SS (T6SS). Although none of the genomes analyzed in this study encoded the full suite of T6SS genes, VrgG alone may allow the export of toxins without the full T6SS gene

array (Hachani et al. 2014). This gene was also present in some of the free-living SOX relatives. Three genes considered essential for T4SS were present in three of the thirteen SOX symbionts of Clade 1 and in both SOX symbionts of Clade 2, but not in any of the free-living or clam SOX. Most of the T4SS present in the mussel SOX symbionts encoded a relaxase that can interact with DNA, supporting a role in conjugation (Abby et al. 2016). In some pathogens, the same T4SS can carry out dual functions in conjugation and host colonization (Dehio 2008).

The BheckSOX of Clade 2 encoded an additional T4SS of the type VirB/D, which was not found in any other mussel SOX symbionts. The BheckSOX VirB/D-T4SS shares a similar genomic architecture with systems used for both conjugation (e.g. *Vibrio parahaemolyticus*), and for host cell invasion and persistence through secretion of toxic effectors (e.g. *Bartonella henselae* str. Houston-1) (Seubert et al. 2003; Schmid et al. 2004; Dehio 2008; Gokulan et al. 2013). A phage integrase was found upstream of the VirB/D T4SS gene cluster, raising the possibility that, just as in pathogens, beneficial bacteria could be acquiring and exchanging secretion systems from bacteriophages (Guy et al. 2013).

Conclusions

The SOX symbionts of deep-sea mussels and sponges encoded a highly diverse array of toxin-related and secretion system genes. Our comparative genomic analyses identified only one toxin class, MARTX, which was common to all mussel SOX symbionts and might therefore be a gene class essential for host-microbe interactions such as recognition, attachment and symbiont uptake in the mussel symbioses. All other TRGs and secretion systems had a heterogeneous distribution in the symbionts we investigated, which attests to the complex and varied routes of genome evolution taken by the members of this closely-related group of symbiotic bacteria. If the SOX symbionts use their species-specific sets of TRGs and secretion systems to interact with their respective hosts, this would be an example of convergent evolution in which free-living bacteria took multiple unique evolutionary trajectories to become intracellular symbionts of animals, depending on the genes they acquired.

TRGs and T4SS that could export protein effectors were not present in free-living SUP05, even though these bacteria are often found in hydrothermal vent plumes in close proximity to mussels (Sylvan et al. 2012; Anantharaman et al. 2014). It is therefore likely that these genes were acquired from other free-living or host-associated bacteria. Gene flow between these bacterial donors, SUP05 bacteria, and SOX symbionts in a ‘free-living’ stage in the environment could lead to the evolution of novel symbiont and free-living lineages (SI Results and Discussion) (Roux et al. 2014, our own unpublished data). Further investigation of

198 horizontal gene transfer and genome evolution in groups of closely related bacteria such as the
199 SUP05 and SOX symbionts, could reveal how free-living bacteria become symbionts.

200 Some pathogen groups such as *Pseudomonas aeruginosa* show a similar pattern to the SOX
201 symbionts we investigated, with species- or strain-specific differences in their genomic
202 complement of toxins and virulence factors (Huber et al. 2016). In *P. aeruginosa*, these
203 genomic differences are clearly reflected in major phenotypic differences such as severity of
204 human disease. At the morphological level, the SOX symbionts of different mussel and sponge
205 species do not show clear differences. However, just as in pathogens, the underlying genomic
206 variation between symbionts could result in differences in the way these diverse symbionts
207 interact with their hosts. For example, some host species seem to consistently carry a higher
208 symbiont load than others (Duperron et al. 2008; Raggi et al. 2013), and this could not only be
209 due to differences in the availability of their energy sources, but also to differences in the rates
210 of symbiont acquisition, maintenance, proliferation and digestion by the host. TRGs and SSs
211 are likely to affect such host-microbe interactions and could thus have a significant impact on
212 the functioning and stability of these symbioses (Huber et al. 2016).

213 ## Acknowledgements

214 We thank the captains, crews and funding agencies of the sampling cruises AT26-23,
215 M64-2, M67-2, M78-2, ATA57, M114-2 and NA043. We thank Christian Borowski, Stephanie
216 Markert, and Charles Fisher for providing samples, Brandon Seah for helpful discussions and
217 Miriam Sadowski for technical assistance. This work was funded by the Max Planck Society,
218 the DFG Cluster of Excellence "The Ocean in the Earth System" at MARUM (University of
219 Bremen), a European Research Council Advanced Grant (BathyBiome, Grant 340535) and a
220 Gordon and Betty Moore Foundation Marine Microbiology Initiative Investigator Award
221 through Grant GBMF3811 to ND, the DAAD through a doctoral grant to LS, and the Vienna
222 Science and Technology Fund (WWTF) through project VRG14-021 to JMP.

223 ## Statement of competing interests

224 The authors declare no competing interests.

225 ## Author contributions

226 LS, ND, and JP conceived the study; LS, RA, and MRB analyzed the data; NL did TEM
227 analysis; LS and JP wrote the paper with contributions and revisions from all coauthors.

References

Abby SS, Cury J, Guglielmini J, Néron B, Touchon M, Rocha EPC. 2016. Identification of protein secretion systems in bacterial genomes. Sci. Rep. 6:23080.

Abby SS, Néron B, Ménager H, Touchon M, Rocha EPC. 2014. MacSyFinder: A Program to mine genomes for molecular systems with an application to CRISPR-Cas Systems. PLOS ONE 9:e110726.

Aktories K. 2011. Bacterial protein toxins that modify host regulatory GTPases. Nat. Rev. Microbiol. 9:487–498.

Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat. Biotechnol. 31:533–538.

Anantharaman K, Breier JA, Sheik CS, Dick GJ. 2012. Evidence for hydrogen oxidation and metabolic plasticity in widespread deep-sea sulfur-oxidizing bacteria. Proc. Natl. Acad. Sci. 110:330–335.

Anantharaman K, Duhaime MB, Breier JA, Wendt KA, Toner BM, Dick GJ. 2014. Sulfur oxidation genes in diverse deep-sea viruses. Science 344:757–760.

Assié A, Borowski C, van der Heijden K, Raggi L, Geier B, Leisch N, Schimak MP, Dubilier N, Petersen JM. 2016. A specific and widespread association between deep-sea *Bathymodiolus* mussels and a novel family of Epsilonproteobacteria. Environ. Microbiol. Rep. 8:805–813.

Aziz R, Bartels D, Best A, DeJongh M, Disz T, Edwards R, Formsma K, Gerdes S, Glass E, Kubal M, et al. 2008. The RAST Server: Rapid annotations using subsystems technology. BMC Genomics 9:75.

Boetzer M, Pirovano W. 2012. Toward almost closed genomes with GapFiller. Genome Biol. 13:R56.

Bouck J, Miller W, Gorrell JH, Muzny D, Gibbs RA. 1998. Analysis of the quality and utility of random shotgun sequencing at low redundancies. Genome Res. 8:1074–1084.

Buchfink B, Xie C, Huson DH. 2014. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12:59–60.

Dehio C. 2008. Infection-associated type IV secretion systems of *Bartonella* and their diverse roles in host cell interaction. Cell. Microbiol. 10:1591–1598.

Di Genova BM, Tonelli RR. 2016. Infection strategies of intestinal parasite pathogens and host cell responses. Front. Microbiol. [Internet] 7. Available from: http://journal.frontiersin.org/article/10.3389/fmicb.2016.00256

Domman D, Collingro A, Lagkouvardos I, Gehre L, Weinmaier T, Rattei T, Subtil A, Horn M. 2014. Massive expansion of ubiquitination-related gene families within the *Chlamydiae*. Mol. Biol. Evol. 31:2890–2904.

Dubilier N, Bergin C, Lott C. 2008. Symbiotic diversity in marine animals: The art of harnessing chemosynthesis. Nat Rev Micro 6:725–740.

Duperron S, Halary S, Lorion J, Sibuet M, Gaill F. 2008. Unexpected co-occurrence of six bacterial symbionts in the gills of the cold seep mussel *Idas* sp. (Bivalvia: Mytilidae). Environ. Microbiol. 10:433–445.

267 Duperron S, Lorion J, Samadi S, Gros O, Gaill F. 2009. Symbioses between deep-sea mussels
268 (Mytilidae: Bathymodiolinae) and chemosynthetic bacteria: Diversity, function and evolution.
269 C. R. Biol. 332:298–310.

270 Faure B, Schaeffer SW, Fisher CR. 2015. Species distribution and population connectivity of deep-sea
271 mussels at hydrocarbon seeps in the Gulf of Mexico. PLoS ONE [Internet] 10. Available from:
272 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4393317/

273 Fisher C, Childress J, Oremland R, Bidigare R. 1987. The importance of methane and thiosulfate in the
274 metabolism of the bacterial symbionts of two deep-sea mussels. Mar. Biol. 96:59–71.

275 Fujiwara Y, Takai K, Uematsu K, Tsuchida S, Hunt J, Hashimoto J. 2000. Phylogenetic characterization
276 of endosymbionts in three hydrothermal vent mussels: Influence on host distributions. Mar.
277 Ecol. Prog. Ser. 208:147–155.

278 Gokulan K, Khare S, Rooney AW, Han J, Lynne AM, Foley SL. 2013. Impact of plasmids, including
279 those encodingVirB4/D4 type IV secretion systems, on *Salmonella enterica* serovar Heidelberg
280 virulence in macrophages and epithelial cells. PLOS ONE 8:e77866.

281 Green ER, Mecsas J. 2016. Bacterial secretion systems – an overview. Microbiol. Spectr. [Internet] 4.
282 Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4804464/

283 Guy L, Nystedt B, Toft C, Zaremba-Niedzwiedzka K, Berglund EC, Granberg F, Näslund K, Eriksson
284 A-S, Andersson SGE. 2013. A gene transfer agent and a dynamic repertoire of secretion
285 systems hold the keys to the explosive radiation of the emerging pathogen *Bartonella*. PLOS
286 Genet. 9:e1003393.

287 Hachani A, Allsopp LP, Oduko Y, Filloux A. 2014. The VgrG proteins are "à la Carte" delivery systems
288 for bacterial Type VI effectors. J. Biol. Chem. 289:17872–17884.

289 Hentschel U, Piel J, Degnan SM, Taylor MW. 2012. Genomic insights into the marine sponge
290 microbiome. Nat. Rev. Microbiol. 10:641–654.

291 Hillman K, Goodrich-Blair H. 2016. Are you my symbiont? Microbial polymorphic toxins and
292 antimicrobial compounds as honest signals of beneficial symbiotic defensive traits. Curr. Opin.
293 Microbiol. 31:184–190.

294 Huber P, Basso P, Reboud E, Attrée I. 2016. *Pseudomonas aeruginosa* renews its virulence factors.
295 Environ. Microbiol. Rep. [Internet]. Available from:
296 http://onlinelibrary.wiley.com/doi/10.1111/1758-2229.12443/abstract

297 Ikuta T, Takaki Y, Nagai Y, Shimamura S, Tsuda M, Kawagucci S, Aoki Y, Inoue K, Teruya M, Satou
298 K, et al. 2015. Heterogeneous composition of key metabolic gene clusters in a vent mussel
299 symbiont population. ISME J. [Internet]. Available from:
300 http://www.nature.com/doifinder/10.1038/ismej.2015.176

301 Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: A novel method for rapid multiple sequence
302 alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.

303 Kaufmann SHE, Dorhoi A. 2016. Molecular determinants in phagocyte-bacteria interactions. Immunity
304 44:476–491.

305 Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz
306 S, Duran C, et al. 2012. Geneious basic: An integrated and extendable desktop software
307 platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649.

Kendall MM, Sperandio V. 2016. What a dinner party! Mechanisms and functions of interkingdom signaling in host-pathogen associations. mBio 7:e01748-15.

Kuo C-H, Ochman H. 2009. The fate of new bacterial genes. FEMS Microbiol. Rev. 33:38–43.

Lang AE, Schmidt G, Schlosser A, Hey TD, Larrinua IM, Sheets JJ, Mannherz HG, Aktories K. 2010. *Photorhabdus luminescens* toxins ADP-Ribosylate actin and RhoA to force actin clustering. Science 327:1139–1142.

Li L. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res. 13:2178–2189.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Yunjie, et al. 2012. SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. GigaScience 1:18.

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, et al. 2014. CDD: NCBI's conserved domain database. Nucleic Acids Res.:gku1221.

Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, et al. 2011. IMG: The integrated microbial genomes database and comparative analysis system. Nucleic Acids Res. 40:D115–D122.

Martinez J, Cogni R, Cao C, Smith S, Illingworth CJR, Jiggins FM. 2016. Addicted? Reduced host resistance in populations with defensive symbionts. Proc. R. Soc. B Biol. Sci. 283:20160778.

McDonald KL. 2014. Rapid embedding methods into epoxy and LR White resins for morphological and immunological analysis of cryofixed biological specimens. Microsc. Microanal. 20:152–163.

McDowell E, Trump B. 1976. Histologic fixatives suitable for diagnostic light and electron microscopy. Arch. Pathol. Lab. Med. 100:405–414.

McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, Dubilier N, Eberl G, Fukami T, Gilbert SF, et al. 2013. Animals in a bacterial world, a new imperative for the life sciences. Proc. Natl. Acad. Sci. 110:3229–3236.

Meier DV, Pjevac P, Bach W, Hourdez S, Girguis PR, Vidoudez C, Amann R, Meyerdierks A. 2017. Niche partitioning of diverse sulfur-oxidizing bacteria at hydrothermal vents. ISME J.

Miller DJ, Hemmrich G, Ball EE, Hayward DC, Khalturin K, Funayama N, Agata K, Bosch T. 2007. The innate immune repertoire in Cnidaria-ancestral complexity and stochastic gene loss. Genome Biol 8:R59.

Montanaro J, Gruber D, Leisch N. 2016. Improved ultrastructure of marine invertebrates using non-toxic buffers. PeerJ 4:e1860.

Moore AD, Held A, Terrapon N, Weiner J, Bornberg-Bauer E. 2014. DoMosaics: Software for domain arrangement visualization and domain-centric analysis of proteins. Bioinformatics 30:282–283.

Moya A, Pereto J, Gil R, Latorre A. 2008. Learning how to live together: Genomic insights into prokaryote-animal symbioses. Nat Rev Genet 9:218–229.

Nelson DC, Hagen KD, Edwards DB. 1995. The gill symbiont of the hydrothermal vent mussel *Bathymodiolus thermophilus* is a psychrophilic, chemoautotrophic, sulfur bacterium. Mar. Biol. 121:487–495.

Nuismer SL, Otto SP. 2004. Host–parasite interactions and the evolution of ploidy. Proc. Natl. Acad. Sci. U. S. A. 101:11036–11039.

Nurk S, Meleshko D, Korobeynikov A, Pevzner P. 2016. metaSPAdes: A new versatile de novo metagenomics assembler. ArXiv160403071 Q-Bio [Internet]. Available from: http://arxiv.org/abs/1604.03071

Pel MJC, Pieterse CMJ. 2013. Microbial recognition and evasion of host immunity. J. Exp. Bot. 64:1237–1248.

Peng Y, Leung HCM, Yiu SM, Chin FYL. 2012. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28:1420–1428.

Pérez-Brocal V, Latorre A, Moya A. 2011. Symbionts and pathogens: What is the difference? In: Dobrindt U, Hacker JH, Svanborg C, editors. Between Pathogenicity and Commensalism. Vol. 358. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 215–243. Available from: http://link.springer.com/10.1007/82_2011_190

Petersen JM, Wentrup C, Verna C, Knittel K, Dubilier N. 2012. Origins and evolutionary flexibility of chemosynthetic symbionts from deep-sea animals. Biol. Bull. 223:123–137.

Ponnudurai R, Kleiner M, Sayavedra L, Petersen JM, Moche M, Otto A, Becher D, Takeuchi T, Satoh N, Dubilier N, et al. 2016. Metabolic and physiological interdependencies in the *Bathymodiolus azoricus* symbiosis. ISME J. [Internet]. Available from: http://www.nature.com/doifinder/10.1038/ismej.2016.124

Raggi L, Schubotz F, Hinrichs K-U, Dubilier N, Petersen JM. 2013. Bacterial symbionts of *Bathymodiolus* mussels and *Escarpia* tubeworms from Chapopote, an asphalt seep in the southern Gulf of Mexico. Environ. Microbiol. 15:1969–1987.

Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection: A toolbox for specialized analyses of microbial genomes and metagenomes. Available from: https://doi.org/10.7287/peerj.preprints.1900v1

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Roux S, Hawley AK, Beltran MT, Scofield M, Schwientek P, Stepanauskas R, Woyke T, Hallam SJ, Sullivan MB. 2014. Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta- genomics. Elife 3.

Rubin-Blum M, Antony CP, Borowski C, Sayavedra L, Pape T, Sahling H, Bohrmann G, Kleiner M, Redmond MC, Valentine DL, et al. 2017. Short-chain alkanes fuel mussel and sponge *Cycloclasticus* symbionts from deep-sea gas and oil seeps. Nat. Microbiol. 2:17093.

Sansonetti P. 2002. Host–pathogen interactions: The seduction of molecular cross talk. Gut 50:iii2–iii8.

Satchell KJF. 2011. Structure and function of MARTX toxins and other large repetitive RTX proteins. Annu. Rev. Microbiol. 65:71–90.

Sayavedra L, Kleiner M, Ponnudurai R, Wetzel S, Pelletier E, Barbe V, Satoh N, Shoguchi E, Fink D, Breusing C, et al. 2015. Abundant toxin-related genes in the genomes of beneficial symbionts from deep-sea hydrothermal vent mussels. eLife [Internet] 4. Available from: http://elifesciences.org/lookup/doi/10.7554/eLife.07966

389 Schmid MC, Schulein R, Dehio M, Denecker G, Carena I, Dehio C. 2004. The VirB type IV secretion
390 system of *Bartonella henselae* mediates invasion, proinflammatory activation and antiapoptotic
391 protection of endothelial cells. Mol. Microbiol. 52:81–92.

392 Seah BKB, Gruber-Vodicka HR. 2015. Gbtools: Interactive visualization of metagenome bins in R.
393 Microb. Physiol. Metab.:1451.

394 Seemann T. 2014. Prokka: Rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069.

395 Seubert A, Hiestand R, De La Cruz F, Dehio C. 2003. A bacterial conjugation machinery recruited for
396 pathogenesis. Mol. Microbiol. 49:1253–1266.

397 Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with
398 thousands of taxa and mixed models. Bioinformatics 22:2688.

399 Sun J, Zhang Yu, Xu T, Zhang Yang, Mu H, Zhang Yanjie, Lan Y, Fields CJ, Hui JHL, Zhang W, et
400 al. 2017. Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes.
401 Nat. Ecol. Evol. 1:0121.

402 Sylvan JB, Toner BM, Edwards KJ. 2012. Life and death of deep-sea vents: Bacterial diversity and
403 ecosystem succession on inactive hydrothermal sulfides. MBio 3:e00279-11.

404 Tellier A, Moreno-Gámez S, Stephan W. 2014. Speed of adaptation and genomic footprints of host–
405 parasite coevolution under arms race and trench warfare dynamics. Evolution 68:2211–2224.

406 Thomas T, Rusch D, DeMaere MZ, Yung PY, Lewis M, Halpern A, Heidelberg KB, Egan S, Steinberg
407 PD, Kjelleberg S. 2010. Functional genomic signatures of sponge bacteria reveal unique and
408 shared features of symbiosis. ISME J. 4:1557–1567.

409 Wentrup C, Wendeberg A, Schimak M, Borowski C, Dubilier N. 2014. Forever competent: Deep-sea
410 bivalves are colonized by their chemosynthetic symbionts throughout their lifetime. Environ.
411 Microbiol.:3699–3713.

412 Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: Interactive visualization of *de novo* genome
413 assemblies: Fig. 1. Bioinformatics 31:3350–3352.

414 Won YJ, Hallam SJ, O'Mullan GD, Pan IL, Buck KR, Vrijenhoek RC. 2003. Environmental acquisition
415 of thiotrophic endosymbionts by deep-sea mussels of the genus *Bathymodiolus*. Appl. Environ.
416 Microbiol. 69:6785.

417 Zhou J, Bruns MA, Tiedje JM. 1996. DNA recovery from soils of diverse composition. Appl. Environ.
418 Microbiol. 62:316–322.

Figures

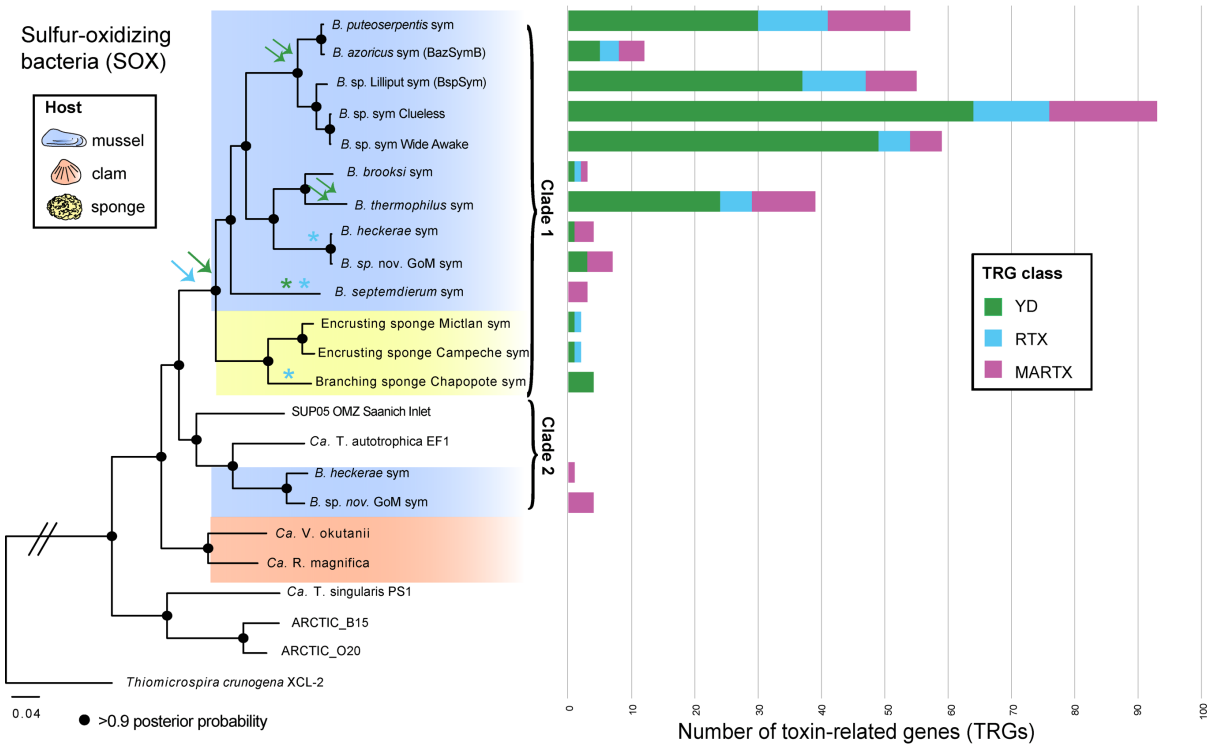

Fig. 1. Phylogenomic tree of symbiotic and free-living SOX, estimated with 38 orthologous protein-coding genes, and the corresponding distribution of TRGs in these SOX. Filled circles represent a posterior probability higher than 0.9. The blue arrow indicates the proposed acquisition of RTX genes, green arrow of YD-repeat genes. All genomes shown in this tree were searched for all TRGs. Single arrows represent possible gene acquisition; double arrows indicate possible gene duplications; asterisks show possible gene loss events. The color of the arrows and stars corresponds to the TRG class. The number of individuals sequenced per species and geographic location is shown in Supplementary Table 1. T = *Ca*. Thioglobus; B = *Bathymodiolus*; sym = symbiont.

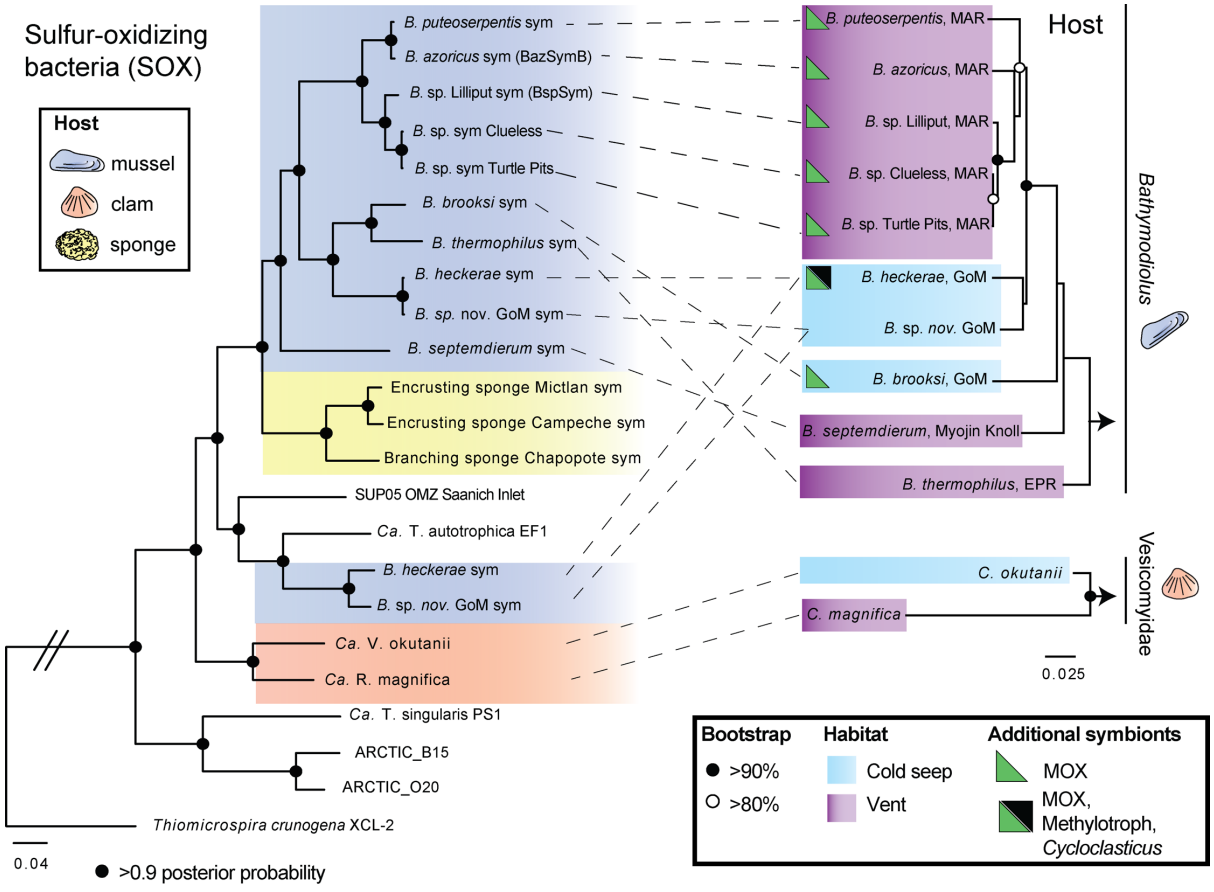

Fig. 2. Bivalve phylogeny and symbiont phylogeny. The maximum-likelihood host phylogeny was reconstructed based on cytochrome oxidase I (COI). Symbiont phylogeny was estimated with 38 orthologous protein-coding genes.

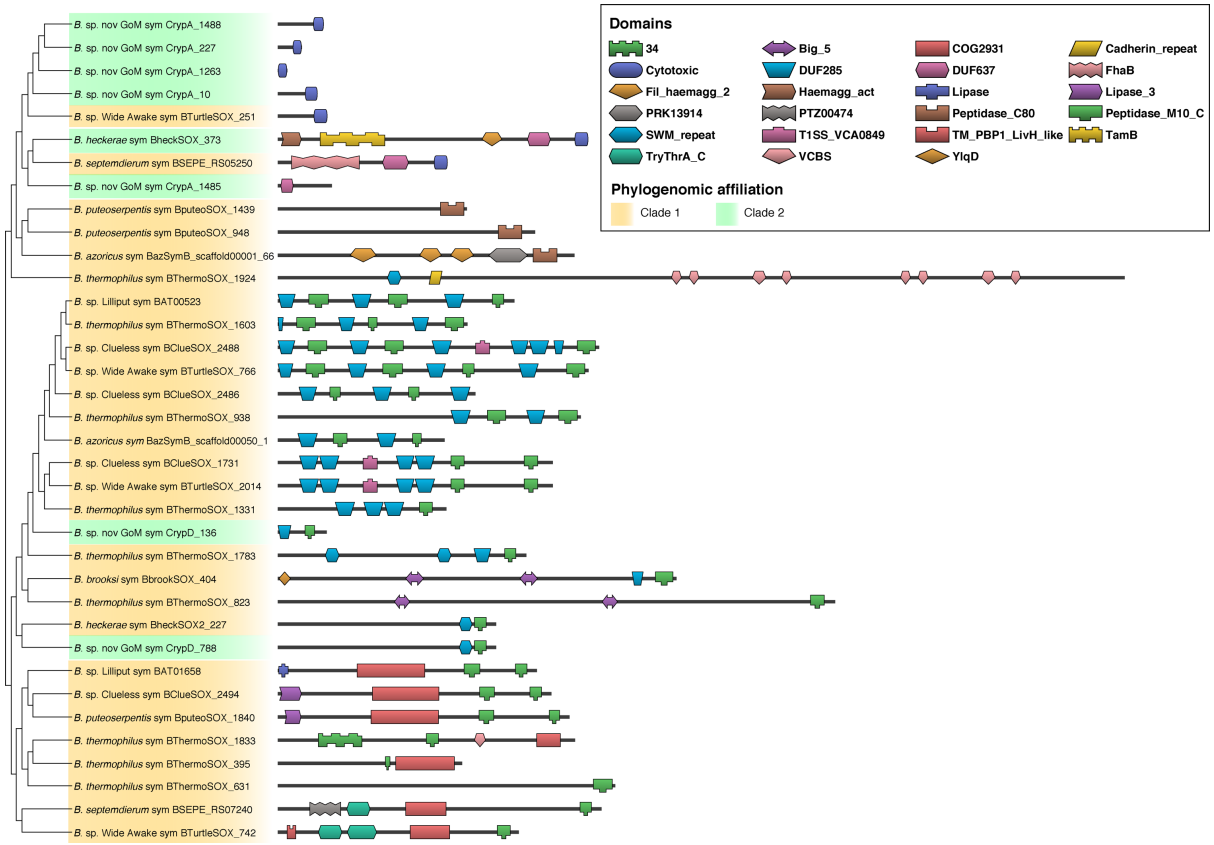

Fig. 3. Domain structure of MARTX-like genes that have the most similar domain architecture to the MARTX-like genes from Clade 2 SOX symbionts (Fig. 1). The tree was estimated with the domain distance among proteins with DoMosaics (Moore et al. 2014). The descriptions of these domains are available in Supplementary Table 2. Locus tags of the MARTX genes from the SOX symbionts are shown at the nodes.

Tables

Table 1. Overview of sulfur-oxidizing bacteria analyzed in this study.

Organism		Sampling site	Ecosystem	GC content (%)	Approx. Sequencing Depth (X)	Completeness (%)	Genome Size (Mbp)	No. of scaffolds (>1000 bp)
Host-associated	*B. azoricus* symbiont[1]	Menez Gwen, MAR	Vent	38.20	8	90.60	1.66	239
	B. sp. *9°* South, Lilliput symbiont[1]	9°S, Lilliput, MAR	Vent	38.23	22	95.39	2.29	52
	B. thermophilus symbiont[+]	Crab-Spa, EPR	Vent	38.4	199	97.86	2.25	149
	B. puteoserpentis symbiont[+]	Logatchev, MAR	Vent	37.67	3600	97.7	2.19	77
	B. sp. 5° South, Clueless symbiont[+]	5°S, Clueless, MAR	Vent	37.81	102	98.52	2.43	383
	B. sp. 5° South, Wide Awake symbiont[+]	5°S, Wide Awake, MAR	Vent	37.76	226	96.55	2.54	382
	B. heckerae symbiont (Clade 2)[+]	Chapopote, GoM	Seep	37.41	557	96.58	1.96	236
	B. heckerae symbiont (Clade 1)[+]	Chapopote, GoM	Seep	38.82	223	97.19	1.49	110
	B. brooksi symbiont[+]	Chapopote, GoM	Seep	36.63	187	97.2	2.82	374
	B. septemdierum symbiont[2]	Myojin Knoll	Vent	38.74	505	98.68	1.47	1
	B. sp. *nov* GoM symbiont (Clade 2)[+,*]	DC673, GoM	Seep	36.91	103	98.01	2.19	322
	B. sp. *nov* GoM symbiont (Clade 1)[+,*]	DC673, GoM	Seep	38.77	26	90.83	1.41	168
	C. okutani[3]	Sagami Bay	Seep	31.59	-	93.58	1.02	1
	C. magnifica[4]	9°N, EPR	Vent	34.03	-	94.84	1.16	1
	Encrusting sponge symbiont[9]	Mictlan, GoM	Seep	38.77	50	96.89	2.20	207
	Encrusting sponge symbiont[9]	Chapopote, GoM	Seep	38.71	50	96.03	2.93	378
	Branching sponge symbiont[9]	Chapopote, GoM	Seep	39.19	500	95.2	2.09	105
Free-living	SUP05[5]	Saanich Inlet	OMZ	39.29	-	85.76	1.37	97
	Candidatus T. singularis PS1[6]	Puget Sound	Pelagic (5 m depth)	37.44	-	98.68	1.71	1
	Candidatus T. autotrophica EF1[7]	Effingham Inlet	Pelagic, redox gradient (60 m depth)	39.14	-	99.18	1.51	1
	Thiomicrospira crunogena XCL-2[8]	EPR	Vent	43.13	-	100	2.43	1

B., *Bathymodiolus*; *C*., *Calyptogena*; T, Thioglobus; GoM, Gulf of Mexico; EPR, East Pacific Rise; MAR, Mid-Atlantic Ridge; OMZ, oxygen minimum zone.

*Symbionts of this species were characterized in this study (SI Results and Discussion); +Sequenced in this study; [1](Sayavedra *et al*., 2015); [2](Ikuta *et al.*, 2015); [3](Kuwahara *et al*., 2007); [4](Newton *et al*., 2007); [5](Walsh *et al*., 2009); [6](Marshall and Morris, 2015); [7](Shah and Morris, 2015); [8](Scott *et al*., 2006); [9]Sequenced by Rubin-Blum et al., (2017), assembled in this study.