Title

Genomic insights into specialized metabolism in the marine actinomycete Salinispora

Journal

Environmental Microbiology, 19(9)

ISSN

1462-2912

Authors

Letzel, Anne-Catrin
Li, Jing
Amos, Gregory CA
et al.

Publication Date

2017-09-01

DOI

10.1111/1462-2920.13867

Peer reviewed

environmental microbiology

Environmental Microbiology (2017) **19**(9), 3660–3673 doi:10.1111/1462-2920.13867

Genomic insights into specialized metabolism in the marine actinomycete *Salinispora*

Anne-Catrin Letzel,[1] Jing Li,[1†] Gregory C.A. Amos,[1] Natalie Millán-Aguiñaga,[1,2] Joape Ginigini,[3] Usama R. Abdelmohsen,[4,5] Susana P. Gaudêncio,[6] Nadine Ziemert,[1,7] Bradley S. Moore[1,8,9] and Paul R. Jensen[1,9*]

[1]*Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, 9500 Gilman Dr, La Jolla, CA 92093, USA.*

[2]*Facultad de Ciencias Marinas, Universidad Autónoma de Baja California, Ensenada, Baja California 22800, Mexico.*

[3]*Institute of Applied Sciences, Faculty of Science, Technology and Environment, University of the South Pacific, Laucala Campus, Private Mail Bag, Suva, Fiji.*

[4]*Department of Botany II, Julius-von-Sachs Institute for Biological Sciences, University of Würzburg, Germany.*

[5]*Department of Pharmacognosy, Faculty of Pharmacy, Minia University, Minia 61519, Egypt.*

[6]*Department of Chemistry, REQUIMTE, LAQV and UCIBIO, Faculty of Science and Technology, Universidade NOVA de Lisboa, Caparica 2529-516, Portugal.*

[7]*Interfaculty Institute of Microbiology and Infection Medicine Tuübingen, University of Tuübingen, Auf der Morgenstelle 28, Tuübingen 72076, Germany.*

[8]*Skaggs School of Pharmacy and Pharmaceutical Sciences, University of 9500 Gilman Dr, La Jolla, CA 92093, USA.*

[9]*Center for Microbiome Innovation, University of California San Diego, La Jolla, CA 92093, USA.*

Summary

Comparative genomics is providing new opportunities to address the diversity and distributions of genes encoding the biosynthesis of specialized metabolites. An analysis of 119 genome sequences representing three closely related species of the marine actinomycete genus *Salinispora* reveals extraordinary biosynthetic diversity in the form of 176 distinct biosynthetic gene clusters (BGCs) of which only 24 have been linked to their products. Remarkably, more than half of the BGCs were observed in only one or two strains, suggesting they were acquired relatively recently in the evolutionary history of the genus. These acquired gene clusters are concentrated in specific genomic islands, which represent hot spots for BGC acquisition. While most BGCs are stable in terms of their chromosomal position, others migrated to different locations or were exchanged with unrelated gene clusters suggesting a plug and play type model of evolution that provides a mechanism to test the relative fitness effects of specialized metabolites. Transcriptome analyses were used to address the relationships between BGC abundance, chromosomal position and product discovery. The results indicate that recently acquired BGCs can be functional and that complex evolutionary processes shape the micro-diversity of specialized metabolism observed in closely related environmental bacteria.

Received 12 April, 2017; revised 18 July, 2017; accepted 21 July, 2017. *For correspondence. E-mail: pjensen@ucsd.edu; Tel. (+01) 858-534-7322; Fax (+01) 858-534-1318. †Present address: College of Marine Life Sciences, Ocean University of China, Qingdao, Shandong, 266003, People's Republic of China.

Introduction

Microorganisms produce small organic compounds called secondary or specialized metabolites. These compounds have been the source of or inspiration for many of today's most important medicines (Berdy, 2005; Newman and Cragg, 2012) and represent a fundamental mechanism by which microorganisms interact with each other and the environment. While the ecological functions of most specialized metabolites remain unknown, they mediate important processes such as quorum sensing (Waters and Bassler, 2005), metal acquisition (Sandy and Butler, 2009) and allelopathy (Wietz *et al.*, 2013). Once considered non-essential for survival, it is now known that specialized metabolites are involved in fundamental physiological processes such as electron transport, raising questions about their 'secondary' nature (Price-Whelan *et al.*, 2006).

Advances in our understanding of the molecular genetics of natural product biosynthesis, coupled with increased

access to genome sequence data, are providing unique opportunities to identify the biosynthetic gene clusters (BGCs) that encode the production of specialized metabolites (Fischbach and Walsh, 2006; Hertweck, 2009). In addition to encoding the enzymes needed for biosynthesis, these gene collectives can also include regulatory elements, transporters and mechanisms of resistance (Cimermancic *et al*., 2014). BGCs account for $>10\%$ of some bacterial genomes (Nett *et al*., 2009; Cimermancic *et al*., 2014), indicating the fundamental importance of this type of metabolism for certain bacteria. The analysis of bacterial genome sequences has led to the surprising observation that even well studied strains can maintain many more BGCs than the compounds discovered from them would suggest (Bentley *et al*., 2002; Udwary *et al*., 2007). BGCs that have not been linked to their products have been termed 'orphan' (Gross, 2007) and are often thought to be transcriptionally inactive or 'silent' under normal laboratory conditions (Seyedsayamdost, 2014; Abdelmohsen *et al*., 2015). Efforts to identify the products of these orphan BGCs is driving a resurgence of interest in natural product research (Winter *et al*., 2011; Kim *et al*., 2015) and the development of techniques such as genome mining (Challis, 2008; Ziemert *et al*., 2016), which take a sequence first approach to natural product discovery. With the increasing availability of online tools (Boddy, 2013; Weber, 2014) and data repositories (Medema *et al*., 2015), BGCs can now be readily identified from sequence data without specific expertise in this form of metabolism.

The BGCs encoding specialized metabolites are said to be among the most rapidly evolving genetic elements known (Fischbach *et al*., 2008). Access to genome sequences from large numbers of closely related strains creates opportunities to compare similar BGCs in different strains and begin to infer the evolutionary processes that create chemical diversity in nature (Freel *et al*., 2011). There is substantial evidence that BGCs are horizontally exchanged among bacteria (Egan *et al*., 1998; Metsa-Ketela *et al*., 2002; Ziemert *et al*., 2014), although the frequency at which this occurs and the mechanisms driving these events remain poorly understood. Nonetheless, horizontal gene transfer is widely recognized as a driving force in bacterial evolution (Ochman *et al*., 2000) and the acquisition of a gene cluster encoding the production of a specialized metabolite provides immediate opportunities to test the effects of the encoded compound on fitness.

The marine actinomycete genus *Salinispora* has proven to be a useful model to address questions related to specialized metabolism and BGC evolution (Jensen *et al*., 2015; Jensen, 2016). The genus is currently comprised of three named species: *S. tropica, S. arenicola* and *S. pacifica* (Maldonado *et al*., 2005; Ahmed *et al*. 2013), which share 99% 16S rRNA sequence identity. They are broadly distributed in ocean sediments and the source of salinosporamide A, a potent proteasome inhibitor that has completed a series of phase I clinical trials for the treatment of cancer (Harrison *et al*., 2016). Evidence that most BGCs were observed infrequently among *Salinispora* strains was used to infer relatively recent acquisition in the evolutionary history of the genus (Ziemert *et al*., 2014). These BGCs were concentrated in genomic islands (Penn *et al*., 2009), flexible regions of the genome enriched in acquired genes that generally encode adaptive traits (Juhas *et al*., 2009). While the integration of BCGs into GIs minimizes the risk of disrupting important cellular pathways (Dobrindt *et al*., 2004), it remains unknown if recently acquired BGCs are drawn to certain genomic islands or if their frequency of occurrence in the population or location on the chromosome affects the likelihood they will be expressed. Furthermore, population genomics provides opportunities to address BGC chromosomal stability, exchange events, and the possibility that some BGCs are degraded over time in addition to simply being gained or lost. To address these questions, we analysed 119 high quality *Salinispora* genome sequences for BGCs associated with specialized metabolism. The goals were to assess BGC diversity and distributions in the context of species-level 'pan-chromosomes' (Chan *et al*., 2015) and to gain new insight into the dynamic nature of BGC evolution.

Results and discussion

BGC diversity, distributions and product discovery

High quality draft genome sequences generated from 119 strains of the marine actinomycete genus *Salinispora* were analysed for BGCs associated with specialized metabolism (Supporting Information Table S1). These strains were largely cultured from marine sediment samples collected from global locations over a 25-year time frame and represent a subset of a larger culture collection that has yielded a wide range of structurally diverse specialized metabolites (Jensen *et al*., 2015). The average number of BGCs identified per genome was significantly greater in *S. arenicola* (25) than in the other two species (Supporting Information Fig. S1, ANOVA, $p = 0.00$), providing genetic support for the observation that this species preferentially invests in interference competition (Patin *et al*., 2015). Further distinguishing the species, we confirmed that *S. tropica* strains consistently maintain four BGCs predicted to encode siderophores (*des*, sid1, sid3 and sid4) relative to two on average in *S. arenicola* (Roberts *et al*., 2012). The specialized metabolites discovered to date from the genus are enriched in compounds of PKS, NRPS and PKS/NRPS hybrid origins (78%) relative to their occurrence in the genomes (48%–60%) (Supporting Information Fig. S1), indicating a bias

Table 1. Compounds and their associated BGCs detected in *Salinispora* strains (numbering as defined in Fig. 1).

No.	Compound	BGC	Linkage	References
16	Arenicolide	PKS28	B	(Williams *et al*., 2007)
20	Arenimycin	arn	B	(Kersten *et al*., 2013b)
24	BE-43547 (APD-CLD)	PKS58	B	(Villadsen *et al*., 2016)
8	Cyanosporaside	cya	G	(Lane *et al*., 2013)
19	Cyclomarin/cyclomarazine	cym	G	(Schultz *et al*., 2008)
3	Desferrioxamine	des	G	(Roberts *et al*. 2012)
17	Enterocin	ent	H	(Bonet *et al*., 2014)
12	Ikarugamycin	ika	B	(Greunke *et al*., 2017)
8	Lomaiviticin	lom	G	(Kersten *et al*., 2013a)
2	Lymphostin	lym	G	(Miyanaga *et al*. 2011)
14	Pacificanone/salinipyrone	spr	G	(Awakawa *et al*., 2015)
23	Retimycin	rtm	B	(Duncan *et al*., 2015)
5	Rifamycin/saliniketal	rif	G	(Wilson *et al*., 2010)
11	Salinichelin	slc	B	(Unpublished)
9	Salinilactam	slm	B	(Udwary *et al*. 2007)
22	Saliniquinone	PKS65	B	(Unpublished)
10	Salinosporamide	sal	G	(Eustáquio *et al*., 2011)
1	Sioxanthin	sio	G	(Richter *et al*., 2015)
13	Sporolide	spo	H	(McGlinchey et al., 2008)
4	Staurosporine	sta	B	(Freel *et al*., 2011)
21	Tambromycin	tbr	B	(Goering *et al*., 2016)
6	Isopimara-8,15-dien-19-ol	terp1	H	(Xu *et al*., 2014)
15	Thiolactomycin	tlm	H	(Tang *et al*., 2015)
18	Tirandalydigin	tdy	B	(Castro-Falcón *et al*., 2016)
25	Mevinolin	NI	-	(Bose *et al*., 2014)
26	Salinipostin	NI	-	(Schulze *et al*., 2015)
27	Arenamide	NI	-	(Asolkar *et al*., 2009)

In the case of BE-43547, the biosynthetic gene cluster (BGC) was detected in *Salinispora* but the compound originally reported from another taxon. Compounds for which the BGCs are unknown are listed as not identified (NI). The method by which compounds and BGCs were linked is reported as bioinformatic (B), gene inactivation (G), or heterologous expression (H).
Note: The pseudopeptide ketomemicin C and its associated BGC was recently identified following heterologous expression from *S. tropica* strain CNB-440 (Ogasawara *et al*., 2016) but not included in these analyses.

towards the isolation of compounds produced by these biosynthetic mechanisms.

Despite sharing 99% 16S rRNA gene sequence identity, we observed extensive diversity in the 119 genomes in the form of 176 distinct BGCs. Of these, it was possible to link 14 to the small molecules they encode based on prior experimental evidence and 10 more based on bioinformatic analysis (*Table* 1). Thus, as has been observed in other actinomycete genomes, a majority of the BGCs detected (80% in this case) have not been linked their products. Despite extensive chemical analysis of this genus (Jensen *et al*., 2015), only three additional compounds have been isolated yet not connected to their BGCs (Asolkar *et al*., 2009; Bose *et al*., 2014; Schulze *et al*., 2015) indicating that a large majority of the biosynthetic potential of the strains sequenced has yet to be realized. *Salinispora* spp. devote a large percentage of their genomes to specialized metabolism ($>$ 10%) (Udwary *et al*., 2007) supporting the importance of these small molecules to their biology.

We next assessed the relative abundance of each BGC in the three *Salinispora* species (Supporting Information Fig. S2). Surprisingly, only four BGCs are conserved at the genus level (Fig. 1A). Of these, only the *sio* BGC has been linked to its biosynthetic product, the carotenoid pigment sioxanthin, which gives these bacteria their characteristic orange pigmentation (Richter *et al*., 2015). At the species level, *S. arenicola* possesses a relatively large number of BGCs that were observed among all strains including six that were not observed in the other two species. This level of conservation among 62 strains isolated from global locations over 25 years provides evidence they are maintained due to strong selective pressures. The products of the BGCs conserved in all *S. arenicola* strains encompass diverse biosynthetic types and include potent antibiotics (rifamycins), cytotoxins (staurosporines) and a diterpene of unknown function discovered via heterologous expression (Fig. 1B) (Jensen *et al*., 2015).

While relatively few BGCs are fixed at the genus or species level, 54% were observed in only one or two strains. More specifically, 61 BGCs were observed in one strain (singletons) while an additional 34 were observed in only two of the 119 strains (Fig. 1, Supporting Information Fig. S3). It can be inferred that these 95 BGCs were acquired

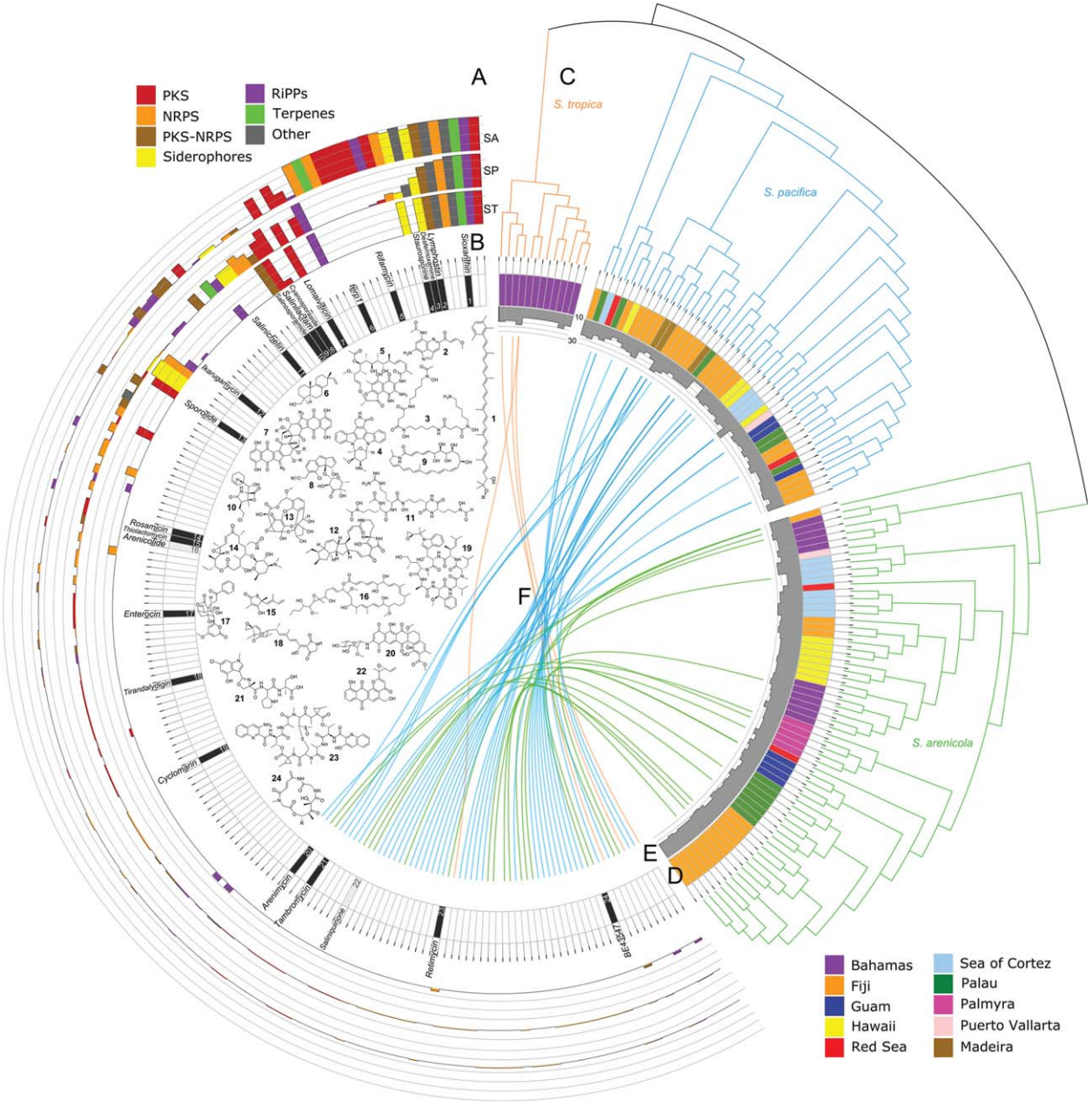

Fig. 1. *Salinispora* specialized metabolism.
A. Outer three circular bar graphs show the relative abundance of each BGC in the three *Salinispora* species arranged counter-clockwise in decreasing order by genus (scale = 0%–100% of strains, lines at 25% intervals, SA = *S. arenicola*, SP = *S. pacifica*, ST = *S. tropica*). Bar colours indicate BGC type (see legend). Only the first four BGCs occur in all 176 strains.
B. Inner circle shows the BGCs for which the products are confirmed (black) or predicted (grey) with the corresponding chemical structures numbered and named: **1** sioxanthin, **2** lymphostin, **3** desferrioxamine B, **4** staurosporine, **5** rifamycin S, **6** diterpene, **7** lomaiviticin A, **8** cyanosporaside A, **9** salinilactam, **10** salinosporamide A, **11** salinichelin A, **12** ikarugamycin, **13** sporolide A, **14** rosamicin A, **15** thiolactomycin, **16** arenicolide B, **17** enterocin, **18** tirandalydigin, **19** cyclomarin D, **20** arenimycin A, **21** tambromycin, **22** saliniquinone A, **23** retimycin A, **24** BE43547A1.
C. Species phylogeny based on 876 concatenated genes (Millán-Aguiñaga et al., 2017).
D. Location from which the strain was isolated (see legend).
E. Histogram shows number of BGCs per strain (scale = 10–30).
F. Plot linking strains on the right with BGCs on left. Coloured links show 'singletons' with the corresponding species (orange = *S. tropica*, blue = *S. pacifica*, green = *S. arenicola*).

by horizontal gene transfer relatively recently in the evolutionary history of the genus thus providing support for the concept that BGCs represent highly mobile genetic elements (Fischbach *et al.*, 2008). The singletons are broadly distributed throughout the phylogeny (Fig. 1C and F) indicating that BGC acquisition is not limited to any specific clades and thus appears to be a common feature of the genus. The remarkable BGC diversity observed among a group of closely related strains represents a major component of the flexible genome and creates a theoretical mechanism by which bacteria can respond to new selective pressures or exploit new environments (Jensen, 2016; Rodriguez-Valera *et al.*, 2016).

The numbers of shared BGCs did not decrease with increasing geographic distance at either the genus or species levels (Supporting Information Fig. S4 and Table S2), as would be expected if there were biogeographic barriers to dispersal. Nonetheless, there are examples of clade-specific BGCs that were only observed at a single location (Supporting Information Fig. S2), suggesting that endemism plays a role in BGC distributions. There is also evidence that BGCs such as PKS9 were acquired independently by different strains, in this case on three occasions by five strains from Fiji (Supporting Information Fig. S5). While these observations are likely affected by sample size, they provide evidence of the complex evolutionary histories of BGCs relative to the strains in which they reside.

Islands of innovation

We next asked if there was conservation in the chromosomal positions of BGCs and if this varied depending on their frequency of occurrence among the strains. To address these questions, we generated pseudo-chromosomes for each strain by mapping sequence contigs onto the closed genomes of a closely related strain and then mapping the position of each BGC onto the assembly. The positions of ca. 75% of the BGCs could be assigned using this approach. We then generated a single pan-chromosome for each species and determined the positional conservation for each BGC (Fig. 2A–C). Surprisingly, the 21 genomic islands originally identified in a comparison of two *Salinispora* genome sequences (Penn *et al.*, 2009) were highly conserved among all 119 strains. Depending on species, these genomic islands housed 72%–89% of the mapped BGCs.

While BGCs are commonly observed in genomic islands (Dobrindt *et al.*, 2004), we asked if acquisition events are targeted to specific genomic islands. In *S. pacifica*, the pan-chromosome clearly reveals that GI10 and GI15 are enriched in BGCs, with 54% of those observed in only one or two strains located in these two islands. Similarly, in *S. arenicola*, GI5, GI15 and GI20 account for 83% of the mapped BGCs observed in only one or two strains. These islands appear to be specialized sites for BGC entry and are located in regions comparable to the chromosomal arms identified in *Streptomyces* linear genomes, which are enriched in BGCs (Bentley *et al.*, 2002; Nett *et al.*, 2009). While the mechanisms directing BGCs to specific GIs remain unknown, these regions contain 79% of the mobile genetic elements that could be mapped onto the pseudo-chromosomes. In particular, GI10 and GI15 are enriched in transposases in all three species, which mirrors the abundance of BGCs in these GIs (Supporting Information Fig. S6). These findings contrast with the location of BGCs in the fungal genus *Penicillium* were they are mainly encoded in the core genome (Nielsen *et al.*, 2017).

Given that numerous specialized metabolites have been reported from *Salinispora* strains, we next asked if the position of the BGC affected the likelihood that the encoded secondary metabolites have been discovered. Of the 24 BGCs that have been experimentally or bioinformatically linked to their products, 18 could be mapped onto the three pan-chromosomes. The positions of these BGCs ranged from highly conserved core regions (e.g., *lym* between GI16-GI17 and *des* between GI14-GI15) to genomic islands such as GI10 and GI15, which are among the most active sites of BGC acquisition (Fig. 2A–C). Products have also been identified from BGCs that are both highly conserved (e.g., *sio*, observed in all 119 genomes) and rare (e.g., *rtm* and *arn*, each observed in one genome). These results provide evidence that recently acquired BGCs can be functional and suggest that location does not affect the likelihood of product discovery. To further explore the effects of genome location on BGC functionality, we asked if certain GIs were hot spots for BGC expression (Fig. 2D–F). For GIs that contain more than one BGC, they were differentially expressed 53% of the time, suggesting that most BGCs are under different regulatory control. Examples include GI10 in *S. pacifica* where *STPKS1* was highly expressed relative to *lan3B*. There were also examples where BGCs located in the same GIs in different species were differentially expressed (e.g., *sio* and NRPS4). These observations suggest that genomic position is not a good predictor of BGC expression. Nonetheless, a majority of BGCs (79%) in GI10 and GI15, the major sites of BGC acquisition, were expressed at low levels in the one cultivation condition tested.

BGC migration

The positions of BGCs are largely stable among the strains sequenced, with 92% of those observed in more than one strain found in the same chromosomal position in all strains (Fig. 2A–C). However, the remaining 8% of BGCs could be mapped to more than one position in different strains, providing insight into the complex evolutionary

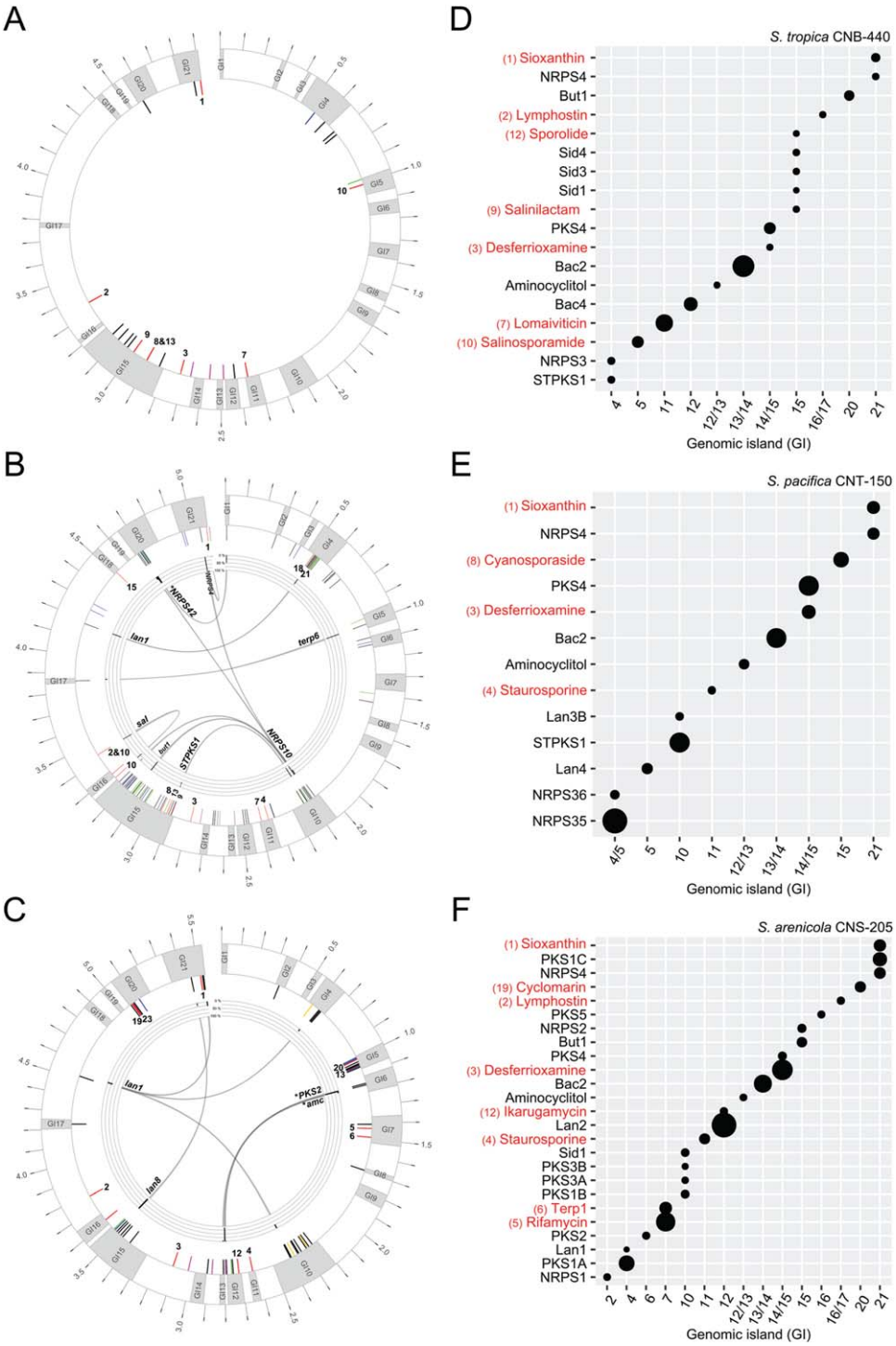

Fig. 2. Pan-chromosomes. (A) *S. tropica*, (B) *S. pacifica* and (C) *S. arenicola* drawn with dnaA at the 12:00 position. Outer ring: genomic islands are shown in grey (labelled GI1–21), outer scale indicates size in Mb. Inner ring: positions of mapped BGCs. Those with known products are in red, genus-specific BGCs in pink, *S. arenicola* specific BGC in yellow, doubletons in green and singletons in blue. Inner ring: percent occurrence of each BGC at that site. Numbers correspond to compounds experimentally linked to their respective BGC as defined in Fig. 1. Expression profiles of BGCs in (D) *S. tropica* CNB-440, (E) *S. pacifica* CNT-150 (DSM 45549) and (F) *S. arenicola* CNS-205. Size of circle corresponds to level of expression in stationary phase. Only expressed BGCs (those above the threshold) are indicated with a circle. BGCs with known products are shown in red (numbering as defined in Fig. 1). Genomic position (*x*-axis) is indicated by GI number or interisland location.

histories of some BGCs within the genus. In all cases, these positional changes were associated with different islands, suggesting the BGCs were either independently acquired or migrated from island to island. The island migration hypothesis is supported by likelihood and phylogenetic analyses in *NRPS42* and *NRPS4* in *S. pacifica* (Supporting Information Figs S7 and S8) and *amc* and *PKS2* in *S. arenicola* (Supporting Information Figs S9 and S10). In all four cases, the likelihood and phylogenetic analyses predict that the BGCs were acquired once and migrated to different chromosomal positions. In the seven remaining cases where a BGC was observed in different chromosomal positions in different strains (*sal*, *STPKS1*, *NRPS10*, *lan1*, *lan8*, *terp6* and *but1*), the most

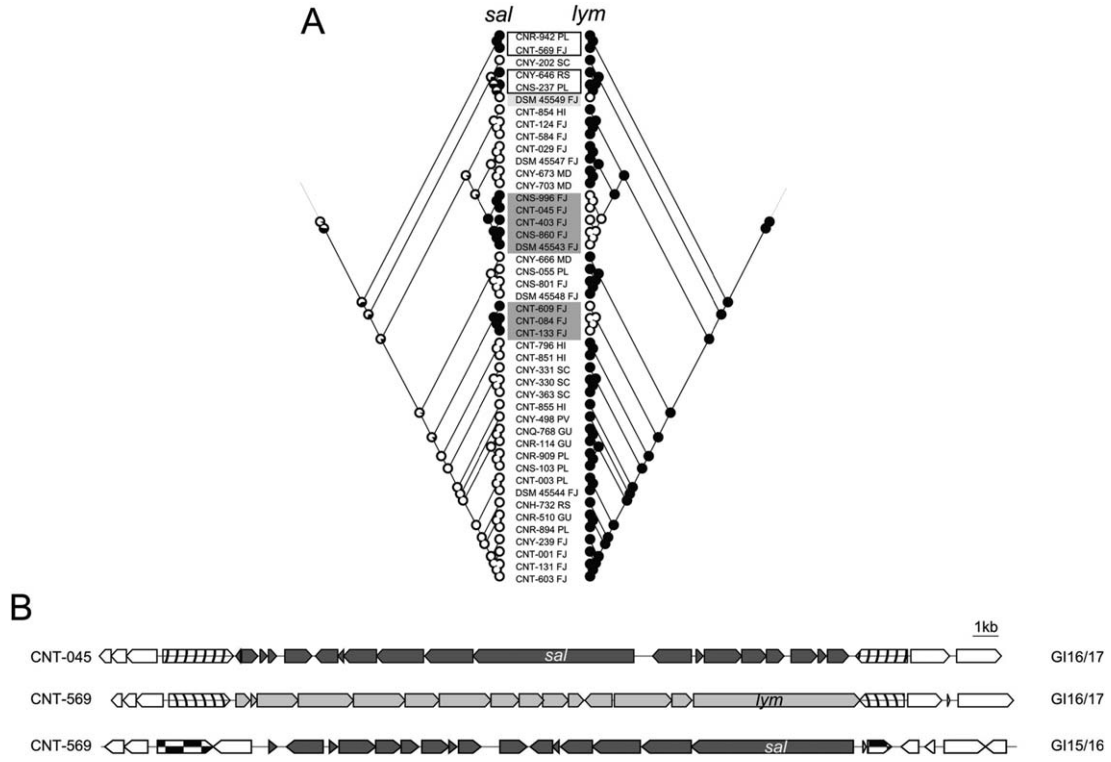

Fig. 3. Exchange of *sal* and *lym* BGCs in *S. pacifica*.
A. Mirrored likelihood analyses of *sal* (left) and *lym* (right) BGCs in *S. pacifica*. These analyses predict that *lym* was present in the common ancestor of the genus while *sal* was acquired more recently in at least four separate events. Grey shaded boxes indicate strains in which the *lym* BGC was replaced by *sal*. Black un-shaded boxes indicate four strains that have both BGCs (CNR-942, CNT-569, CNY-646 and CNS-237), whereas the light grey box indicates the only strain lacking both BGCs (DSM 45549).
B. For strains containing either *sal* or *lym*, the BGCs are located in the same inter-island region between GI16 and GI17 as exemplified here by *sal* in strains CNT-045 and *lym* in CNT-569. For strains containing both *sal* and *lym*, *lym* is located in the normal conserved position (GI16/GI17) while *sal* is located in the inter-island region between GI15 and GI16 as exemplified by strain CNT-569. SC: Sea of Cortez, PV: Puerto Vallarta, RS: Red Sea, FJ: Fiji, HI: Hawaii, PM: Palmyra, GU: Guam, PL: Palau, MD: Madeira archipelago).

parsimonious explanation is that it was acquired more than once (Supporting Information Figs S11–S21, Fig. 4A). In total, 11 BGCs varied in their genomic positions. Sixty-four percent of these could be linked to independent acquisition events and 36% to intragenomic migration events. It remains unknown if BGCs observed in core regions of the genome originally entered via genomic islands and subsequently migrated to these conserved regions where they are less likely to be lost or disrupted by subsequent acquisition events (Dobrindt *et al.*, 2004).

BGC exchange

As previously reported for enediyne BGCs (Ziemert *et al.*, 2014), we observed clear examples of BGC exchange between the highly conserved lymphostin (*lym*) BGC (Miyanaga *et al.*, 2011), which was observed in 110 of 119 genomes and the salinosporamide K version of the *sal* BGC (Supporting Information Fig. S21) (Eustáquio *et al.*, 2011), which was observed in eight of the nine *S. pacifica* strains that lacked the *lym* BGC (Fig. 3A). Likelihood analyses predict that *lym* was vertically inherited from an *S. pacifica* ancestor and subsequently maintained in a majority of strains while *sal* was acquired more recently in at least four independent events. In all eight cases where *sal* replaces *lym*, the BGCs occur in the same chromosomal position (Fig. 3B) supporting the concept that these BGCs were exchanged. Interestingly, four strains maintain both the *sal* and *lym* BGCs and in all cases, *lym* is observed in its conserved genomic position while *sal* has been inserted in the core region between GI15 and 16. Unlike previous observations of enediyne BGC exchange, genes within *sal* and *lym* do not share high levels of sequence identity suggesting that homologous recombination is not involved. There is also evidence for the exchange of *Salinispora* BGCs that encode siderophores (Ziemert *et al.*, accepted). It would be interesting to know if the products of *sal* and *lym* were functionally equivalent in the strains that possess them.

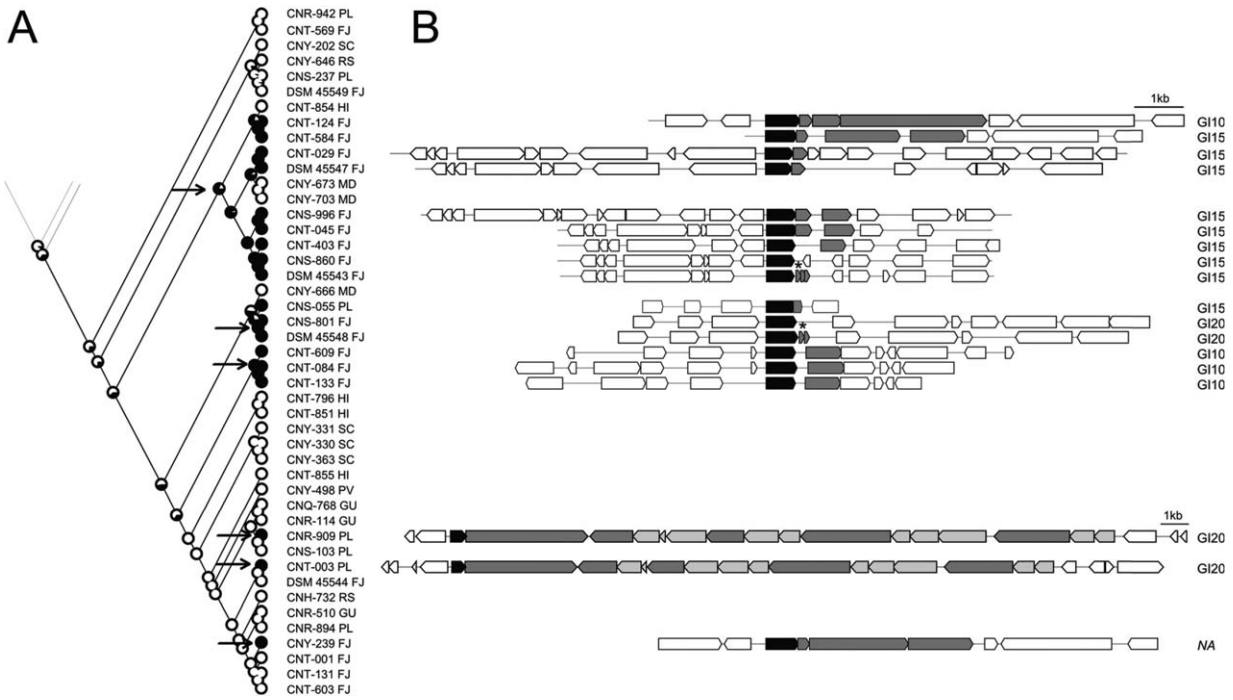

Fig. 4. NRPS10 degradation in *S. pacifica*.
A. Likelihood analysis for the conserved regulatory gene predicts six independent acquisition events (arrows). Abbreviations after strain number indicate isolation site of the strain (PL: Palau, FJ: Fiji, RS: Red Sea, HI: Hawaii, MD: Madeira archipelago, SC: Sea of Cortez, PV: Puerto Vallarta, GU: Guam).
B. NRPS10 gene cluster organization. The 'complete' BGC is observed in CNT-003 and CNR-909. The conserved regulator gene is shown in black, NRPS related genes in dark grey, tailoring enzymes in light grey and flanking genes in white. Position of BGC in genome shown next to cluster. GI: genomic island, NA: GI could not be assigned, *annotated as pseudogene.

BGC degradation

In addition to BGCs being observed in different chromosomal positions, there is evidence that some are degraded. We searched for BGC remnants by extracting all *Salinispora* genes annotated as 'pseudogenes' from the IMG/ER database and screening this pool for annotations related to secondary metabolism. This led to the detection of pseudogene *salpac_2984* in *S. pacifica* strain DSM 45543, which was also annotated as a 'condensation-domain containing protein'. This 139-amino acid pseudogene shares 94% sequence identity with a region of the 1420 amino acid NRPS protein from NRPS10, a BGC identified in *S. pacifica* strains CNT-003 and CNR-909. Using this gene as a BLASTP query in IMG/ER, proteins with $>$ 90% sequence identity to *salpac_2984* were identified in 14 additional *Salinispora* genomes. In addition to this NRPS fragment, a putative regulator gene also associated with NRPS10 was linked to the *Salpac_2984* pseudogene. Using the regulator gene as a BLASTP template, three additional strains containing portions of NRPS10 but lacking the *Salpac_2984* pseudogene could also be identified. Detailed analyses of the strains that only retained the regulator gene showed that parts of the NRPS could also be detected; however, they were not annotated as an open reading frame. Similar patterns were observed for the fungal BGC encoding bikaverin, which showed different stages of BGC decay and preservation of functional regulatory genes (Campbell *et al.*, 2012).

Maximum likelihood analysis supports the independent acquisition of NRPS10 at six points in the evolutionary history of *S. pacifica* (Fig. 4A). The 'complete' and partial versions of NRPS10 showed similar GC-content (69.0%–69.6%), which closely matched the average genome content (69.7%–69.9%), suggesting the acquisition events occurred among closely related strains. Only two strains contained the complete BGC (Fig. 4B) while 14 strains contained partial BGCs. Phylogenetic analysis of the regulatory gene revealed that the sequences observed in CNR-909 and CNT-003 (complete BGC) are most closely related to those detected in the partial BGCs observed in CNT-584 and CNT-124 (Supporting Information Fig. S20). These strains are distantly related in the species phylogeny, suggesting that the partial versions of NRPS10 were exchanged among these strains or their close relatives.

Genome sequences derived from large numbers of closely related bacteria are providing unique opportunities to infer the complex evolutionary processes that contribute to the flexible genome (Rodriguez-Valera *et al.*, 2016). The

genus *Salinispora* invests heavily in specialized metabolism, with most BGCs acquired relatively recently in the evolutionary history of the genus. Acquired BGCs are targeted to specific genomic islands and can replace each other using what appears to be a plug and play type model of evolution that provides a mechanism to test the effects of specialized metabolites on fitness. Gaining a better understanding of the ecological functions of these compounds will provide insight into the selective pressures that maintain BGCs within specific lineages and how acquisition events shape population structure.

Experimental procedures

Genome sequencing, assembly and pathway identification

The strains used in this study are listed in Supporting Information Table S1. The genome sequences can be accessed via the IMG Genome ID number on the Joint Genome Institute's Integrated Microbial Genomes website (IMG/ER, https://img.jgi.doe.gov). Genome sequencing and assembly were carried out as previously described (Ziemert *et al.*, 2014) as part of three independent JGI sequencing projects using the standard approaches employed at the time. AntiSMASH 2.0 (Blin *et al.*, 2013) was used for terpene, RiPP, and 'other' BGC identification while PKS and NRPS BGCs were detected using NaPDoS with default settings (Ziemert *et al.*, 2012). Pathways were grouped into OBUs as described previously (Ziemert *et al.*, 2014) and further investigated using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and IMG/ER.

Pan-chromosome assembly

Pseudo-chromosomes were generated for each strain and GI boundaries defined as previously described (Ziemert *et al.*, 2014). Approximately 25% of the BGCs could not be mapped to the pseudo-chromosomes because the flanking regions did not extend into the genomic core. The average size of the GIs and conserved core regions were used to generate figurative representations of the chromosome for each species, which we termed 'pan-chromosomes' (Chan *et al.*, 2015). BGCs were mapped onto the pan-chromosomes according their position in the assemblies. The positions of BGCs that were present within the same GI or core region in more than one strain were averaged. Pan-chromosomes and Fig. 1 were visualized using 'circos' (Krzywinski *et al.*, 2009) and iTOL (Letunic and Bork, 2016). Genes encoding mobile genetic elements (MGEs) were extracted from the IMG/ER database via gene searches using keywords 'transposase', 'integrase' and 'recombinase'. The positions of MGEs in the assemblies were assigned in the same manner as the BGCs.

Phylogeny, ancestral state reconstruction and statistics

The species phylogeny (Fig. 1) and the ancestral state reconstructions were based on 876 concatenated genes common to all 119 genome sequences (Millán-Aguiñaga *et al.*, 2017). Maximum likelihood trees for core biosynthetic genes derived from individual BGCs were constructed with default settings in FastTree (Price *et al.*, 2010) using a protein alignment (MUSCLE) generated by Geneious Pro v5.5.9. Three outgroups were chosen from the top ten BLASTP hits in NCBI. Ancestral state reconstruction was carried out with Mesquite as previously described (Maddison and Maddison, 2011; Ziemert *et al.*, 2014). T-test and ANOVA analyses were carried out in R using default settings.

BGC migration and biogeography

Mesquite was used to determine if BGCs observed at different chromosomal positions in different strains could be attributed to single or multiple acquisition events. If a single acquisition event was predicted, phylogenies of core biosynthetic genes derived from those BGCs were generated. If the phylogenies were congruent, it was inferred that the BGC had moved positions subsequent to the acquisition event. Distances between collection sites were calculated using GPS coordinates (Supporting Information Table S1) of sites (http://www8.nau.edu/cvm/latlongdist.html). Numbers of shared BGCs were determined by comparing all collections sites with each other. The following BGCs were excluded: Genus-specific BGCs (*sio*, PKS4, bac2, *amc*), butyrolactone 1 (salinipostin), butyrolactone 2a and b, *S. arenicola* specific BCGs (rif, terp1, PKS1A, PKS3A, PKS3B, PKS1B) and NRPS2 (conserved in 61/62 *S. arenicola*).

BGC expression

Transcriptome analyses were performed using established methods. In brief, RNA was sent to the US Department of Energy Joint Genome Institute (JGI) for sequencing, quality control and read mapping. Libraries for RNA-Seq were prepared using the TruSeq RNA Sample Prep Kit V2 (Illumina) with sequencing performed on an Illumina HiSEQ 2500 instrument. The sequencing effort generated $>3 \times 10^7$ paired end reads (100 bp) per replicate, which provided over $100\times$ coverage of each *Salinispora* genome. Raw reads were evaluated for artifact sequences using BBDuk by kmer matching (kmer = 25) (https://sourceforge.net/projects/bbmap/). Quality trimming was performed using the phred trimming method set at Q10 (Ewing *et al.*, 1998) and reads under 45 bases removed. Raw reads from each library were aligned to their respective reference genome using BWA (Li and Durbin, 2010). If a read mapped to more than one location, it was ignored. FeatureCounts was used to generate raw gene counts (Liao *et al.*, 2014). Mapped reads were visualized using BamView in Artemis (Rutherford *et al.*, 2000). The number of reads per kilobase of transcript per million mapped reads (RPKM) was used to normalize raw data in Artemis (Mortazavi *et al.*, 2008). BGC expression levels were derived from average values calculated for key biosynthetic genes. These included PKS, NRPS, terpene synthase, precursor peptide (bacteriocin) and LanM (lantibiotic) genes. Additional genes associated with key biosynthetic operons were checked to confirm the expression levels.

Acknowledgements

This work was funded by the NIH under grants R01GM085770 to PRJ and BSM and U19TW007401 to PRJ; and the Foundation for Science and Technology (FCT) under grants PTDC/QUIQUI/119116/2010 and IF/00700/2014 to SPG. Genome sequencing was conducted by the U.S. Department of Energy Joint Genome Institute and supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02- 05CH11231. We acknowledge the government and people of Fiji for permission to collect samples in their local waters. The authors declare no conflict of interest.

References

Abdelmohsen, U.R., Grkovic, T., Balasubramanian, S., Kamel, M.S., Quinn, R.J., and Hentschel, U. (2015) Elicitation of secondary metabolism in actinomycetes. *Biotechnol Adv* **33:** 798–811.

Ahmed, L., Jensen, P., Freel, K., Brown, R., Jones, A., Kim, B.-Y., and Goodfellow, M. (2013) Salinispora pacifica sp. nov., an actinomycete from marine sediments. *Antonie Van Leeuwenhoek* **103:** 1069–1078.

Asolkar, R.N., Freel, K.C., Jensen, P.R., Fenical, W., Kondratyuk, T.P., Park, E.-J., and Pezzuto, J.M. (2009) Arenamides A-C, cytotoxic NF-κ-B inhibitors from the marine actinomycete *Salinispora arenicola*. *J Nat Prod* **72:** 396–402.

Awakawa, T., Crüsemann, M., Munguia, J., Ziemert, N., Nizet, V., Fenical, W., and Moore, B.S. (2015) Salinipyrone and pacificanone are biosynthetic by-products of the rosamicin polyketide synthase. *ChemBioChem* **16:** 1443–1447.

Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., *et al.* (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417:** 141–147.

Berdy, J. (2005) Bioactive microbial metabolites. A personal view. *J Antibiot* **58:** 1–26.

Blin, K., Medema, M.H., Kazempour, D., Fischbach, M.A., Breitling, R., Takano, E., and Weber, T. (2013) antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res* **41:** W204–W212.

Boddy, C. (2013) Bioinformatics tools for genome mining of polyketide and non-ribosomal peptides. *J Ind Microbiol Biotechnol* **41:** 1–8.

Bonet, B., Teufel, R., CruüSemann, M., Ziemert, N., and Moore, B.S. (2014) Direct capture and heterologous expression of *Salinispora* natural product genes for the biosynthesis of enterocin. *J Nat Prod* **78:** 539–542.

Bose, U., Hodson, M.P., Shaw, P.N., Fuerst, J.A., and Hewavitharana, A.K. (2014) Bacterial production of the fungus-derived cholesterol-lowering agent mevinolin. *Biomed Chrom* **28:** 1163–1166.

Campbell, M.A., Rokas, A., and Slot, J.C. (2012) Horizontal transfer and death of a fungal secondary metabolic gene cluster. *Gen Biol Evol* **4:** 289–293.

Castro-Falcón, G., Hahn, D., Reimer, D., and Hughes, C.C. (2016) Thiol probes to detect electrophilic natural products based on their mechanism of action. *ACS Chem Biol* **11:** 2328–2336.

Challis, G.L. (2008) Mining microbial genomes for new natural products and biosynthetic pathways. *Microbiol* **154:** 1555–1569.

Chan, A.P., Sutton, G., DePew, J., Krishnakumar, R., Choi, Y., Huang, X.-Z., *et al.* (2015) A novel method of consensus pan-chromosome assembly and large-scale comparative analysis reveal the highly flexible pan-genome of *Acinetobacter baumannii*. *Genome Biol* **16:** 143.

Cimermancic, P., Medema, M.H., Claesen, J., Kurita, K., Brown, L.C.W., Mavrommatis, K., *et al.* (2014) Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158:** 412–421.

Dobrindt, U., Hochhut, B., Hentschel, U., and Hacker, J. (2004) Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* **2:** 414–424.

Duncan, K.R., Crüsemann, M., Lechner, A., Sarkar, A., Li, J., Ziemert, N., *et al.* (2015) Molecular networking and pattern-based genome mining improves discovery of biosynthetic gene clusters and their products from *Salinispora* species. *Chem Biol* **22:** 460–471.

Egan, S., Wiener, P., Kallifidas, D., and Wellington, E.M. (1998) Transfer of streptomycin biosynthesis gene clusters within streptomycetes isolated from soil. *Appl Environ Microbiol* **64:** 5061–5063.

Eustáquio, A.S., Nam, S.-J., Penn, K., Lechner, A., Wilson, M.C., Fenical, W., *et al.* (2011) The discovery of salinosporamide K from the marine bacterium "*Salinispora pacifica*" by genome mining gives insight into pathway evolution. *ChemBioChem* **12:** 61–64.

Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8:** 175–185.

Fischbach, M.A., and Walsh, C.T. (2006) Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev* **106:** 3468–3496.

Fischbach, M.A., Walsh, C.T., and Clardy, J. (2008) The evolution of gene collectives: how natural selection drives chemical innovation. *Proc Nat Acad Sci USA* **105:** 4601–4608.

Freel, K.C., Nam, S.-J., Fenical, W., and Jensen, P.R. (2011) Evolution of secondary metabolite genes in three closely related marine actinomycete species. *Appl Environ Microbiol* **77:** 7261–7270.

Freel, K.C., Edlund, A., and Jensen, P.R. (2012) Microdiversity and evidence for high dispersal rates in the marine actinomycete 'Salinispora pacifica'. *Environ Microbiol* **14:** 480–493.

Goering, A.W., McClure, R.A., Doroghazi, J.R., Albright, J.C., Haverland, N.A., Zhang, Y., *et al.* (2016) Metabologenomics: Correlation of microbial gene clusters with metabolites drives discovery of a nonribosomal peptide with an unusual amino acid monomer. *ACS Central Sci* **2:** 99–108.

Greunke, C., Glöckle, A., Antosch, J., and Gulder, T.A. (2017) Biocatalytic total synthesis of ikarugamycin. *Angew Chem Inter Ed* **56:** 4351–4355.

Gross, H. (2007) Strategies to unravel the function of orphan biosynthesis pathways: recent examples and future prospects. *Appl Microbiol Biotechnol* **75:** 267–277.

Harrison, S.J., Mainwaring, P., Price, T., Millward, M.J., Padrik, P., Underhill, C.R., *et al.* (2016) Phase I clinical trial

of marizomib (NPI-0052) in patients with advanced malignancies including multiple myeloma: study NPI-0052–102 final results. *Clin Cancer Res* **22:** 4559–4566.

Hertweck, C. (2009) The biosynthetic logic of polyketide diversity. *Angew Chem Int Ed* **48:** 4688–4716.

Jensen, P.R. (2016) Natural products and the gene cluster revolution. *Trends Microbiol* **24:** 968–977.

Jensen, P.R., Moore, B.S., and Fenical, W. (2015) The marine actinomycete genus *Salinispora*: a model organism for secondary metabolite discovery. *Nat Prod Rep* **32:** 738–751.

Juhas, M., van der Meer, J.R., Gaillard, M., Harding, R.M., Hood, D.W., and Crook, D.W. (2009) Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev* **33:** 376–393.

Kersten, R.D., Lane, A.L., Nett, M., Richter, T.K.S., Duggan, B.M., Dorrestein, P.C., and Moore, B.S. (2013a) Bioactivity-guided genome mining reveals the lomaiviticin biosynthetic gene cluster in *Salinispora tropica*. *ChemBioChem* **14:** 955–962.

Kersten, R.D., Ziemert, N., Gonzalez, D.J., Duggan, B.M., Nizet, V., Dorrestein, P.C., and Moore, B.S. (2013b) Glycogenomics as a mass spectrometry-guided genome-mining method for microbial glycosylated molecules. *Proc Natl Acad Sci USA* **110:** 4407–4416.

Kim, E., Moore, B.S., and Yoon, Y.J. (2015) Reinvigorating natural product combinatorial biosynthesis with synthetic biology. *Nat Chem Biol* **11:** 649–659.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., *et al*. (2009) Circos: an information aesthetic for comparative genomics. *Gen Res* **19:** 1639–1645.

Lane, A.L., Nam, S.-J., Fukuda, T., Yamanaka, K., Kauffman, C.A., Jensen, P.R., *et al*. (2013) Structures and comparative characterization of biosynthetic gene clusters for cyanosporasides, enediyne-derived natural products from marine actinomycetes. *J Am Chem Soc* **135:** 4171–4174.

Letunic, I., and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucl Acids Res* **44:** 242–245.

Li, H., and Durbin, R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26:** 589–595.

Liao, Y., Smyth, G.K., and Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30:** 923–930.

Maddison, W.P., and Maddison, D.R. (2011) *Mesquite: A Modular System for Evolutionary Analysis*, Version 2.75. Available at http://mesquiteproject.org.

Maldonado, L.A., Stach, J.E.M., Pathom-aree, W., Ward, A.C., Bull, A.T., and Goodfellow, M. (2005) Diversity of cultivable actinobacteria in geographically widespread marine sediments. *Antonie Van Leeuwenhoek Int J Gen Mol Microbiol* **87:** 11–18.

McGlinchey, R.P., Nett, M., and Moore, B.S. (2008) Unraveling the biosynthesis of the sporolide cyclohexenone building block. *J Am Chem Soc* **130:** 2406–2407.

Medema, M.H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J.B., Blin, K., *et al*. (2015) Minimum information about a biosynthetic gene cluster. *Nat Chem Biol* **11:** 625–631.

Metsa-Ketela, M., Halo, L., Munukka, E., Hakala, J., Mantsala, P., and Ylihonko, K. (2002) Molecular evolution of aromatic polyketides and comparative sequence analysis of polyketide ketosynthase and 16S ribosomal DNA genes from various *Streptomyces* species. *Appl Environ Microbiol* **68:** 4472–4479.

Millán-Aguiñaga, N., Chavarria, K.L., Ugalde, J.A., Letzel, A.-C., Rouse, G.W., and Jensen, P.R. (2017) Phylogenomic insight into *Salinispora* (Bacteria, Actinobacteria) species designations. *Sci Rep* **7:** 3564.

Miyanaga, A., Janso, J.E., McDonald, L., He, M., Liu, H., Barbieri, L., *et al*. (2011) Discovery and assembly-line biosynthesis of the lymphostin pyrroloquinoline alkaloid family of mTOR inhibitors in *Salinispora* bacteria. *J Am Chem Soc* **133:** 13311–13313.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5:** 621–628.

Nett, M., Ikeda, H., and Moore, B.S. (2009) Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat Prod Rep* **26:** 1362–1384.

Newman, D.J., and Cragg, G.M. (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J Nat Prod* **75:** 311–335.

Nielsen, J.C., Grijseels, S., Prigent, S., Ji, B., Dainat, J., Nielsen, K.F., *et al*. (2017) Global analysis of biosynthetic gene clusters reveals vast potential of secondary metabolite production in *Penicillium* species. *Nat Microbiol* **2:** 17044.

Ochman, H., Lawrence, J.G., and Groisman, E.A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405:** 299–304.

Ogasawara, Y., Kawata, J., Noike, M., Satoh, Y., Furihata, K., and Dairi, T. (2016) Exploring peptide ligase orthologs in Actinobacteria - Discovery of pseudopeptide natural products, ketomemicins. *ACS Chem Biol* **11:** 1686–1692.

Patin, N.V., Duncan, K.R., Dorrestein, P.C., and Jensen, P.R. (2015) Competitive strategies differentiate closely related species of marine actinobacteria. *ISME J* **10:** 478–490.

Penn, K., Jenkins, C., Nett, M., Udwary, D.W., Gontang, E.A., McGlinchey, R.P., *et al*. (2009) Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria. *ISME J* **3:** 1193–1203.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One* **5:** e9490.

Price-Whelan, A., Dietrich, L.E., and Newman, D. (2006) Rethinking secondary 'metabolism': physiological roles for phenazine antibiotics. *Nat Chem Biol* **2:** 71–78.

Richter, T.K., Hughes, C.C., and Moore, B.S. (2015) Sioxanthin, a novel glycosylated carotenoid, reveals an unusual subclustered biosynthetic pathway. *Environ Microbiol* **17:** 2158–2171.

Roberts, A.A., Schultz, A.W., Kersten, R.D., Dorrestein, P.C., and Moore, B.S. (2012) Iron acquisition in the marine actinomycete genus *Salinispora* is controlled by the desferrioxamine family of siderophores. *FEMS Microbiol Lett* **335:** 95–103.

Rodriguez-Valera, F., Martin-Cuadrado, A.-B., and López-Pérez, M. (2016) Flexible genomic islands as drivers of genome evolution. *Curr Opin Microbiol* **31:** 154–160.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* **16:** 944–945.

Sandy, M., and Butler, A. (2009) Microbial iron acquisition: marine and terrestrial siderophores. *Chem Rev* **109:** 4580–4595.

Schultz, A.W., Oh, D.C., Carney, J.R., Williamson, R.T., Udwary, D.W., Jensen, P.R., *et al*. (2008) Biosynthesis and structures of cyclomarins and cyclomarazines, prenylated cyclic peptides of marine actinobacterial origin. *J Am Chem Soc* **130:** 4507–4516.

Schulze, C.J., Navarro, G., Ebert, D., DeRisi, J., and Linington, R.G. (2015) Salinipostins A–K, long-chain bicyclic phosphotriesters as a potent and selective antimalarial chemotype. *J Org Chem* **80:** 1312–1320.

Seyedsayamdost, M.R. (2014) High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. *Proc Natl Acad Sci USA* **111:** 7266–7271.

Tang, X., Li, J., Millán-Aguiñaga, N., Zhang, J.J., O'Neill, E.C., Ugalde, J.A., *et al*. (2015) Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. *ACS Chem Biol* **10:** 2841–2849.

Udwary, D.W., Zeigler, L., Asolkar, R.N., Singan, V., Lapidus, A., Fenical, W., *et al*. (2007) Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc Natl Acad Sci USA* **104:** 10376–10381.

Vidgen, M.E., Hooper, J.N.A., and Fuerst, J.A. (2012) Diversity and distribution of the bioactive actinobacterial genus Salinispora from sponges along the Great Barrier Reef. *Antonie Van Leeuwenhoek* **101:** 603–618.

Villadsen, N.L., Jacobsen, K.M., Keiding, U.B., Weibel, E.T., Christiansen, B., Vosegaard, T., *et al*. (2017) Synthesis of ent-BE-43547A1 reveals a potent hypoxia-selective anticancer agent and uncovers the biosynthetic origin of the APD-CLD natural products. *Nat Chem* **9:** 264–272.

Waters, C.M., and Bassler, B.L. (2005) Quorum sensing: cell-to-cell communication in bacteria. *Annu Rev Cell Dev Biol* **21:** 319–346.

Weber, T. (2014) In silico tools for the analysis of antibiotic biosynthetic pathways. *Int J Med Microbiol* **304:** 230–235.

Wietz, M., Duncan, K., Patin, N., and Jensen, P. (2013) Antagonistic interactions mediated by marine bacteria: the role of small molecules. *J Chem Ecol* **39:** 879–891.

Williams, P.G., Miller, E.D., Asolkar, R.N., Jensen, P.R., and Fenical, W. (2007) Arenicolides A-C, 26-membered ring macrolides from the marine actinomycete *Salinispora arenicola*. *J Org Chem* **72:** 5025–5034.

Wilson, M.C., Gulder, T.A.M., Mahmud, T., and Moore, B.S. (2010) Shared biosynthesis of the saliniketals and rifamycins in *Salinispora arenicola* is controlled by the sare1259-encoded cytochrome P450. *J Am Chem Soc* **132:** 12757–12765.

Winter, J.M., Behnken, S., and Hertweck, C. (2011) Genomics-inspired discovery of natural products. *Curr Opin Chem Biol* **15:** 22–31.

Xu, M., Hillwig, M.L., Lane, A.L., Tiernan, M.S., Moore, B.S., and Peters, R.J. (2014) Characterization of an orphan diterpenoid biosynthetic operon from *Salinispora arenicola*. *J Nat Prod* **77:** 2144–2147.

Ziemert, N., and Jensen, P.R. (2012) Phylogenetic approaches to natural product structure prediction. *Methods Enzymol* **517:** 161–182.

Ziemert, N., Lechner, A., Wietz, M., Millan-Aguinaga, N., Chavarria, K.L., and Jensen, P.R. (2014) Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*. *Proc Natl Acad Sci USA* **111:** 1130–1139.

Ziemert, N., Alanjary, M., and Weber, T. (2016) The evolution of genome mining in microbes–a review. *Nat Prod Rep* **33:** 988–1005.

Supporting information

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

Table S1. Strain information including 16S sequence types (Freel *et al*., 2012; Vidgen *et al*., 2012; Millán-Aguiñaga *et al*., 2017) and location of origin (CB: Caribbean, SC: Sea of Cortez, PV: Puerto Vallarta, RS: Red Sea, FJ: Fiji, HI: Hawaii, PM: Palmyra, GU: Guam, PL: Palau, MD: Madeira archipelago, *sponge isolates).

Table S2. GPS coordinates for calculation of distances between collection sites.

Fig. S1. *Salinispora* specialized metabolism. A. Boxplot of BGCs per genome (median in red): boxes indicate upper and lower quartile, lines indicate minimum and maximum number of BGCs per genome, circles indicate outliers. B. Species-level BGC distributions (red = polyketide synthase (PKS), orange = non-ribosomal peptide synthetase (NRPS), brown = PKS-NRPS hybrid, purple = ribosomally synthesized and post-translationally modified peptide (RiPP), green = terpene, yellow = siderophore, grey = other). C. Genus-level BGC distributions determined bioinformatically (top) and based on isolated compounds (bottom).

Fig. S2. Presence-absence matrix of BGCs in *Salinispora* spp. The species phylogeny is based on core genes with no evidence of recombination (Millán-Aguiñaga *et al*., 2017). The perpendicular axis lists the 176 BGCs detected in these strains. Coloured boxes indicate the presence of a BGC in a genome. BGCs are listed in decreasing order of abundance. Colours correspond to the location from which the strain was isolated (purple = Caribbean, orange = Fiji, green = Palau, blue = Guam, yellow = Hawaii, pink = Palmyra, light blue- = Sea of Cortez, light pink = Puerto Vallarta, brown = Madeira archipelago, red = Red Sea). Location specific BGCs highlighted with a black box.

Fig. S3. Rank-abundance curve of BGCs in *Salinispora* (176 BGCs in total including 61 singletons and 34 that were observed in only two strains).

Fig. S4. Pairwise comparison of shared BGCs (y-axis) and geographic distance (km) between collection locations (*x*-axis). All species include *S. arenicola*, *S. pacifica* and *S. tropica*. Only *S. arenicola* and *S. pacifica* are shown as individual species as *S. tropica* is only present at one location.

Fig. S5. A likelihood analysis reveals the distribution of the PKS9 BGC (black circles) within the *S. pacifica* lineage. From this analysis, it can be inferred that PKS9 was acquired independently on three occasions by strains CNT-403, DSM45548, and by the common ancestor of strains CNT-609 and CNT-084, and CNT-133. All of these strains were isolated from samples collected in Fiji (FJ).

Fig. S6. Distribution of MGEs (in percent) within GIs and core regions. NA = percentage of MGEs that could not be mapped to a specific location. Colours indicate species (red = *S. tropica*, yellow = *S. pacifica*, blue = *S. arenicola*) (A) Transposase distribution per GI, (B) Integrase distribution per GI, (C) Recombinase distribution per GI.

Fig. S7. (A) Likelihood analysis provides evidence that *NRPS42* was acquired once in the *Salinispora* phylogeny. This BGC occurs in genomic island GI1 in *S. pacifica* strains CNS-996 and CNT-045 and GI20 in strains CNS-860 and DSM 45543 suggesting it has changed position (migrated). Strain names are followed by the location from which they were isolated (CB: Caribbean, SC: Sea of Cortez, PV: Puerto Vallarta, RS: Red Sea, FJ: Fiji, HI: Hawaii, PM: Palmyra, GU: Guam, PL: Palau, MD: Madeira archipelago). (B) The maximum likelihood phylogeny of the non-ribosomal peptide synthetase gene in *NRPS42* is congruent with the species phylogeny, supporting vertical inheritance following acquisition. Light-grey boxes indicate strains with the BGC in GI1, dark-grey boxes indicate strains with the BGC in GI20.

Fig. S8. (A) Likelihood analysis predicts the presence of *NRPS4* in the common ancestor of the genus. This BGC is observed in GI21 in all strains except CNS-237, where it is observed in GI10, and CNS-860 and CNS-133, where it is observed in the conserved region between GI21 and GI1. (B) The maximum likelihood phylogeny of the non-ribosomal peptide synthetase gene in *NRPS4* is congruent with the species phylogeny, supporting vertical inheritance and the subsequent migration of this BGC in the highlighted strains. *S. tropica* and *S. arenicola* clades are collapsed. Light-grey-shaded box indicates strain CNS-237, where NRPS4 was located in GI10. Dark shaded box indicates strains CNS-860 and CNT-133, where NRPS was located in the conserved region between GI21 and GI1. Strain names are followed by the location from which they were isolated (CB: Caribbean, SC: Sea of Cortez, PV: Puerto Vallarta, RS: Red Sea, FJ: Fiji, HI: Hawaii, PM: Palmyra, GU: Guam, PL: Palau, MD: Madeira archipelago).

Fig. S9. (A) Likelihood analysis of strains possessing *amc* between GI12 and GI13 predicts that this BGC was present in the common ancestor of the genus (right side). A similar analysis of the five strains possessing *amc* in GI6 (left side) predicts that it was acquired (i.e., migrated to) this new location independently on five occasions. (B) The maximum likelihood phylogeny of *amcB* (3-dehydroquinate synthase, Sare_2489) in *S. arenicola* is congruent with the species phylogeny with the exception of one strain, which supports BGC migration in the five strains as opposed to independent acquisition events. Light-grey-shaded boxes indicate those strains in which the BGC has migrated to GI6. Strain names are followed by the location from which they were isolated (CB: Caribbean, SC: Sea of Cortez, PV: Puerto Vallarta, RS: Red Sea, FJ: Fiji, HI: Hawaii, PM: Palmyra, GU: Guam, PL: Palau, MD: Madeira archipelago).

Fig. S10. (A) Likelihood analysis of the *PKS2* BGC predicts it was present in a common ancestor of the genus, maintained in *S. arenicola*, lost in the other two species, and subsequently re-acquired in four *S. pacifica* strains. The BGC is located in GI6 in most *S. arenicola* strains but shows evidence of migration to GI13 and other positions. Strain names are followed by the location from which they were isolated (CB: Caribbean, SC: Sea of Cortez, PV: Puerto Vallarta, RS: Red Sea, FJ: Fiji, HI: Hawaii, PM: Palmyra, GU: Guam, PL: Palau, MD: Madeira archipelago). (B) Maximum likelihood phylogeny of the keto-acyl-ACP synthase amino acid sequence from *PKS2*. Grey-shaded boxes indicate *S. arenicola* strains in which the BGC is located outside of GI6.

Fig. S11. (A) Likelihood analysis predicts multiple acquisition events for *lan1*. Many of these are associated with different regions of the chromosome. (B) Maximum likelihood phylogeny of the *lan1* lantibiotic dehydratase amino acid sequence is incongruent with the species phylogeny supporting multiple acquisition events. Grey-shaded boxes indicate strains (SA = *S. arenicola*, SP = *S. pacifica*) with differing genomic position of the *lan1* BGC. Genomic islands (GIs) shown next to box. Strain names are followed by the location from which they were isolated (CB: Caribbean, SC: Sea of Cortez, PV: Puerto Vallarta, RS: Red Sea, FJ: Fiji, HI: Hawaii, PM: Palmyra, GU: Guam, PL: Palau, MD: Madeira archipelago).

Fig. S12. Maximum likelihood analysis of the *but1* BGC provides evidence that it was present in a common ancestor of the genus. The location of the BGC in different chromosomal position supports migration. Chromosomal location of the BGC is shown next to the strains. Strain names are followed by the location from which they were isolated (CB: Caribbean, SC: Sea of Cortez, PV: Puerto Vallarta, RS: Red Sea, FJ: Fiji, HI: Hawaii, PM: Palmyra, GU: Guam, PL: Palau, MD: Madeira archipelago).

Fig. S13. Maximum likelihood phylogeny of the A-factor amino acid sequence of from *but1* BGC. *S. tropica* and *S. arenicola* clades are collapsed. Grey-shaded boxes indicate *S. pacifica* strains with differing genomic positions for *but1* (GI10 instead of GI15). Genomic islands (GIs) shown next to box. For strains encircled with a black box, the chromosomal position could not be assigned.

Fig. S14. Maximum likelihood analysis of *terp6* BGC supports independent acquisition events instead of intragenomic movement. Chromosomal location of the BGC is shown next to strains. Strain names are followed by the location from which they were isolated (CB: Caribbean, SC: Sea of Cortez, PV: Puerto Vallarta, RS: Red Sea, FJ: Fiji, HI: Hawaii, PM: Palmyra, GU: Guam, PL: Palau, MD: Madeira archipelago).

Fig. S15. Maximum likelihood phylogeny of polyprenyl disphosphate synthase amino acid sequences from the *terp6* BGC in *S. pacifica*. Light-grey-shaded boxes indicate *S. pacifica* strains in which the gene is located in a different chromosomal position (Genomic island 17 vs. GI5).

Fig. S16. Likelihood analysis of the *STPKS1* BGC shows evidence for independent acquisition events. Chromosomal location of the BGC (Genomic island) is shown next to the strains. Strain names are followed by the location from which they were isolated (CB: Caribbean, SC: Sea of Cortez, PV: Puerto Vallarta, RS: Red Sea, FJ: Fiji, HI: Hawaii, PM: Palmyra, GU: Guam, PL: Palau, MD: Madeira archipelago).

Fig. S17. Maximum likelihood phylogeny of beta-keto acyl synthase amino acid sequence from the *STPKS1* BGC in *S. pacifica* and *S. tropica*. Light-grey-shaded box indicates the *S. pacifica* strain with the BGC in GI15 while dark grey indicates GI10.

Fig. S18. Likelihood analysis of the *lan8* BGC shows evidence for multiple acquisition events (black circles). Strain names are followed by the location from which they were

isolated (CB: Caribbean, SC: Sea of Cortez, PV: Puerto Vallarta, RS: Red Sea, FJ: Fiji, HI: Hawaii, PM: Palmyra, GU: Guam, PL: Palau, MD: Madeira archipelago).

Fig. S19. Maximum likelihood phylogeny of the two genes from the *lan8* BGC. Grey boxes indicate strains where the genomic position of the BGC could be assigned. Genomic islands (GIs) shown next to box. A. Phylogeny of the putative precursor peptide. B. Phylogeny of the thiopeptide modifying enzyme.

Fig. S20. Maximum likelihood phylogeny of a regulatory gene from *NRPS10*. Different shaded boxes represent different GIs (shown next to box). NA: genomic position could not be determined.

Fig. S21. Likelihood analysis of the *sal* and *lym* BGCs. Strain names are followed by the location from which they were isolated (CB: Caribbean, SC: Sea of Cortez, PV: Puerto Vallarta, RS: Red Sea, FJ: Fiji, HI: Hawaii, PM: Palmyra, GU: Guam, PL: Palau, MD: Madeira archipelago).