Comparative transcriptomics as a guide to natural product discovery and biosynthetic gene cluster functionality

Gregory C. A. Amos[a], Takayoshi Awakawa[a,b], Robert N. Tuttle[a], Anne-Catrin Letzel[a], Min Cheol Kim[a], Yuta Kudo[a], William Fenical[a,c], Bradley S. Moore[a,c,d], and Paul R. Jensen[a,c,1]

[a]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093; [b]Graduate School of Pharmaceutical Sciences, The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan; [c]Center for Microbiome Innovation, University of California, San Diego, La Jolla, CA 92093; and [d]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093

Edited by Jon Clardy, Harvard Medical School, Boston, MA, and accepted by Editorial Board Member Stephen J. Benkovic November 14, 2017 (received for review August 15, 2017)

Bacterial natural products remain an important source of new medicines. DNA sequencing has revealed that a majority of natural product biosynthetic gene clusters (BGCs) maintained in bacterial genomes have yet to be linked to the small molecules whose biosynthesis they encode. Efforts to discover the products of these orphan BGCs are driving the development of genome mining techniques based on the premise that many are transcriptionally silent during normal laboratory cultivation. Here, we employ comparative transcriptomics to assess BGC expression among four closely related strains of marine bacteria belonging to the genus *Salinispora*. The results reveal that slightly more than half of the BGCs are expressed at levels that should facilitate product detection. By comparing the expression profiles of similar gene clusters in different strains, we identified regulatory genes whose inactivation appears linked to cluster silencing. The significance of these subtle differences between expressed and silent BGCs could not have been predicted a priori and was only revealed by comparative transcriptomics. Evidence for the conservation of silent clusters among a larger number of strains for which genome sequences are available suggests they may be under different regulatory control from the expressed forms or that silencing may represent an underappreciated mechanism of gene cluster evolution. Coupling gene expression and metabolomics data established a bioinformatic link between the salinipostins and their associated BGC, while genetic manipulation established the genetic basis for this series of compounds, which were previously unknown from *Salinispora pacifica*.

Salinispora | transcriptomics | biosynthetic gene cluster | specialized metabolism

Bacteria are an important source of the small-molecule natural product drugs employed in modern medicine (1). While interest in natural product drug discovery has waned in past decades (2), major advances in our understanding of how these compounds are assembled, coupled with ready access to genome sequence data, have provided unique opportunities to return to microbial natural products research using more informed discovery approaches (3). This paradigm shift was facilitated by the general observation that genes encoding natural product biosynthesis are clustered on the bacterial chromosome and readily identifiable using a variety of bioinformatic approaches (4). One consistent observation from the analysis of bacterial genome sequences is that, even for well-studied strains (5), a majority of biosynthetic gene clusters (BGCs) remain orphan (i.e., they have yet to be linked to the small molecules whose biosynthesis they encode) (6, 7). This surprising observation inspired the development of "genome mining" techniques that seek to find the products of unassigned BGCs (3, 8, 9). Although putting estimates on the number of orphan BGCs in the microbial world is difficult, a recent survey of 1,154 genomes predicted the presence of over 10,000 distinct BGCs (10), which is ∼10-fold greater than the total number of experimentally characterized BGCs listed in the Minimum Information about a Biosynthetic Gene cluster (MIBiG) repository as of this writing (mibig.secondarymetabolites.org/repository.html) (11).

Current genome mining efforts are often based on the assumption that the products of orphan BGCs remain obscured due to insufficient levels of gene expression (12–14). In response, considerable effort has been devoted to awakening "silent" BGCs using either synthetic biology (15) or culture-based approaches (16). While transcript analyses have confirmed that some BGCs are silent under standard laboratory conditions (17), global DNA microarrays (18) and proteomics (19) have provided some of the first hints that a majority of BGCs can be expressed in wild-type strains, suggesting they required no special effort for activation. Furthermore, proteomics targeting specific BGC classes (20) or correlating protein expression (21) or BGC distributions (22) with secondary metabolite production are providing useful approaches to identify active BGCs and link them to their products, while new techniques such as OASIS (23) and PrISM (20) are providing opportunities to assess biosynthetic gene activity. Despite these advances, transcriptomics has seldom been employed to assess

Significance

Genomics has revealed that even well-studied bacteria maintain many more biosynthetic gene clusters (BGCs) predicted to encode specialized metabolites than expected based on product discovery. These orphan BGCs are often assumed to be transcriptionally silent. Here, we show that a majority of the 46 BGCs observed in four strains of the marine actinomycete *Salinispora* are transcribed at levels that should facilitate product detection. In five cases, similar BGCs were differentially expressed among strains, suggesting that simple presence or absence analyses are not good predictors of metabolic output. Highly expressed BGCs were bioinformatically linked to their products, including a series of salinipostins not previously reported from *Salinispora pacifica*. Subsequent genetic experiments established a formal link between salinipostins and their cognate BGC.

Author contributions: G.C.A.A., T.A., Y.K., B.S.M., and P.R.J. designed research; G.C.A.A., T.A., R.N.T., A.-C.L., M.C.K., and Y.K. performed research; G.C.A.A., T.A., M.C.K., W.F., B.S.M., and P.R.J. analyzed data; and G.C.A.A., T.A., W.F., B.S.M., and P.R.J. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. J.C. is a guest editor invited by the Editorial Board.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. NC_009953.1, NZ_ARBB00000000.1, NZ_AQZW00000000, and NC_009380.1).

[1]To whom correspondence should be addressed. Email: pjensen@ucsd.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1714381115/-/DCSupplemental.

global BGC expression, leaving major gaps in our understanding of the relationships between orphan BGCs and specialized metabolite discovery.

Theory suggests that the initiation of specialized metabolism is linked to the onset of development as triggered by a culture's nutrient status (24). There is evidence that BGC expression in *Streptomyces* spp. and *Burkholderia thailandensis* is silenced by repressors such as DasR and ScmR, respectively, which act as a master switch for antibiotic production (24, 25). One of the most striking features of this regulation, which includes both pleotropic and pathway-specific regulators (26), remains its diversity and complexity (27). While much is to be gained from a better understanding of regulation, it remains unclear if transcriptional silence is the primary mechanism by which strains containing large numbers of BGCs do not deliver on their biosynthetic potential when cultured in the laboratory.

The marine actinomycete genus *Salinispora* is a rich source of diverse natural products (28). To date, at least 29 distinct compound classes have been reported from the three named species, including salinosporamide A (29, 30), which is advancing through clinical trials for the treatment of cancer (31). Of these compounds, 19 have been experimentally linked to their respective BGCs, while three more have been linked using bioinformatic approaches (28). Despite these significant efforts, a recent analysis of 119 *Salinispora* genomes revealed 176 distinct BGCs, the vast majority of which remain orphan (32). The aim of this study was to address the relationships between orphan and transcriptionally silent BGCs by assessing global transcription patterns in four *Salinispora* strains in conjunction with metabolomic studies to assess the relationships between gene expression and compound detection. The results provide a baseline for functional levels of BGC expression and shed new light on how growth phase, levels of conservation, and regulatory mechanisms impact the likelihood of product discovery. Our data indicate that many orphan BGCs are expressed at levels sufficient for product detection, while subtle differences in BGC content and expression levels suggest that "inactivation" may be more common than previously recognized. We identified a highly expressed, yet orphan, BGC and provide experimental evidence linking it to salinipostin biosynthesis, which is reported here from *Salinispora pacifica*.

Results

Global BGC Expression. An analysis of 119 *Salinispora* genomes sequenced in collaboration with the Joint Genome Institute (https://img.jgi.doe.gov/) led to the selection of four strains that contained 49 different BGCS, including 13 that could be linked to the small molecules whose biosynthesis they encode (7, 32–44) (Fig. 1). The remaining 36 BGCs (74%) were categorized as orphan since their products had yet to be determined at the start of this study. While the large number of orphan BGCs observed is similar to that reported for other bacteria (10), it remains unknown if this is due to transcriptional inactivity during laboratory cultivation or to the myriad factors that can hinder small-molecule detection. To explore the relationships between orphan BGCs and transcription levels, global transcriptome analyses were performed at both exponential (96 h) and stationary (216 h) growth phases in four *Salinispora* strains (*SI Appendix*, Fig. S1). A baseline to distinguish between silent and expressed BGCs was established at 27.1 reads per kilobase of transcript per million mapped reads (RPKM) based on the lowest RPKM values associated with the detection of a known *Salinispora* metabolite in comparison to the highest levels where compounds were not detected (*SI Appendix*, Table S1). In support of this threshold, staurosporine could be detected by mass spectral analysis in an extract of *Salinispora arenicola* CNS-205 when the *sta* BGC was expressed at 27.1 RPKM, but not in a similar extract from *S. pacifica* CNT-150 expressed at 12.0 RPKM. While this empirical threshold was established based on the average expression levels of key biosynthetic genes in each BGC, it may vary depending upon the analytical methods employed and the structural features of individual compounds that facilitate their detection.

BGC				Strain			
No	Name	Product	Ref.	CNS-991	CNS-205	CNB-440	CNT-150
1	Amc	TBD	NA				
2	Bac2	TBD	NA				
3	Bac4	TBD	NA				
4	**But1/*spt***	**Salinipostins***	NA				
5	***Cya***	**Cyanosporaside**	28				
6	***Cyl***	**Cyclomarin B**	29				
7	***Des***	**Desferrioxamine E**	30				
8	***Enc***	**Enterocin**	38				
9	FAS1	TBD	NA				
10	FAS5	TBD	NA				
11	Lan1	TBD	NA				
12	Lan2	TBD	NA				
13	Lan3b	TBD	NA				
14	Lan4	TBD	NA				
15	***Lom***	**Lomaiviticin**	31				
16	Lan6	TBD	NA				
17	***Lym***	**Lymphostin**	32				
18	NRPS1	TBD	NA				
19	NRPS2	TBD	NA				
20	NRPS3	TBD	NA				
21	NRPS4	TBD	NA				
22	NRPS14	TBD	NA				
23	NRPS35	TBD	NA				
24	NRPS36	TBD	NA				
25	PKS1A	TBD	NA				
26	PKS1B	TBD	NA				
27	PKS1C	TBD	NA				
28	PKS2	TBD	NA				
29	PKS3A	TBD	NA				
30	PKS3B	TBD	NA				
31	PKS4	TBD	NA				
32	PKS5	TBD	NA				
33	PKS7	TBD	NA				
34	PKS15	TBD	NA				
35	PKS16	TBD	NA				
36	PKS17	TBD	NA				
37	PKS30	TBD	NA				
38	PKSNRPS2	TBD	NA				
39	***Slm***	**Salinilactam**	7				
40	***Sal***	**Salinosporamide**	33				
41	***Rif***	**Rifamycin S**	39				
42	Sid1	TBD	NA				
43	Sid3	TBD	NA				
44	Sid4	TBD	NA				
45	***Sio***	**Sioxanthin**	34				
46	***Spo***	**Sporolide**	35				
47	***Sta***	**Staurosporine**	36				
48	STPKS1	TBD	NA				
49	***Ido***	**Isopimara-8,15-dien-19-ol**	37				

Fig. 1. BGC distributions among strains. BGCs in bold font have been linked to their product. *But1/*spt* was assigned as part of this study. TBD, BGC products remain to be determined. Unassigned BGCs were named based on bioinformatic predictions (Lan, lantibiotic; NRPS, nonribosomal peptide synthetase; PKS/NRPS, hybrid; Sid, siderophore). Box colors: white, BGC not present; gray, BGC silent; red, BGC expressed; red with diagonal lines, BGC expressed and products detected. CNB-440: *S. tropica*, CNT-150: *S. pacifica*, CNS-991: S. *arenicola*, CNS-205: *S. arenicola*. Isopimara-8,15-dien-19-ol from the *ido* BGC may not represent the end product of this BGC, and thus was excluded from calculations of product detection from characterized BGCs. NA, not applicable.

The transcription data revealed consistent yet unexpected results for all four strains. In the case of *Salinispora tropica* strain CNB-440, 10 of 19 BGCs were transcribed during at least one growth phase at levels that were considerably greater (average of 449 RPKM) than the threshold established for compound detection (Fig. 2). Twelve orphan BGCs were identified in this strain, and five of these were expressed above the threshold, accounting for 50% of the expressed BGCs (Table 1). In addition

Table 1. Summary of BGC expression data

Species (strain)	BGCs	Expressed BGCs (% of total)	Orphan BGCs (% of total)	Orphan/expressed BGCs (% of total expressed)
S. tropica (CNB-440)	19	10 (52.6)	12 (63.2)	5 (50.0)
S. arenicola (CNS-205)	26	13 (50.0)	19 (73.1)	7 (53.4)
S. arenicola (CNS-991)	28	12 (42.9)	22 (78.6)	7 (58.3)
S. pacifica (CNT-150)	18	13 (72.2)	13 (72.2)	9 (69.2)

to the observation that most BGCs were not silent, eight were expressed in exponential phase at levels that often exceeded those detected in the stationary phase. The products of five of the 10 expressed BGCs had been experimentally verified at the time of this study (Fig. 1), yet only sioxanthin and salinilactam could be detected, with the former based on its diagnostic UV spectrum (*SI Appendix*, Fig. S2*A*) and the latter based on a close match to published UV maxima (7). Our inability to detect lomaiviticins or salinosporamides, despite the high levels of BGC expression, could be due to the inherent limitations associated with liquid chromatography (LC)/MS analyses.

Similar expression levels were observed for *S. arenicola* strain CNS-205, with 13 of 26 BGCs expressed, including seven of 19 orphan BGCs (Table 1). In this case, the products of four of the five experimentally verified BGCs were detected when expressed above the threshold level (Fig. 2 and *SI Appendix*, Fig. S2 *A–D*). The previously characterized product isopimara-8,15-dien-19-ol was not detected and may not represent the end product of the *ido* BGC (42). It was therefore excluded from the calculations. The lowest proportion of expressed BGCs was observed in *S. arenicola* strain CNS-991 (12 of 28), while the highest was in CNT-150 (13 of 18), which included nine of 13 orphan BGCs. From these two strains, the products of seven of the nine characterized BGCs were detected when expressed above the threshold (*SI Appendix*, Fig. S2 *A* and *E–H*). In summary across the four strains and two time points, 25 (51.0%) of the 49 BGCs were expressed in at least one strain, while each strain transcribed, on average, 54.4% of its biosynthetic potential (Fig. 1

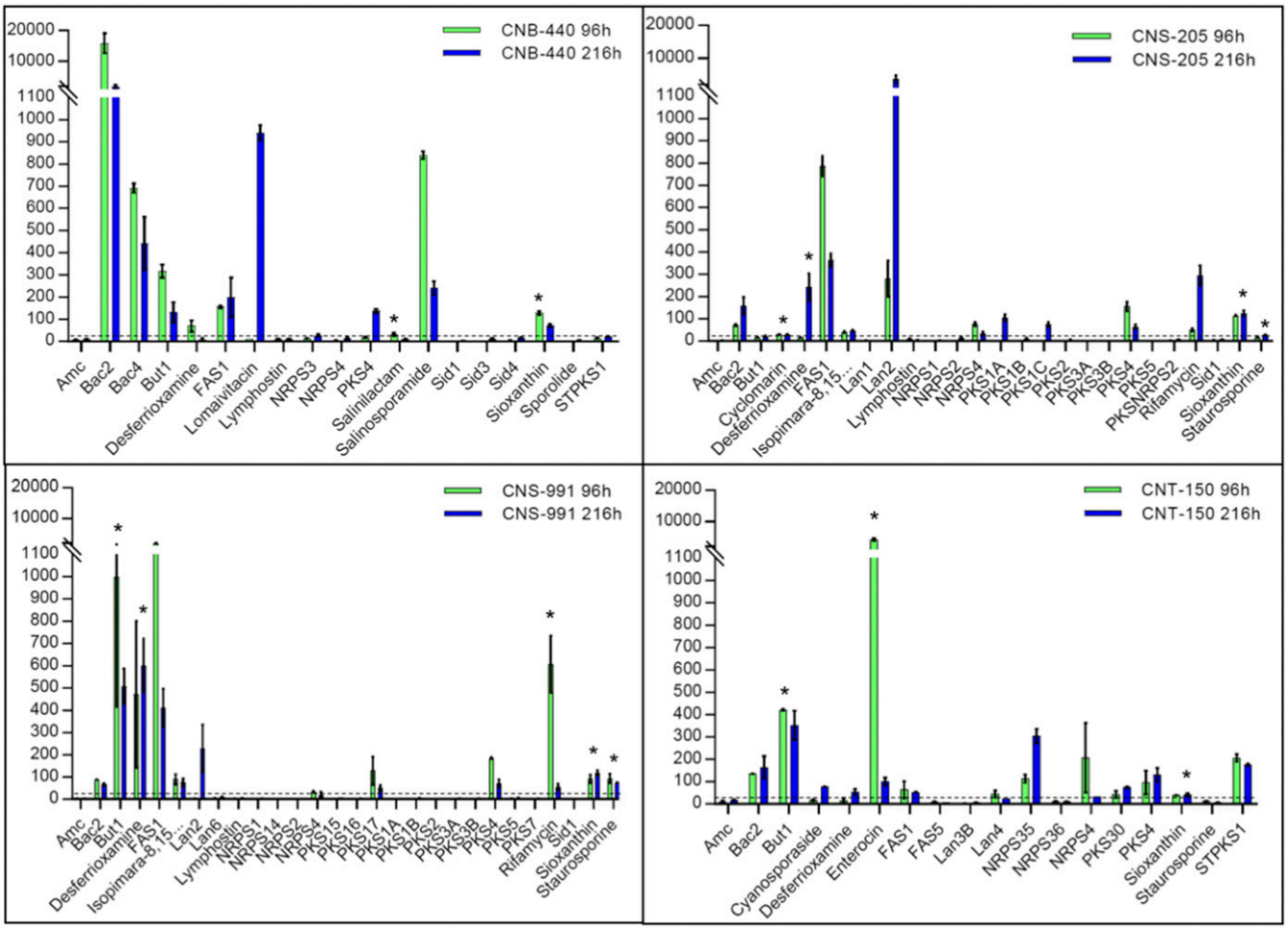

Fig. 2. BGC transcription during exponential (96 h, green) and stationary (216 h, blue) growth in four *Salinispora* strains. Expression levels (*y* axis) are expressed as RPKM. The threshold for distinguishing between silent and expressed BGCs is indicated by a dashed line. The BGCs observed in each strain are listed on the *x* axis and assigned formal names in cases where the products are known. But1 was experimentally verified as part of this study. BGCs for which the products were detected are indicated with an asterisk (*).

and Table 1). Nearly 40% (14 of 36) of the orphan BGCs were expressed in at least one strain (Fig. 1), and 75% of all expressed BGCs were expressed across both growth phases (Fig. 2). In most cases, transcription was a good indicator of compound production, with known compounds detected in 12 of 18 cases (67%) in which a BGC was transcribed above the established detection threshold.

Differential Expression of Shared BGCs. Of the 49 distinct BGCs observed in the four strains, 23 were shared between at least two strains (Fig. 1). Grouping similar BGCs into "operational biosynthetic units" (45) or gene cluster families (4) based on the prediction that they encode similar metabolites is the primary mechanism used to assess biosynthetic diversity among strains (45, 46). While examining groups of similar BGCs has provided insight into how these gene collectives evolve (47–49), this process overlooks the possibility that some shared BGCs are not equally expressed. Thus, the simple presence or absence of a BGC may not be a sufficient predictor of biosynthetic potential. Of the 23 shared BGCs identified in the four *Salinispora* strains, 13 were expressed in at least one of the strains and five of these (38%) were differentially expressed among strains (Fig. 1). Although there was no a priori mechanism to predict which version of a BGC would be expressed based on gene content alone, a comparison of expressed and nonexpressed BGCs revealed subtle differences that provide important clues as to why some BGCs were silent. One example is STPKS1, an orphan BGC that was expressed in both growth phases in *S. pacifica* strain CNT-150 but was silent in *S. tropica* strain CNB-440 (Fig. 2). The expressed version of this BGC includes an *araC* activator (50) upstream from the core polyketide synthase (PKS) operon (Fig. 3). In the silent BGC, this activator has been replaced with a family IS285 transposase (51) and the upstream ORFs share no homology with the expressed cluster. Given that mobile genetic elements such as transposes often flank BGCs, it would be difficult to predict that the *S. tropica* CNB-440 version of STPKS1 is silent if it was the only version of the BGC to be sequenced. By coupling comparative genomic and transcriptomic data, it becomes possible to predict which versions of a BGC are silent and to develop testable hypotheses as to why. Surprisingly, when the distribution of STPKS1 is expanded to a larger number of strains for which genome sequences are available, the silent version is largely maintained within a well-supported clade in the *S. tropica* species phylogeny (*SI Appendix*, Fig. S3). This conservation was unexpected, given that the loss of a key regulatory element suggests that this BGC may be permanently silenced. An alternative hypothesis is that the silent BGC is under different regulatory control. While this remains to be tested, differentiating between expressed and silent BGCs can be particularly useful as a guide to selecting strains for product discovery or heterologous expression.

A second example of differential BGC expression was observed between the two *S. arenicola* strains included in this study. In this case, PKS1A, which was predicted using the NaPDoS webtool (52) to encode the biosynthesis of an enediyne, was expressed in strain CNS-205 but silent in strain CNS-991 (Fig. 3). Comparative analysis of the two BGCs revealed an upstream regulatory element (sigma factor) that was expressed in CNS-205 but silent in CNS-991. In CNS-991, the sigma factor was flanked by hypothetical proteins, inverted in orientation, and in a distinct gene neighborhood relative to CNS-205, suggesting that this change may be linked to the lack of BGC expression. Once again, the distribution of the nonexpressed version of the BGC was highly conserved within the *S. arenicola* phylogeny (*SI Appendix*, Fig. S4), suggesting it may under different regulatory control.

Another example of differential BGC expression was observed with PKS4, which codes for a type II PKS predicted to produce the black spore pigment (7) that is typical of *Salinispora* and *Micromonospora* species. While this BGC was expressed in all four strains, different portions of the BGC were expressed in different strains (Fig. 4). More specifically, the entire BGC was expressed in both *S. tropica* strain CNB-440 and *S. pacifica* strain CNT-150. However, in both *S. arenicola* strains (CNS-205 and CNS-991), only the left half of the BGC (PKS4A) was expressed, while the right half (PKS4B) was silent. A comparison of the BGCs associated with these two expression patterns revealed that the fully expressed BGCs maintain either one (*S. pacifica*) or two (*S. tropica*) *luxR* homologs, while the partially expressed versions maintain only small genes encoding hypothetical proteins of unknown

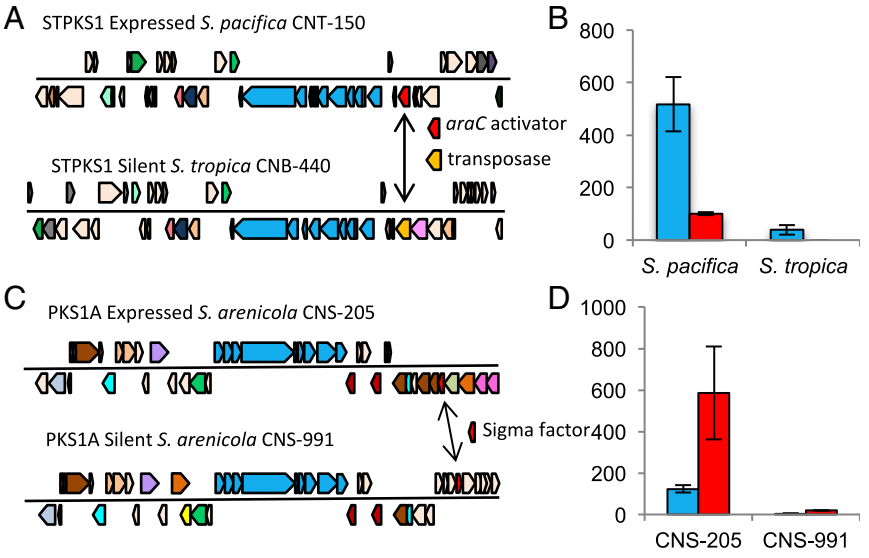

Fig. 3. STPKS1 and PKS1A BGCs. (*A*) Comparative genomics of the STPKS1 BGC observed in *S. pacifica* strain CNT-150 and *S. tropica* strain CNB-440. Blue ORFs indicate the core biosynthetic operon, including PKS genes. (*B*) Differences in the expression (RPKM) of core biosynthetic genes (blue) and *araC* regulator (red) between *S. pacifica* and *S. tropica*. Similar differences in expression were observed across the downstream portions of the BGCs. (*C*) Comparative genomics of the PKS1A BGC observed in two *S. arenicola* strains. Blue ORFs indicate the core biosynthetic operon, including PKS genes. Red ORFs indicate regulator elements, including a sigma factor. (*D*) Differences in the expression (RPKM) of core biosynthetic genes (blue) and sigma factor (red) between *S. arenicola* CNS-205 and *S. arenicola* CNS-991.

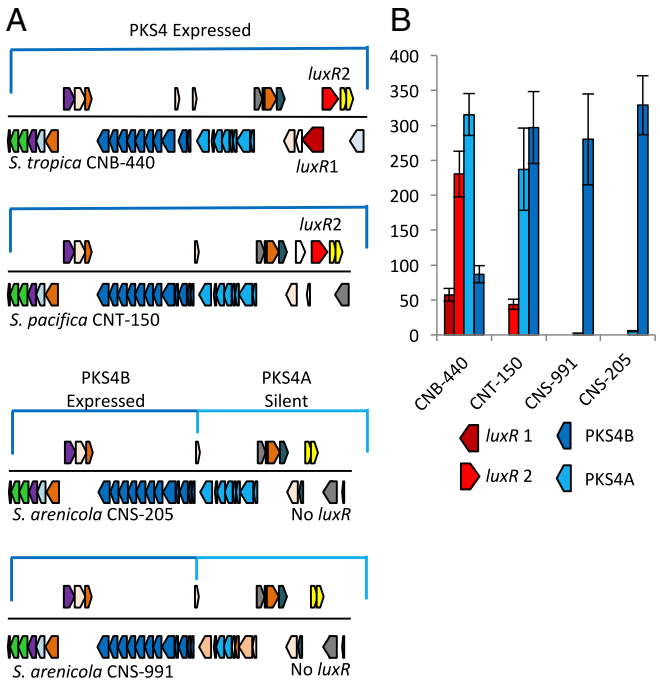

Fig. 4. PKS4 expression. (*A*) Comparison of the PKS4 BGC in four *Salinispora* strains. Dark red, *luxR* homolog 1; light red, *luxR* homolog 2; bright yellow, conserved *crcB* proteins (ion exporters). The BGC was divided into left (PKS4B, dark blue) and right (PKS4A, light blue) regions based on the expression patterns. (*B*) Expression levels of PKS4A (light blue) and PKS4B (dark blue) reported in RPKM.

function in place of the *luxR* regulators. Outside of these upstream differences, all four versions of the BGC are largely identical. These observations suggest that only one copy of *luxR* is needed for PKS4A expression, the deletion of both *luxR* regulatory elements is associated with the silencing of the PKS4A portion of the BGC, and control of the PKS4B portion of the BGC is regulated by a different mechanism. These differences are conserved between *S. arenicola* and *S. tropica* at the species level for all *Salinispora* strains for which genome sequences are available, while some *S. pacifica* strains show additional variations. There were clear phenotypic differences associated with the PKS4 expression patterns, with darkening observed in liquid cultures of *S. tropica* and *S. pacifica* (entire BGC expressed), suggesting they are producing the spore pigment. In contrast, the *S. arenicola* cultures, which only expressed the PKS4B portion, remained bright orange throughout the fermentation (*SI Appendix*, Fig. S5), suggesting that this portion of the BGC is not sufficient for pigment production. In addition, essential sporulation genes such as *ssgA* (53) were highly expressed in *S. tropica* and *S. pacifica* but silent in both *S. arenicola* strains, providing further support that these latter two strains were not sporulating and that the PKS4A component of the BGC is critical for spore pigment biosynthesis.

The fourth example of differential expression was observed in the *sta* BGC and provides a different paradigm in which it was not possible to infer meaningful correlations between BGC gene content and expression patterns. The production of staurosporine was previously recognized as a consistent phenotypic trait of *S. arenicola* (54), with the associated BGC (*sta*) fixed among globally distributed strains (45). Staurosporine production has also been reported sporadically among *S. pacifica* strains (47). The transcriptome analyses revealed that the *sta* BGC was expressed in both *S. arenicola* strains, while the *S. pacifica* version was silent (Fig. 5). Correspondingly, staurosporine was detected in the culture extracts of both *S. arenicola* strains but not in *S. pacifica* CNT-150 (Fig. 1). While the gene content and organization of the *sta* BGC is largely the same in all three strains, the sequence identity of the *malT* regulatory gene shared 99% amino acid identity between the two *S. arenicola* strains yet only 82% with *S. pacifica* strain CNT-150 (Fig. 5). Furthermore, this regulator was 99% conserved among all 62 *S. arenicola* strains for which genome sequences are available, while the 45 *S. pacifica* strains differed by up to 13%. While it remains unknown if *malT* is associated with the lack of expression in CNT-150, these results illustrate the challenges associated with predicting which BGCs will be expressed under a given set of growth conditions.

The final example of a shared but differentially expressed BGC is NRPS4, which was expressed in *S. arenicola* and *S. pacifica* but not in *S. tropica* (CNB-440). A comparison of the NRPS4 BGC in the four strains reveals numerous differences in gene content, including a number of upstream regulatory genes (*SI Appendix*, Fig. S6). In particular, a xenobiotic response element was missing from the silent *S. tropica* strain but expressed in the other three strains in conjunction with the core biosynthetic operon. While the absence of this element in

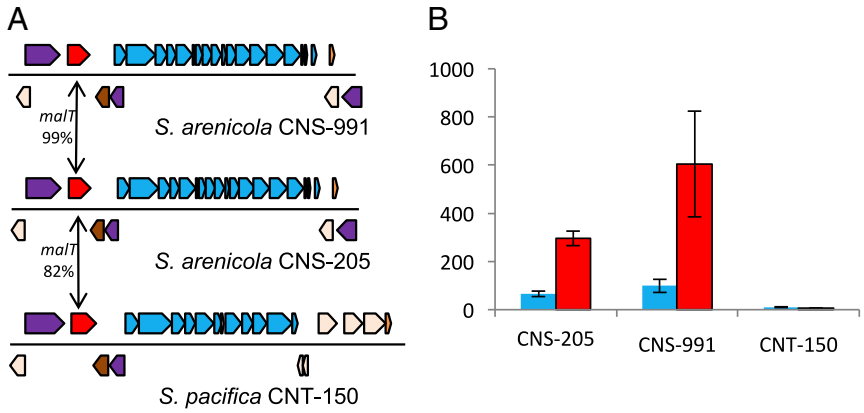

Fig. 5. Staurosporine BGC. (*A*) Comparative analysis of the staurosporine (*sta*) BGC in three *Salinispora* strains. Core biosynthetic operon (blue) and *malT* (red, *luxR* homolog) were observed in both *S. arenicola* strains and *malT*-like (82% amino acid identity) ATP-dependent transcriptional regulator in *S. pacifica* CNT-150. The upstream purple ORF present in both *S. arenicola* strains but absent in *S. pacifica* is annotated as a major facilitator superfamily transporter. These are known drug efflux proteins that contribute to antibiotic resistance. (*B*) Average expression of the *sta* biosynthetic operon in each of the three strains (blue) and expression of the regulatory element (red) are illustrated. Expression values are given in RPKM.

S. tropica may play a role in the lack of BGC expression, there are many other genetic differences that could be involved. In all five of these examples, the ability to differentiate among silent and expressed versions of similar BGCs was made possible by analyzing transcriptome data from closely related strains.

Relationships Between BGC Conservation and Expression. Comparative genomics has revealed that some *Salinispora* BGCs are conserved at the genus or species level (32, 45, 54), suggesting they encode ecologically important traits (55). This led us to explore the concept that conserved BGCs would be more commonly expressed since they potentially encode traits that provide a selective advantage. To test this hypothesis, we first calculated the frequency at which the 49 BGCs detected in the four strains occurred in the 119 *Salinispora* genomes (*SI Appendix*, Table S2). A comparison of expression relative to the frequency of occurrence revealed that BGCs conserved at the genus level (present in all strains) were expressed more frequently than those that were not fully conserved ($\chi^2 = 6.46$, $P = 0.011$). Furthermore, the frequency at which BGCs occurred within the genus was correlated with the likelihood of expression (Wald = 5.097, $P = 0.024$; $\chi^2 = 5.357$, $P = 0.021$). Finally, we could show that BGCs conserved at the species level (i.e., in all strains of that species) were expressed significantly more frequently than nonconserved BGCs ($\chi^2 = 75.938$, $\chi^2 = 71.357$, $\chi^2 = 55.97$, $P < 0.001$ in all cases). These results suggest that BGC conservation can provide a useful proxy for the probability of expression.

Metabolomics. Strain CNT-150 possessed the largest proportion of expressed BGCs (Table 1), including two orphan BGCs (NRPS35 and PKS30) that were unique to this strain and one (STPKS1) that was only expressed in this strain (Fig. 2). CNT-150 was therefore selected for metabolomic analyses. We previously reported the use of pattern-based genome mining to link tandem MS (MS/MS)-generated parent ions to BGC distributions (22). Here, we complimented this approach with transcriptome data in an effort to facilitate product discovery. A molecular network generated using extracts from CNT-150 and an existing library of *Salinispora* MS/MS data (56) revealed a number of parent ions that were unique to this strain, including enterocin, which could be readily identified using the Global Natural Products Social Molecular Networking (GNPS) dereplication function (57) (Fig. 6). There were numerous other parent ions of interest, including some molecular families that could not be linked to any known *Salinispora* compounds. In an attempt to identify these compounds, a larger culture extract was generated. Enterocin was readily identified in this new extract, along with a series of peaks that could not be identified by comparison with an in-house library of UV absorption spectra (*SI Appendix*, Fig. S7). Purification of the associated compounds followed by NMR and mass spectral analyses led to the identification of 10 members of the recently described salinipostin family (58) (*SI Appendix*, Figs. S8 and S9), a compound not known to be produced by *S. pacifica*. The salinipostins were readily detected using HPLC and UV detection, and could subsequently be identified as single nodes in the MS/MS-based network analysis (Fig. 6). Interestingly, the yields of salinipostin C from strain CNT-150 (0.5 mg/L) were ∼1,000-fold greater than those originally reported (0.4 μg/L) (58), suggesting it is the product of one of the highly expressed BGCs.

Experimental Validation of the Salinipostin BGC. The six highly expressed BGCs in CNT-150 that had not been linked to their products (bac2, NRPS4, NRPS35, STPKS1, PKS4, and but1) represented obvious candidates for salinipostin biosynthesis. Bac2, NRPS4, and NRPS35 could be ruled out, given that they are predicted to encode peptides, while STPKS1 is predicted to encode an enedyine based on a phylogenetic analysis of the KS

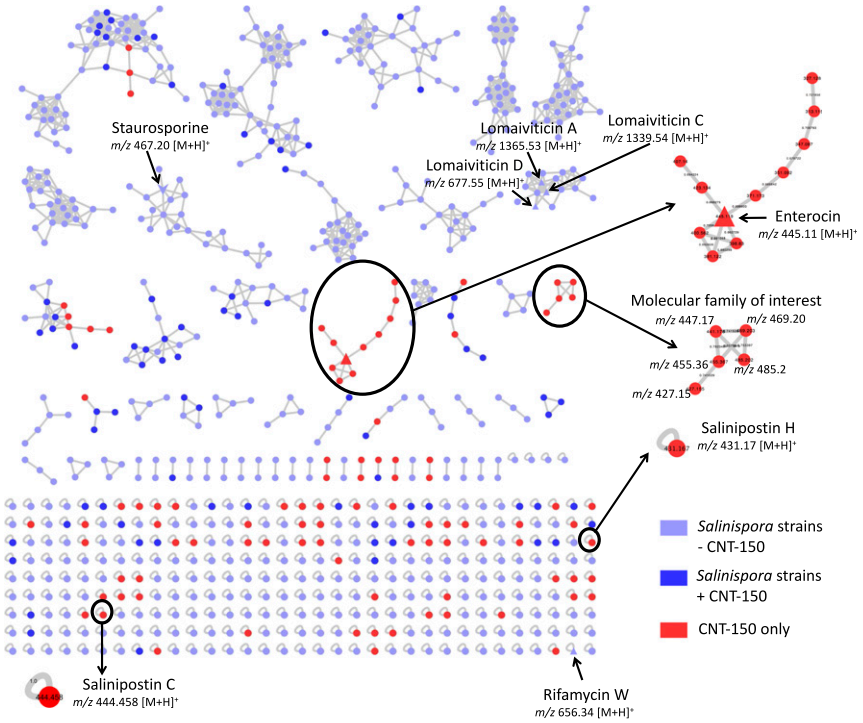

Fig. 6. MS/MS molecular network. CNT-150 culture extracts were networked with a database of MS/MS data previously acquired from *Salinispora* strains. Nodes containing fragmentation spectra that matched library reference spectra are represented as triangles or arrowheads and labeled with the compound name. Nodes containing parent ions not observed in CNT-150 are colored purple, nodes observed in the database and CNT-150 are colored blue, and nodes containing parent ions unique to CNT-150 are colored red.

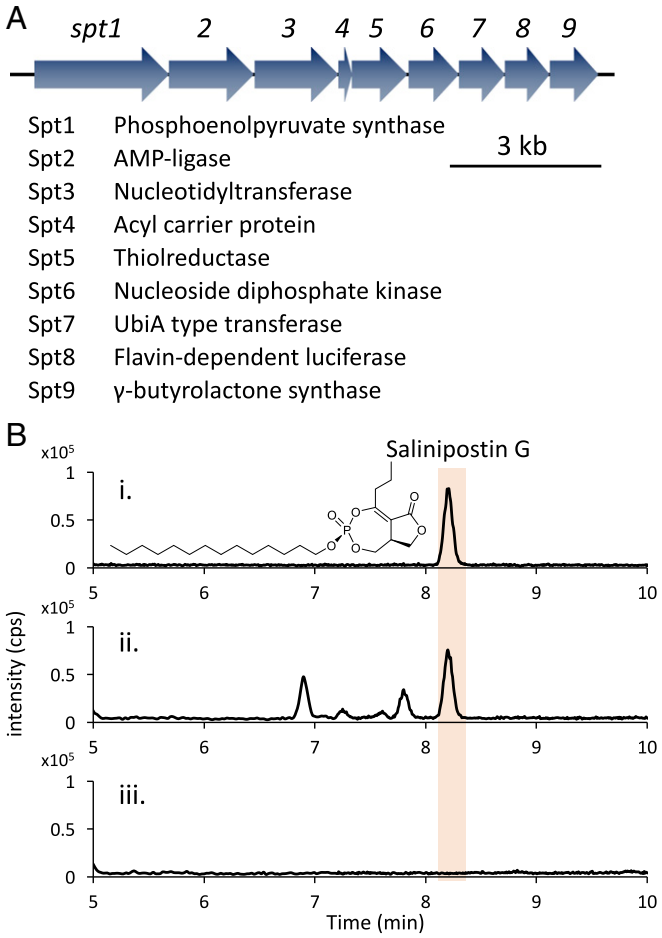

Fig. 7. Experimental characterization of the but1 (*spt*) BGC. (*A*) Salinipostin (*spt*) BGC. (*B*) Extracted ion chromatograms for (*i*) salinipostin G standard; (*ii*) extract of wild-type *S. tropica* CNB-440 (peaks at 6.8–7.8 min show $[M+H]^+$ and $[M+Na]^+$ spectra identical with salinipostin G, and therefore likely to correspond to different salinipostin isomers); and (*iii*) the extract of *S. tropica* CNB-440/Δ*spt9*.

domain. PKS4 is a type II PKS predicted to encode the biosynthesis of the black spore pigment and can likewise be ruled out. This leaves the highly expressed but1 BGC as the best candidate for salinipostin biosynthesis. AntiSMASH (https://antismash.secondarymetabolites.org/) annotated the nine-ORF but1 BGC (Fig. 7*A*) as encoding for the biosynthesis of a butyrolactone, a class of compounds that includes A-factor (59), and other five-membered ring lactone signaling agents from *Streptomyces* spp. that structurally resemble the salinipostins. The but1 BGC was observed in all four *Salinispora* strains and expressed in three (Fig. 1). With information about the compounds in hand, they could subsequently be detected in extracts of two strains in which the BGC was expressed (Fig. 1 and *SI Appendix*, Fig. S2 *I* and *J*). To establish a formal link between but1 and salinipostin biosynthesis, we deleted the gene *spt9* in *S. tropica* CNB-440 by double-crossover homologous recombination. LC/MS analysis of the resulting mutant *S. tropica* CNB-440/Δ*spt9* revealed the selective loss of salinipostin G (Fig. 7*B*), thereby confirming that but1 (now renamed *spt*) codes for salinipostin biosynthesis.

Discussion

The resurgence of interest in natural products research is being driven by increased access to genome sequencing, a better understanding of the molecular genetics of natural product biosynthesis, and the observation that even well-studied bacteria maintain considerable unrealized biosynthetic potential (60). Given the implications for natural product discovery, it is somewhat surprising that the relationships between orphan BGCs (those that have not been linked to their products) and silent BGCs (those transcribed at levels below the threshold at which their products can be detected) remain poorly characterized. Here, we provide evidence that 42–72% of the BGCs observed in four strains grown under one set of standard laboratory conditions were expressed at levels that should facilitate product discovery. This does not support the frequently cited concept that most BGCs are silent but, instead, suggests that our failure to detect considerable biosynthetic potential is linked to other factors, including the extraction and analytical techniques employed (56) and subjective decisions about what constitutes a natural product worthy of isolation and structure elucidation. These factors likely come into play for the highly expressed bac2 and bac4 BGCs in CNB-440. These are predicted to encode ribosomal peptides, which are often not recovered during small-molecule discovery efforts. Our observation that most BGCs are not silent supports previous observations made with the myxobacterium *Myxococcus xanthus* using both global microarrays and proteomics (18, 19). While the temporal relationships between BGC expression and specialized metabolite assembly remain largely unknown, our detection of products in 12 of 18 cases where experimentally characterized BGCs were expressed above the threshold level established for this study indicates a strong correlation between these processes.

During bacterial cultivation, it is generally accepted that most specialized metabolism is linked to nutrient limitation following the growth phase (27, 61). Here, we show that 75% of the expressed BGCS were transcribed in both the exponential and stationary phases of growth. This unexpected result suggests that, at least at the level of transcription, the genus *Salinispora* does not conform to this traditional model. There is evidence to suggest that many BGCs in the model organisms *Streptomyces coelicolor* A3 (2, 62) and *M. xanthus* (18) are also expressed in the growth phase, and thus that this phenomenon may be more common than generally recognized. The frequent detection of expressed BGCs early in the growth cycle supports the concept that the "secondary" nature of these compounds should be reconsidered (63).

A variety of tools are now available to identify BGCs from DNA sequence data (4). The ability to group similar BGCs into gene cluster families (46, 64) or operational biosynthetic units (45) based on the prediction that they encode the biosynthesis of similar specialized metabolites is a challenging but critical aspect of assessing BGC diversity and distributions among sequenced genomes (10) or environments (65). Here, we show that biosynthetic potential is more complicated than the simple presence or absence of a BGC, as more than one-third (38%) of the BGCs expressed in at least one strain were differentially expressed between strains. Many of these differentially expressed BGCs could be distinguished based on subtle differences in regulatory elements that appear to account for the expression profiles. In the case of STPKS1 in *S. tropica* strain CNB-440, the replacement of an *araC* activator with a transposase suggests that this BGC may be permanently silenced. While this remains to be verified, it is noteworthy that the detection of a transposase at the border of a BGC is more likely to be interpreted as evidence of horizontal gene transfer than inactivation (4). This speaks to the value of comparing the expression levels of similar BGCs across multiple strains, which can yield clues as to which variants are more likely to yield small-molecule products and why.

It is more difficult to explain why some silent BGCs show evidence of conservation (*SI Appendix*, Figs. S2 and S4). It remains possible that lost regulatory elements are complimented *in trans*, thus rendering what appears to be a silent BGC functional under a different set of conditions. If so, this raises the intriguing

possibility that BGCs predicted to encode the biosynthesis of the same compounds may be expressed in response to different environmental cues, and thus serve different ecological functions. While there is evidence that a single biosynthetic pathway yields products with different ecological functions (66), the concept that a single product may serve different functions under different conditions adds considerable complexity to our understanding of chemical ecology (67, 68). Alternatively, there may be benefits associated with maintaining certain components of a silent BGC, such as a resistance gene, thereby selecting against loss of the entire cluster. There is evidence of BGC degradation in the *Salinispora* genomes (32), suggesting that silencing may represent a first step in the gradual breakdown and ultimate loss of some BGCs. While these interpretations remain speculative, BGC inactivation has been reported in fungi (69) and warrants further consideration in the strains studied here.

While many orphan BGCs were expressed, the observation that some are differentially expressed among strains helps explain the historical importance of strain selection. While there is no way to predict a priori which BGCs will be expressed under a given set of conditions, comparative genomics coupled with expression profiling can help prioritize strains and identify key regulatory elements. This has important implications for synthetic biology, as the likelihood for successful heterologous expression is greater if the BGC is functional in the native organism. Furthermore, transcriptome profiles can provide an effective method to prioritize strains for study. This approach facilitated the identification of a series of compounds not previously reported from *S. pacifica* and the experimental verification of its associated BGC. Population genomics is providing valuable insight into the diversity, distributions, and evolution of natural product BGCs. Comparative transcriptomics adds a new level of understanding to genome mining efforts by providing insight into which BGCs are expressed and clues as to why some remain silent.

Methods

Genomic and Phylogenetic Analyses. A total of 119 *Salinispora* genome sequences, generated in collaboration with the Joint Genome Institute, were downloaded from the Integrated Microbial Genomes & Microbiomes (IMG/ER) database (https://img.jgi.doe.gov/). BGCs were annotated using antiSMASH and manual blast analyses using key biosynthetic genes as queries (7, 45). Four strains were selected for transcriptome analyses based on the number of different BGCs they maintained. These strains were S. *arenicola* CNS-205 (accession no. NC_009953.1), *S. arenicola* CNS-991 (accession no. NZ_ARBB00000000.1), *S. pacifica* CNT-150 (accession no. NZ_AQZW00000000), and *S. tropica* CNB-440 (accession no. NC_009380.1). Similar BGCs in different strains were assigned the same BGC identifier (e.g., STPKS1) based on the sequence identity of relevant sequence tags such as ketosynthase and condensation domains (45) and the architecture of key biosynthetic operons. A species phylogeny was generated from nucleotide sequences for 10 single-copy genes (*dnaA*, *gyrB*, *pyrH*, *recA*, *pgi*, *trpB*, *atpD*, *sucC*, *rpoB*, and *topA*) extracted from the 119 *Salinispora* genome sequences, aligned using Muscle in Geneious Pro Version 5.5.6 and concatenated using Mesquite (v2.74) (70). The nucleotide model GTR + G was used to create a maximum likelihood tree using RaxML implemented on the CIPRES Science Portal (71).

Strain Cultivation. Growth curves were generated for each of the four strains to establish time points for transcriptome and metabolome analyses. Starter cultures were inoculated from frozen stocks into 50 mL of medium A1 (10 g·L^{-1} starch, 4 g·L^{-1} yeast extract, 2 g·L^{-1} peptone, and 1 L of 0.2-μm filtered seawater) in 250 mL flasks and grown for 5 d at 25 °C with shaking at 160 rpm (New Brunswick Innova 2300). Starter cultures (1 mL) were then used to inoculate each strain into triplicate flasks containing 50 mL of A1 and glass beads to reduce clumping. Optical density (600 nm) was monitored at 24-h intervals, with three readings averaged for each replicate culture at each time point. Based on the results of the growth curves, 96 h and 216 h were selected as time points for the transcriptome and metabolome analyses.

Transcriptome Analyses. At each time point, RNA was extracted from 5 mL of culture following an acid phenol/chloroform/isoamyl alcohol procedure (72). RNA was sent to the US Department of Energy Joint Genome Institute for sequencing, quality control, and read mapping as previously described (32). In brief, Illumina HiSeq 2500 sequencing generated >3 × 10^7 paired-end reads (100 bp) per replicate. Using BBDuk (https://sourceforge.net/projects/bbmap/), raw reads were evaluated for artifact sequences by kmer matching (kmer = 25). Quality trimming was performed using the phred trimming method set at Q10 (73), with reads under 45 bases removed. Raw reads were aligned to their respective reference genome using Burrows–Wheeler Aligner (BWA) (74). FeatureCounts was used to generate raw gene counts (75). Mapped reads were visualized using BamView in Artemis (76). The number of RPKM was used to normalize raw data in Artemis (77). BGC expression levels were derived from average values calculated for key biosynthetic genes. These included PKS, nonribosomal peptide synthetase, terpene synthase, precursor peptide (bacteriocin), and LanM (lantibiotic) genes. Additional genes associated with key biosynthetic operons were checked to confirm the expression levels.

Metabolomics. Fifteen milliliters of culture was harvested and extracted with an equal volume of ethyl acetate from each of the four strains at the same time points used for the transcriptome analyses. The ethyl acetate layers were collected, dried in vacuo, dissolved in methanol (MeOH) at a final concentration of 2 mg/mL, and analyzed using an Agilent 6530 Accurate-Mass Q-TOF spectrometer coupled to an Agilent 1260 LC system with a Phenomenex Kinetex C18 reversed-phase column (2.6 mm, 100 mm × 4.6 mm). The LC conditions included 0.1% formic acid and a flow rate of 0.7 mL·min^{-1}: 1–2 min [10% acetonitrile (MeCN) in H_2O], 2–14 min (10–100% MeCN), and 14–15 min (100% MeCN). The divert valve was set to waste for the first 2 min. The Q-TOF MS settings during the LC gradient were as follows: 200–1,600 *m/z* positive ion mode; 3/s MS scan rate; 5/s MS/MS scan rate; 20-eV fixed collision energy; 300 °C source gas temperature; 11-L·min^{-1} gas flow; 45-psig nebulizer; and scan source parameters VCap = 3,000, fragmentor = 100, skimmer1 = 65, and OctopoleRFPeak = 750. The MS was auto-tuned using Agilent tuning solution in positive mode before each measurement.

The extracts were also run on an Agilent 1260 HPLC instrument coupled with an Agilent 6230 TOF mass spectrometer with Jet Stream electrospray ionization source operated under positive ion mode with the following parameters: VCap = 3,500 V, 160-V fragmentor voltage, 500-V nozzle voltage, 325 °C drying gas temperature, 325 °C sheath gas temperature, 7-L·min^{-1} drying gas flow rate, 10-L·min^{-1} sheath gas flow rate, and 40-psi nebulizer pressure. Chromatographic separations were performed at room temperature on a Phenomenex EVO C-18 column (2.1-mm i.d. × 50-mm length, 2.6-μm particle size). The LC conditions included 0.1% formic acid and a flow rate of 0.7 mL·min^{-1}: 1–2 min (10% MeCN in H_2O), 2–14 min (10–100% MeCN), and 14–20 min (100% MeCN). The divert valve was set to waste for the first 2 min, and the gradient was followed by a 2-min post-run. HPLC-UV data were compared with an in-house library to facilitate compound identification. MS/MS data were analyzed with MassHunter software (Agilent) and compared with known compounds and crude extract spectral libraries stored in the GNPS database (57).

The MS/MS data were converted to mzXML from Agilent MassHunter data files (.d) using the Trans-Proteomic pipeline (78). Files were then uploaded to the Mass Spectrometry Interactive Virtual Environment (MASSive) server (https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp) via the file transfer protocol client Filezilla (https://filezilla-project.org). Uploaded files were accessed and networked using the GNPS pipeline (57) using previously validated parameters (22, 79). Only parent ions that fragmented at least two times were considered in the network. Fragmentation spectra with a cosine similarity score above 0.65 and six matching fragment ions were connected with edges. Library matches needed to share at least eight peaks and a similarity score above 0.55. Networks were visualized using Cytoscape (www.cytoscape.org/cy3.html) with the built-in solid layout.

Compound Isolation and Structure Determination. *S. pacifica* CNT-150 was inoculated into 25 mL of A1 media from a frozen glycerol stock and cultured for 4 d at room temperature (28 °C) with shaking at 160 rpm, after which 5 mL was inoculated into 3 × 250-mL flasks containing 50 mL of A1. After 4 d at room temperature with shaking (160 rpm), 10 mL of these starter cultures was inoculated into 10 × 2.8-L Fernbach flasks each containing 1 L of A1. After 9 d with shaking at 160 rpm at room temperature, the cultures were extracted 1:1 with ethyl acetate, the organic layers were combined, and the solvent was removed in vacuo. The crude CNT-150 extract was fractionated using C-18 reversed-phase vacuum column chromatography and a MeOH/H_2O and step gradient elution (with 20%, 40%, 60%, 80%, and 100% MeOH in H_2O) to yield five fractionations, each of which was further analyzed by reversed-phase HPLC with an Phenomenex Luna C18 (2) 5-μm 100 × 4.6-mm

column using a linear gradient of 10–100% aqueous MeCN for 20 min with elution and thereafter isocratic with the same solvent for 10 min with UV detection (254 nm). LC (diode array detection) data were analyzed with ChemStation software (Agilent) and compared with an in-house spectral library of known *Salinispora* compounds.

Statistical Analysis. All statistics were performed using SPSS version 24 (IBM). The χ^2 test was used to test for significant differences in the frequency of expressed core BGCs vs. nonconserved BGCs. A binomial logistic regression was used to determine the relationship between the phylogenetic level of conservation of a BGC and the likelihood of expression, with the Wald test used to determine variable significance.

Salinipostin BGC Functional Studies. The gene deletion vector pIJ773-Δ*spt9* (procedure for construction is described below) was introduced into *S. tropica* CNB-440 via intergeneric conjugation, and the Apr^R exoconjugates from double crossover were selected as previously described (34). *S. tropica* CNB-440/Δ*spt9* was isolated from Apr^s clones obtained from the repeated subculture of Apr^R clones on the antibiotic-free plate. DNA fragment A containing the upstream 3.0-kb region of *spt9* was amplified by PCR with primer 1, 5′-cccccgggctgcaggaattc**acctcgaacgcgcctactggcac**-3′ (annealing sequence boldfaced) and primer 2, 5′-tgccatcgatgatca**tagccgggtgtcggtcatcgattg**-3′, and fragment B containing the downstream 3.0-kb region of *spt9* was amplified by PCR with primer 3, 5′-accgacacccggcta**tgatcatcgatggcaccctgtcgac**-3′ and primer 4, 5′-ccagcctacacatcgaattc**gatcggcacggcgatggtggacag**-3′. With Gibson assembly (New England Biolabs), fragments A and B were assembled with pIJ773-derived fragment amplified by primer 5, 5′-gaattcctgcagcccgggggatc-3′ and primer 6, 5′-gaattcgatgtgtaggctggag-3′, to give pIJ773-Δ*spt9*.

Wild-type *S. tropica* CNB-440 and the mutant prepared above were cultured in 250-mL Erlenmeyer flasks containing 50 mL of medium A1 + BFe [A1 + 1 g·L^{-1} $CaCO_3$, 40 mg·L^{-1} $Fe_2(SO_4)_3$•$4H_2O$, 100 mg·L^{-1} KBr] at 28 °C with shaking at 230 rpm (New Brunswick Innova 2300). Sterile Amberlite XAD-7 resin (1 g) was added to each flask after 48 h, and the fermentation was continued for an additional 4 d. The resin was collected by filtration using cheese cloth, washed with distilled water, and soaked in acetone for 2 h. The extract was concentrated in vacuo, and the resultant residue was dissolved in MeOH. After filtration, an aliquot of the extract was subjected to LC/MS using a Phenomenex Kinetex XB-C_{18} column (2.6 μm, 100 × 4.6 mm) with the following conditions: positive mode electrospray ionization, 0.1% formic acid in H_2O, flow rate 1.0 mL·min^{-1}, 0–8 min (80–90% MeCN in H_2O), 8–13 min (90–100% MeCN), 13–15 min (100% MeCN) with the divert valve set to waste for the first 3 min. The ion at *m/z* 445.2714 (salinipostin G) corresponding to the $[M+H]^+$ ion was analyzed in the extracted ion chromatograms.

ACKNOWLEDGMENTS. We thank K. S. Ryan, Y. L. Du, and R. G. Linington for valuable discussions and analytical assistance with the salinipostin BGC studies. We thank N. Millán-Aguiñaga for assistance with the phylogenetic analyses. This work was supported by NIH Grants GM085770, 2U19TW007401, and S10-OD010640, and postdoctoral fellowships from the Uehara Memorial Foundation (to T.A.) and the Japan Society for Promotion of Science (to Y.K.). Sequencing was conducted by the U.S. Department of Energy Joint Genome Institute and supported by the Office of Science of the US Department of Energy under Contract DEAC02-05CH11231.

1. Newman DJ, Cragg GM (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J Nat Prod* 75:311–335.
2. Harvey AL (2008) Natural products in drug discovery. *Drug Discov Today* 13:894–901.
3. Bachmann BO, Van Lanen SG, Baltz RH (2014) Microbial genome mining for accelerated natural products discovery: Is a renaissance in the making? *J Ind Microbiol Biotechnol* 41:175–184.
4. Medema MH, Fischbach MA (2015) Computational approaches to natural product discovery. *Nat Chem Biol* 11:639–648.
5. Zhang JJ, Moore BS (2015) Digging for biosynthetic dark matter. *Elife* 4:e06453.
6. Bentley SD, et al. (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417:141–147.
7. Udwary DW, et al. (2007) Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc Natl Acad Sci USA* 104: 10376–10381.
8. Corre C, Challis GL (2009) New natural product biosynthetic chemistry discovered by genome mining. *Nat Prod Rep* 26:977–986.
9. Lautru S, Deeth RJ, Bailey LM, Challis GL (2005) Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. *Nat Chem Biol* 1:265–269.
10. Cimermancic P, et al. (2014) Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* 158:412–421.
11. Medema MH, et al. (2015) Minimum information about a biosynthetic gene cluster. *Nat Chem Biol* 11:625–631.
12. Rutledge PJ, Challis GL (2015) Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat Rev Microbiol* 13:509–523.
13. Okada BK, Seyedsayamdost MR (2017) Antibiotic dialogues: Induction of silent biosynthetic gene clusters by exogenous small molecules. *FEMS Microbiol Rev* 41:19–33.
14. Chiang Y-M, Chang S-L, Oakley BR, Wang CC (2011) Recent advances in awakening silent biosynthetic gene clusters and linking orphan clusters to natural products in microorganisms. *Curr Opin Chem Biol* 15:137–143.
15. Yamanaka K, et al. (2014) Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc Natl Acad Sci USA* 111:1957–1962.
16. Tanaka Y, Hosaka T, Ochi K (2010) Rare earth elements activate the secondary metabolite-biosynthetic gene clusters in Streptomyces coelicolor A3(2). *J Antibiot (Tokyo)* 63:477–481.
17. Behnken S, Hertweck C (2012) Cryptic polyketide synthase genes in non-pathogenic Clostridium SPP. *PLoS One* 7:e29609.
18. Bode HB, et al. (2009) Identification of additional players in the alternative biosynthesis pathway to isovaleryl-CoA in the myxobacterium *Myxococcus xanthus*. *ChemBioChem* 10:128–140.
19. Schley C, Altmeyer MO, Swart R, Müller R, Huber CG (2006) Proteome analysis of *Myxococcus xanthus* by off-line two-dimensional chromatographic separation using monolithic poly-(styrene-divinylbenzene) columns combined with ion-trap tandem mass spectrometry. *J Proteome Res* 5:2760–2768.
20. Bumpus SB, Evans BS, Thomas PM, Ntai I, Kelleher NL (2009) A proteomics approach to discovering natural products and their biosynthetic pathways. *Nat Biotechnol* 27: 951–956.
21. Gubbens J, et al. (2014) Natural product proteomining, a quantitative proteomics platform, allows rapid discovery of biosynthetic gene clusters for different classes of natural products. *Chem Biol* 21:707–718.
22. Duncan KR, 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.
23. Meier JL, et al. (2009) An orthogonal active site identification system (OASIS) for proteomic profiling of natural product biosynthesis. *ACS Chem Biol* 4:948–957.
24. Rigali S, et al. (2008) Feast or famine: The global regulator DasR links nutrient stress to antibiotic production by *Streptomyces*. *EMBO Rep* 9:670–675.
25. Mao D, Bushin LB, Moon K, Wu Y, Seyedsayamdost MR (2017) Discovery of *scmR* as a global regulator of secondary metabolism and virulence in *Burkholderia thailandensis* E264. *Proc Natl Acad Sci USA* 114:E2920–E2928.
26. Liu G, Chater KF, Chandra G, Niu G, Tan H (2013) Molecular regulation of antibiotic biosynthesis in streptomyces. *Microbiol Mol Biol Rev* 77:112–143.
27. Bibb MJ (2005) Regulation of secondary metabolism in streptomycetes. *Curr Opin Microbiol* 8:208–215.
28. Jensen PR, Moore BS, Fenical W (2015) The marine actinomycete genus *Salinispora*: A model organism for secondary metabolite discovery. *Nat Prod Rep* 32:738–751.
29. Feling RH, et al. (2003) Salinosporamide A: A highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *salinospora*. *Angew Chem Int Ed Engl* 42:355–357.
30. Ogasawara Y, et al. (2016) Exploring peptide ligase orthologs in Actinobacteria–Discovery of pseudopeptide natural products, ketomemicins. *ACS Chem Biol* 11: 1686–1692.
31. Butler MS, Robertson AA, Cooper MA (2014) Natural product and natural product derived drugs in clinical trials. *Nat Prod Rep* 31:1612–1661.
32. Letzel AC, et al. (2017) Genomic insights into specialized metabolism in the marine actinomycete *Salinispora*. *Environ Microbiol* 19:3660–3673.
33. Lane AL, 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.
34. Schultz AW, et al. (2008) Biosynthesis and structures of cyclomarins and cyclomarazines, prenylated cyclic peptides of marine actinobacterial origin. *J Am Chem Soc* 130:4507–4516.
35. Roberts AA, Schultz AW, Kersten RD, Dorrestein PC, Moore BS (2012) Iron acquisition in the marine actinomycete genus *Salinispora* is controlled by the desferrioxamine family of siderophores. *FEMS Microbiol Lett* 335:95–103.
36. Kersten RD, et al. (2013) Bioactivity-guided genome mining reveals the lomaiviticin biosynthetic gene cluster in *Salinispora tropica*. *ChemBioChem* 14:955–962.
37. Miyanaga A, 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.
38. Eustáquio AS, et al. (2009) Biosynthesis of the salinosporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from S-adenosyl-L-methionine. *Proc Natl Acad Sci USA* 106:12295–12300.
39. Richter TKS, Hughes CC, Moore BS (2014) Sioxanthin, a novel glycosylated carotenoid reveals an unusual subclustered biosynthetic pathway. *Environ Microbiol* 17: 2158–2171.
40. McGlinchey RP, Nett M, Moore BS (2008) Unraveling the biosynthesis of the sporolide cyclohexenone building block. *J Am Chem Soc* 130:2406–2407.
41. Onaka H, Taniguchi S, Igarashi Y, Furumai T (2002) Cloning of the staurosporine biosynthetic gene cluster from *Streptomyces* sp. TP-A0274 and its heterologous expression in *Streptomyces lividans*. *J Antibiot (Tokyo)* 55:1063–1071.
42. Xu M, et al. (2014) Characterization of an orphan diterpenoid biosynthetic operon from *Salinispora arenicola*. *J Nat Prod* 77:2144–2147.
43. Bonet B, Teufel R, Crüsemann M, Ziemert N, Moore BS (2015) Direct capture and heterologous expression of *Salinispora* natural product genes for the biosynthesis of enterocin. *J Nat Prod* 78:539–542.

44. Wilson MC, Gulder TAM, Mahmud T, Moore BS (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.
45. Ziemert N, et al. (2014) Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*. *Proc Natl Acad Sci USA* 111:E1130–E1139.
46. Nguyen DD, et al. (2013) MS/MS networking guided analysis of molecule and gene cluster families. *Proc Natl Acad Sci USA* 110:E2611–E2620.
47. Freel KC, Nam S-J, Fenical W, Jensen PR (2011) Evolution of secondary metabolite genes in three closely related marine actinomycete species. *Appl Environ Microbiol* 77:7261–7270.
48. Nielsen JC, et al. (2017) Global analysis of biosynthetic gene clusters reveals vast potential of secondary metabolite production in *Penicillium* species. *Nat Microbiol* 2: 17044.
49. Cubillos-Ruiz A, Berta-Thompson JW, Becker JW, van der Donk WA, Chisholm SW (2017) Evolutionary radiation of lanthipeptides in marine cyanobacteria. *Proc Natl Acad Sci USA* 114:E5424–E5433.
50. Martin RG, Rosner JL (2001) The AraC transcriptional activators. *Curr Opin Microbiol* 4:132–137.
51. Hu P, et al. (1998) Structural organization of virulence-associated plasmids of Yersinia pestis. *J Bacteriol* 180:5192–5202.
52. Ziemert N, et al. (2012) The natural product domain seeker NaPDoS: A phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS One* 7: e34064.
53. van Wezel GP, et al. (2000) ssgA is essential for sporulation of *Streptomyces coelicolor* A3(2) and affects hyphal development by stimulating septum formation. *J Bacteriol* 182:5653–5662.
54. Jensen PR, Williams PG, Oh DC, Zeigler L, Fenical W (2007) Species-specific secondary metabolite production in marine actinomycetes of the genus *Salinispora*. *Appl Environ Microbiol* 73:1146–1152.
55. Patin NV, Duncan KR, Dorrestein PC, Jensen PR (2016) Competitive strategies differentiate closely related species of marine actinobacteria. *ISME J* 10:478–490.
56. Crüsemann M, et al. (2017) Prioritizing natural product diversity in a collection of 146 bacterial strains based on growth and extraction protocols. *J Nat Prod* 80: 588–597.
57. Wang M, et al. (2016) Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* 34: 828–837.
58. Schulze CJ, Navarro G, Ebert D, DeRisi J, Linington RG (2015) Salinipostins A-K, long-chain bicyclic phosphotriesters as a potent and selective antimalarial chemotype. *J Org Chem* 80:1312–1320.
59. Horinouchi S, Beppu T (1994) A-factor as a microbial hormone that controls cellular differentiation and secondary metabolism in *Streptomyces griseus*. *Mol Microbiol* 12: 859–864.
60. Nett M, Ikeda H, Moore BS (2009) Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat Prod Rep* 26:1362–1384.
61. Horinouchi S (2007) Mining and polishing of the treasure trove in the bacterial genus streptomyces. *Biosci Biotechnol Biochem* 71:283–299.
62. Jeong Y, et al. (2016) The dynamic transcriptional and translational landscape of the model antibiotic producer *Streptomyces coelicolor* A3(2). *Nat Comm* 7:11605.
63. Price-Whelan A, Dietrich LE, Newman DK (2006) Rethinking 'secondary' metabolism: Physiological roles for phenazine antibiotics. *Nat Chem Biol* 2:71–78.
64. Doroghazi JR, et al. (2014) A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat Chem Biol* 10:963–968.
65. Reddy BVB, Milshteyn A, Charlop-Powers Z, Brady SF (2014) eSNaPD: A versatile, web-based bioinformatics platform for surveying and mining natural product biosynthetic diversity from metagenomes. *Chem Biol* 21:1023–1033.
66. Gallagher KA, et al. (2017) Ecological implications of hypoxia-triggered shifts in secondary metabolism. *Environ Microbiol* 19:2182–2191.
67. Kubanek J, et al. (2002) Multiple defensive roles for triterpene glycosides from two Caribbean sponges. *Oecologia* 131:125–136.
68. Schmitt TM, Hay ME, Lindquist N (1995) Constraints on chemically mediated coevolution: Multiple functions for seaweed secondary metabolites. *Ecology* 76: 107–123.
69. Campbell MA, Rokas A, Slot JC (2012) Horizontal transfer and death of a fungal secondary metabolic gene cluster. *Genome Biol Evol* 4:289–293.
70. Maddison WP, Maddison DR (2011) Mesquite: A modular system for evolutionary analysis. Version 2.75. Available at mesquiteproject.org. Accessed July 1, 2014.
71. Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Gateway Computing Environments Workshop (GCE)*. Available at https://www.phylo.org/. Accessed July 1, 2014.
72. Nieselt K, et al. (2010) The dynamic architecture of the metabolic switch in Streptomyces coelicolor. *BMC Genomics* 11:10.
73. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8:175–185.
74. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589–595.
75. Liao Y, Smyth GK, Shi W (2014) featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30:923–930.
76. Rutherford K, et al. (2000) Artemis: Sequence visualization and annotation. *Bioinformatics* 16:944–945.
77. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5:621–628.
78. Deutsch EW, et al. (2010) Trans-Proteomic Pipeline supports and improves analysis of electron transfer dissociation data sets. *Proteomics* 10:1190–1195.
79. Scheubert K, et al. (2017) Significance estimation for large scale untargeted metabolomics annotations. bioRxiv:10.1101/109389.