

## **Supplementary Information for** Incipient genome erosion and metabolic streamlining for antibiotic production in a defensive symbiont.

Taras Y. Nechitaylo<sup>1,§</sup>, Mario Sandoval-Calderón<sup>2,§</sup>, Tobias Engl<sup>1,2</sup>, Natalie Wielsch<sup>3</sup>, Diane M. Dunn<sup>4</sup>, Alexander Goesmann<sup>5</sup>, Erhard Strohm<sup>6</sup>, Aleš Svatoš<sup>3</sup>, Colin Dale<sup>7</sup>, Robert B. Weiss<sup>4</sup>, Martin Kaltenpoth<sup>1,2,4,6,7,8,\*</sup>

<sup>1</sup>Research Group Insect Symbiosis, Max Planck Institute for Chemical Ecology, Jena, Germany

<sup>2</sup>Department of Evolutionary Ecology, Institute of Organismic and Molecular Evolution, Johannes Gutenberg University, 55128 Mainz, Germany.

<sup>3</sup>Research Group Mass Spectrometry, Max Planck Institute for Chemical Ecology, Jena, Germany.

<sup>4</sup>Department of Human Genetics, University of Utah, Salt Lake City, Utah, USA.

<sup>5</sup>Research Group Computational Genomics, CeBiTec, Bielefeld University, Germany.

<sup>6</sup>Department of Zoology, University of Regensburg, Regensburg, Germany.

<sup>7</sup>Department of Biology, University of Utah, Salt Lake City, Utah, USA.

<sup>8</sup>Department of Insect Symbiosis, Max Planck Institute for Chemical Ecology, Jena, Germany.

§equal contributions

\*Corresponding author

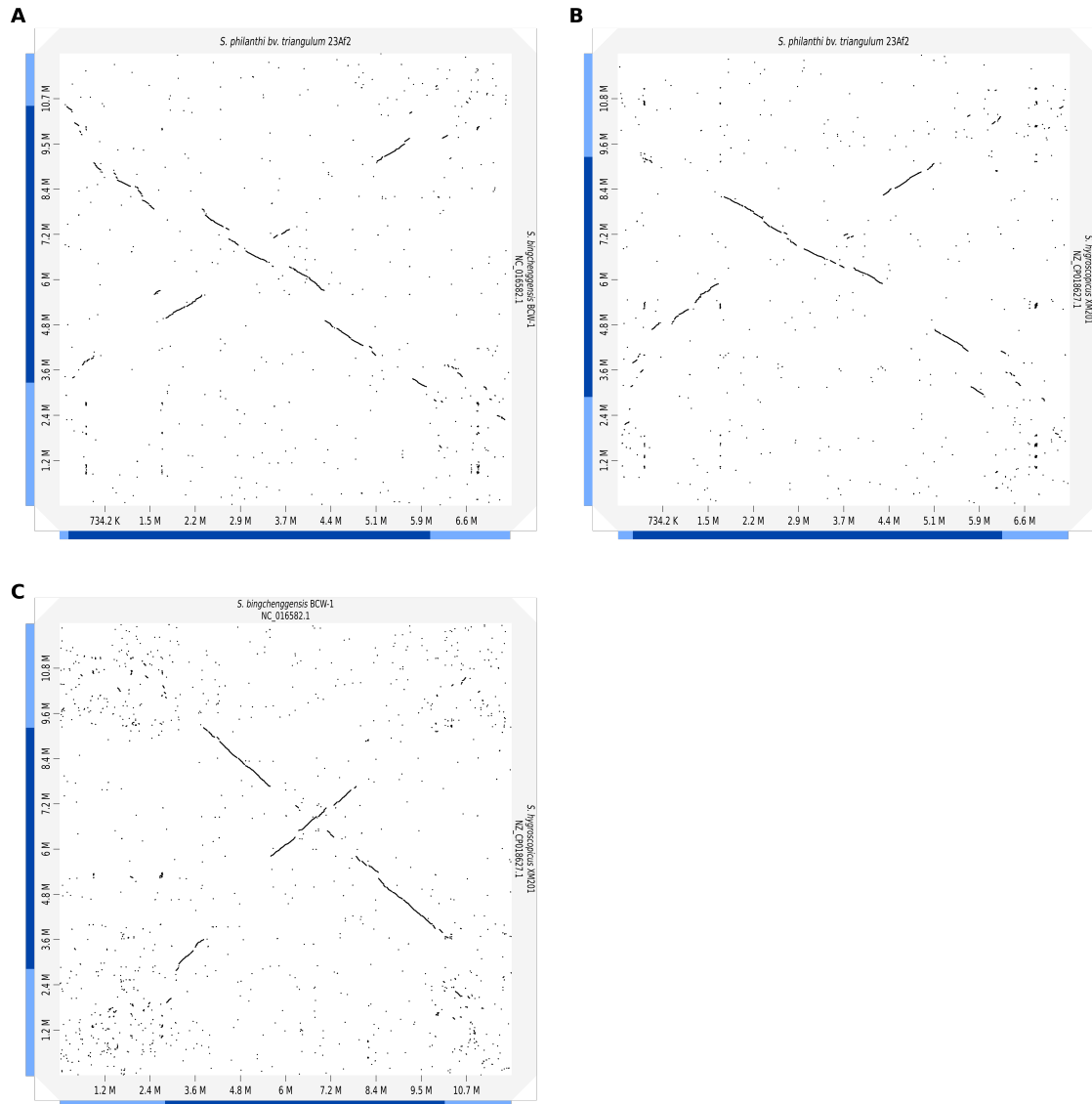
Email: [kaltenpoth@ice.mpg.de](mailto:kaltenpoth@ice.mpg.de)

### **This PDF file includes:**

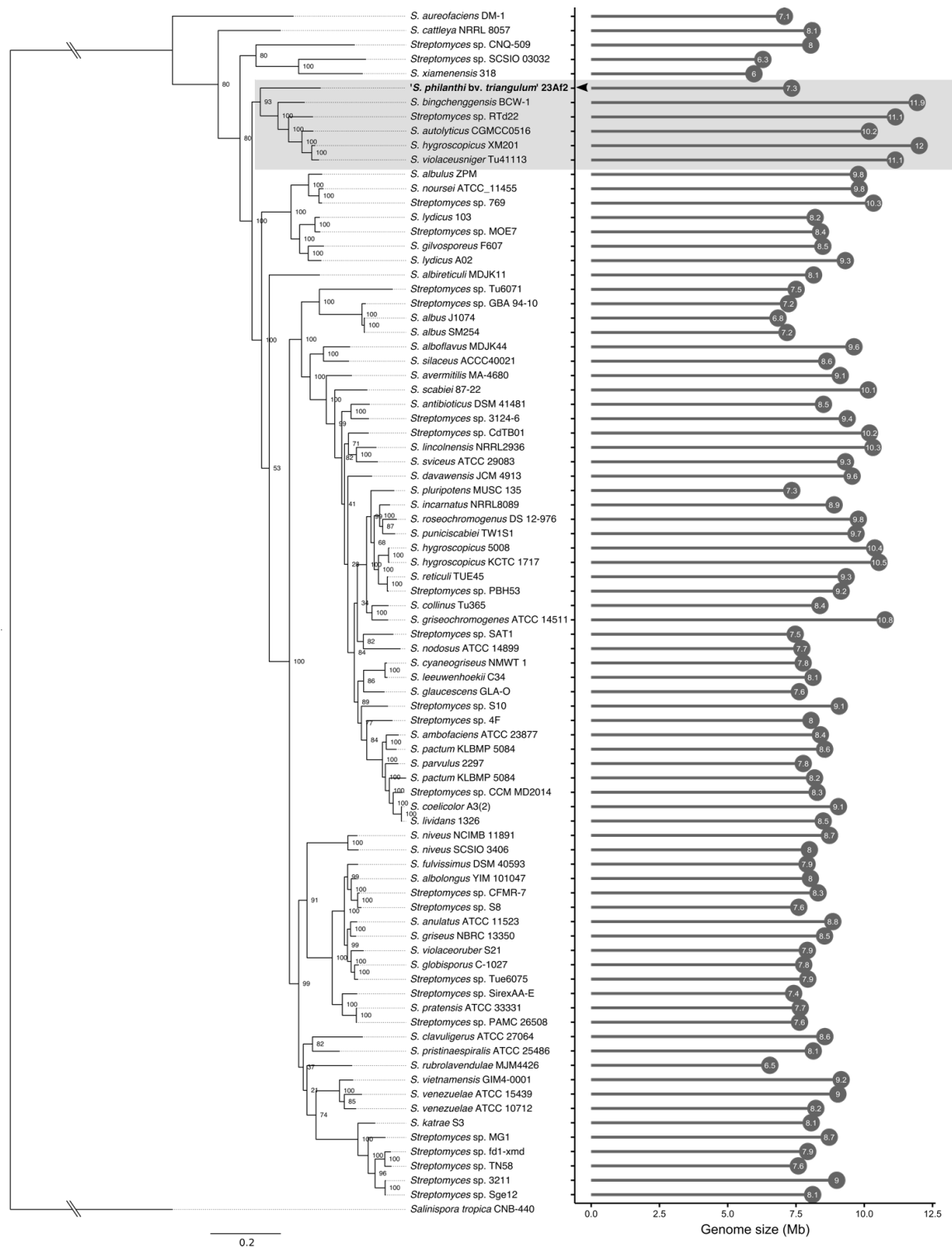
Figures S1 to S7

Table S1

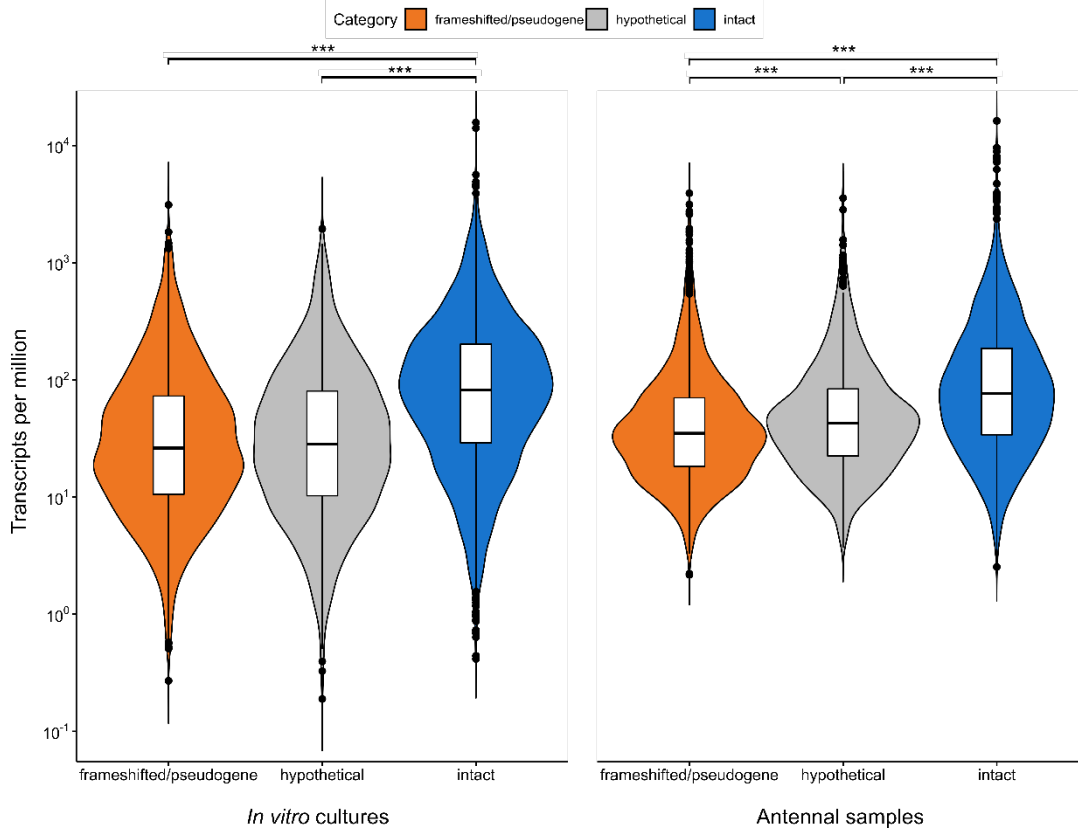
SI References



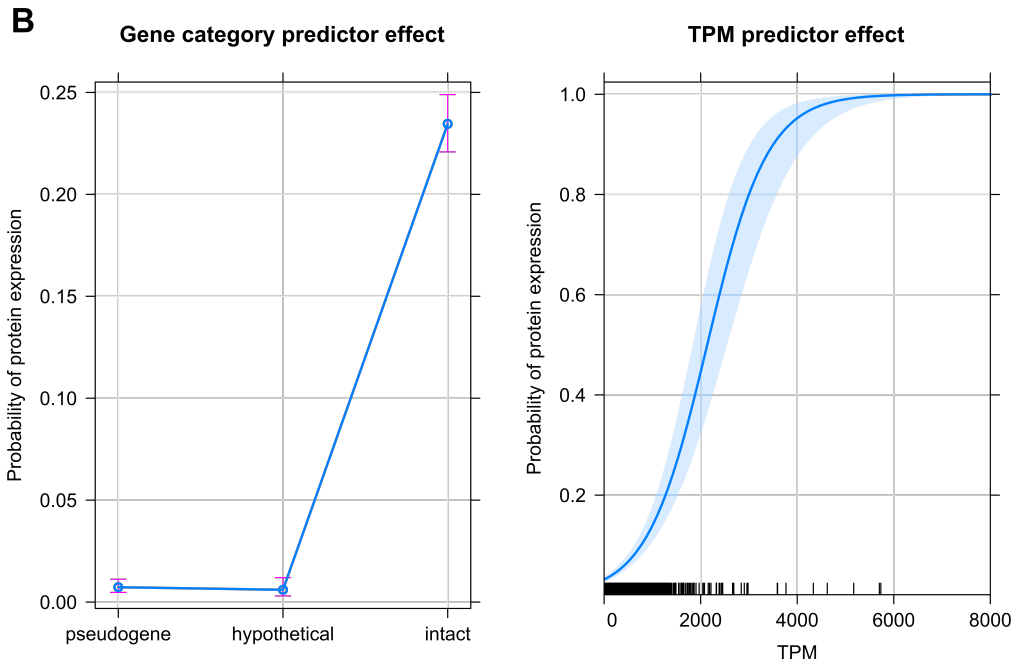
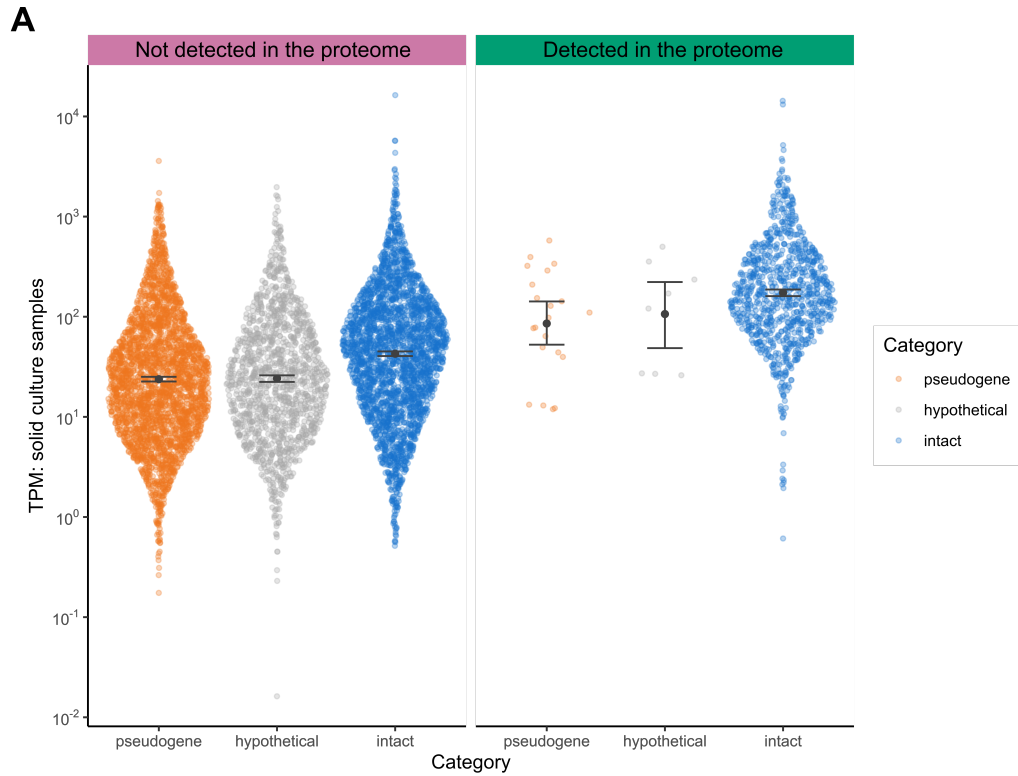
**Figure S1.** Synteny plots of “*S. philanthi*” bv. *triangulum* vs. *S. bingchenggensis* and *S. hygrosopicus* (**A** and **B**); and *S. bingchenggensis* vs. *S. hygrosopicus* (**C**). The dot-plots were performed with D-Genies (1). Position in the corresponding chromosome is indicated in the x and y axis. The “core” of the reference chromosome is indicated in dark blue, and the “arms” in light blue.



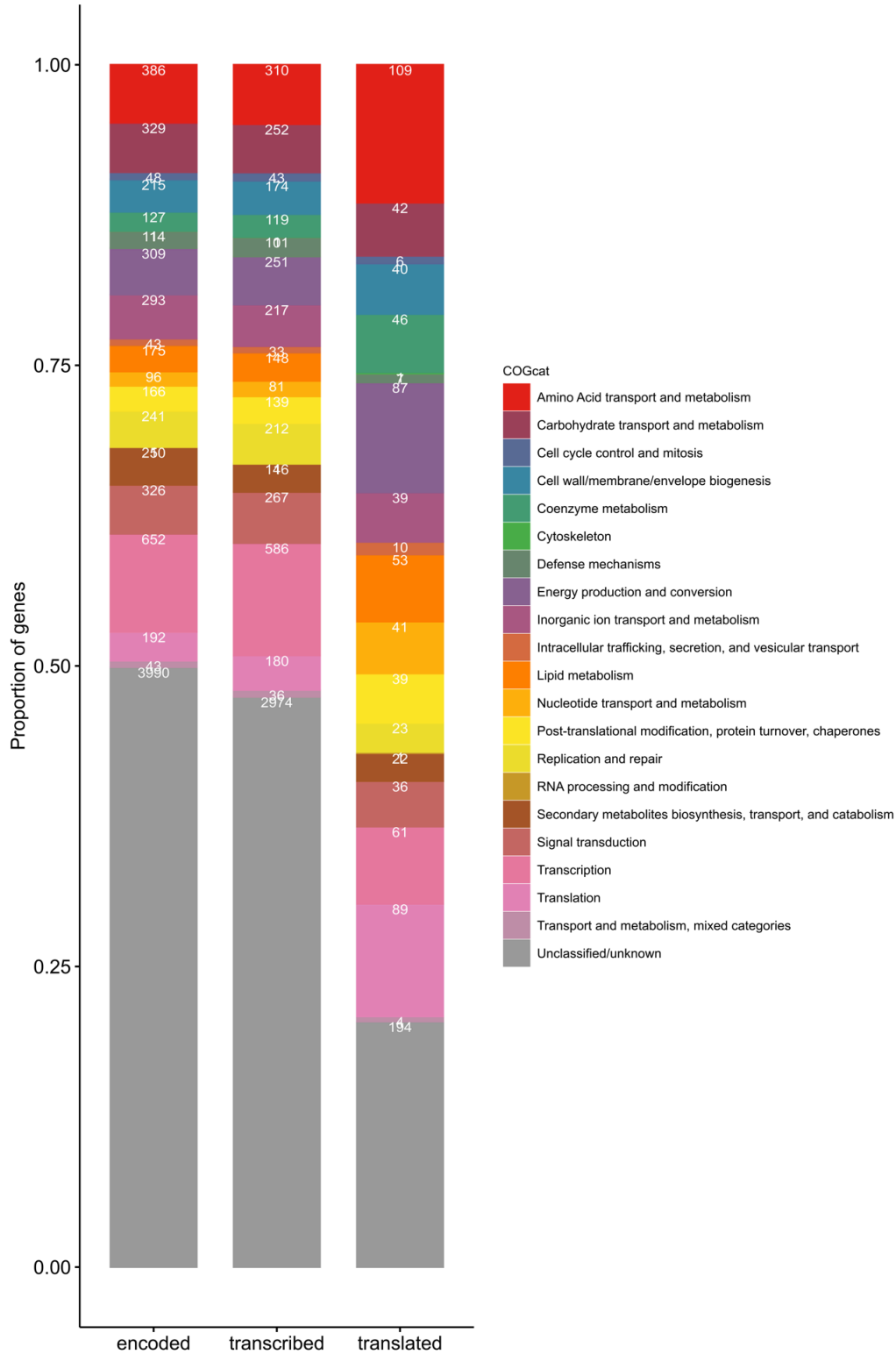
**Figure S2.** Phylogenomic tree of “*S. philanthi*” and other *Streptomyces* species with fully sequenced genomes. The tree was constructed using PhyML (2) from a concatenated set of 85 genes obtained with Phyla-AMPHORA (3). Bipartition support was obtained from 100 bootstrap pseudoreplicates. *Salinispora tropica* was used to root the tree. Length of the branches is proportional to the number of substitutions per site. Genome size of each species is indicated in the graph on the right side. The clade containing the beewolf symbiont (arrowhead) is highlighted with grey shading



**Figure S3.** Expression of intact genes, pseudogenes and hypothetical protein-coding genes in “*S. philanthi*” *in vitro* cultures and antennal samples, respectively. Frameshifted and hypothetical protein coding genes showed lower expression levels in comparison to intact genes. TPM values are shown for the RNAseq experiments of *in vitro* cultures and antennal samples. Dunn post hoc test for Kruskal-Wallis with multiple comparisons was performed, p-values adjusted with the Benjamini-Hochberg method. \*\*\* indicates comparisons where  $p < 10^{-5}$ .

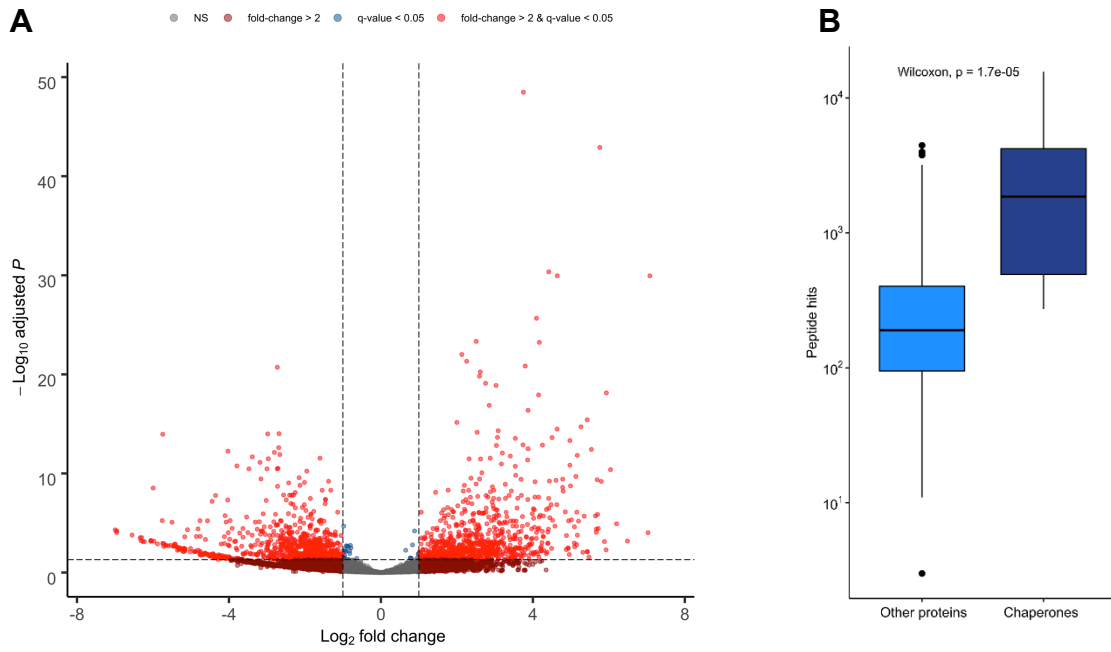


**Figure S4. A.** Sina plots showing the TPM of genes in the transcriptome, separated according to whether they were detected in the proteome or not. Using a generalized linear model we found that having an intact coding sequence is a strong predictor of whether their corresponding peptides were found in the proteomic analysis (binomial logistic regression,  $p < 2e-16$ ). Strength of transcription, as measured through gene TPM values is also a predictor of protein expression ( $p < 2e-16$ ). **B.** Predictor effect plots (4) on the probability of protein expression, showing the effect of gene intactness (left) and gene TPM (right).

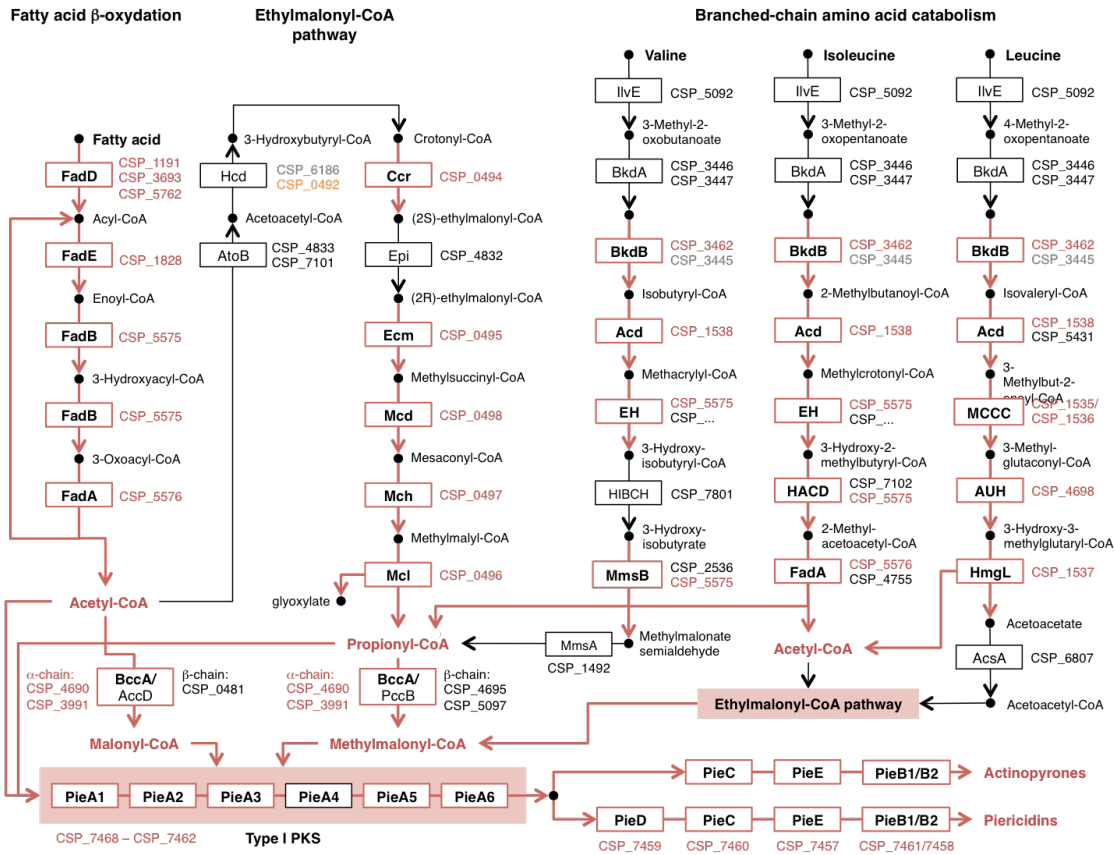


**Figure S5.** Classification of all protein coding genes in the genome of “*S. philanthi*” according to their COG categories, in comparison to the subset of genes which are transcribed or translated. The absolute number is indicated inside the bars.

*S. philanthi* 23Af2 *in vitro* cultures vs. antennal samples



**Figure S6. A.** Volcano plot showing the differentially expressed genes between antennal and *in vitro* “*S. philanthi*” samples. Genes with a fold change > 2 and a q-value < 0.05 are shown in light red. **B.** Peptide hits to molecular chaperones in comparison to the rest of the proteome from *in vitro* cultures of “*S. philanthi*”.



**Figure S7.** Metabolic network linking fatty acid and branched-chain amino acid catabolism with ptericidin/actinopyrone biosynthesis, showing in red the genes upregulated in the antennae. Genes with frameshifts are in orange and genes downregulated in antenna are in grey.



**Table S1.** Secondary metabolite gene clusters found in the genome of “*S. philanthi*” bv. *triangulum* 23Af2. These biosynthetic clusters were identified using antiSMASH v.5.0.0 (5).

Region	Type	From	To	Most similar known cluster	Similarity	MIBiG BGC-ID
Region 1	siderophore	351,102	363,597	-	-	-
Region 2	T1PKS	407,045	448,901	Azalomycin F	52%	BGC0001523
Region 3	bacteriocin	509,320	518,188	-	-	-
Region 4	lanthipeptide	847,667	870,630	Griselimycin	7%	BGC0001414
Region 5	T1PKS	1,643,642	1,687,932	Zincophorin	76%	BGC0001828
Region 6	terpene	1,708,094	1,727,585	Merochlorin	7%	BGC0001083
Region 7	NRPS-like	1,761,747	1,802,830	Echosides	76%	BGC0000340
Region 8	terpene	2,030,141	2,049,379	Xiamycin	13%	BGC0000666
Region 9	terpene	2,685,354	2,706,064	Geosmin	100%	BGC0001181
Region 10	arylpolyene	3,563,466	3,604,356	Skylamycin	10%	BGC0000429
Region 11	terpene	3,819,940	3,839,905	-	-	-
Region 12	other	4,420,192	4,472,315	A-503083	5%	BGC0000288
Region 13	siderophore	4,822,135	4,835,992	Ficellomycin	5%	BGC0001593
Region 14	bacteriocin	5,098,868	5,110,643	-	-	-
Region 15	T3PKS	5,125,171	5,166,259	Thiotetronate TŸ 3010	8%	BGC0001352
Region 16	NRPS	5,234,055	5,279,555	Diisonitrile antibiotic SF2768	66%	BGC0001574
Region 17	ectoine	5,603,927	5,614,331	Ectoine	100%	BGC0000853
Region 18	NRPS	6,118,594	6,186,851	Cahuitamycins	33%	BGC0001351
Region 19	NRPS	6,310,372	6,357,386	Griseobactin	41%	BGC0000368
Region 20	terpene	6,509,367	6,534,065	Hopene	61%	BGC0000663
Region 21	T1PKS	6,752,447	6,835,492	<b>Piericidin A1</b>	100%	BGC0000124
Region 22	terpene	6,828,880	6,853,613	Isorenieratene	71%	BGC0000664
Region 23	T1PKS	6,850,019	6,895,522	-	-	-

## SI References

1. Cabanettes F & Klopp C (2018) D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* 6:e4958.
2. Guindon S, *et al.* (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology* 59(3):307-321.
3. Wang Z & Wu M (2013) A phylum-level bacterial phylogenetic marker database. *Molecular biology and evolution* 30(6):1258-1262.
4. Fox J (2003) Effect Displays in R for Generalised Linear Models. *2003* 8(15):27.
5. Blin K, *et al.* (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* 47(W1):W81-W87.