Omics Technologies to Understand Activation of a Biosynthetic Gene Cluster in *Micromonospora* sp. WMMB235: Deciphering Keyicin Biosynthesis

Deepa Acharya,[a] Ian Miller,[a] Yusi Cui,[a] Doug R. Braun,[a] Mark E. Berres,[b] Matthew J. Styles,[c] Lingjun Li,[a] Jason Kwan,[a] Scott R. Rajski,[a] Helen E. Blackwell,[c] and Tim S. Bugni[a,*]

[a]Pharmaceutical Sciences Division, University of Wisconsin-Madison, Madison, Wisconsin 53705, USA

[b]Bioinformatics Resource Center, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

[c]Department of Chemistry, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

Table of Contents

Parameters for AntiSmash: The fasta nucleotide sequence was uploaded on AntiSmash v3.0 with ClusterFinder on, "Use ClusterFinder algorithm for BGC border prediction" on, Minimum cluster size in CDS 5, Minimum number of biosynthetic-related PFAM domains 5, Minimum ClusterFinder probability at 60% and using the "All on" for extra features which uses the following programs: KnownClusterBlast, ClusterBlast, SubClusterBlast, smCoG analysis, Active Site Finder, Detect TTA codons, Whole genome PFAM analysis.

Transcriptomics details:
Read filtering done by removing low abundance genes defined as those with an average read count below a threshold of 1.0 in two or more samples. Samples in RNASeq experiment were normalized by the method of trimmed mean of M-values (TMM), resulting in more uniform distributions centered on a common median.

Table S1: The total RNA concentrations yielded after the RNAeasy Plus Mini Kit extraction were all between 300-500 ng/ml concentration and had A_{260}/A_{280} values ~ 2 suggesting good quality RNA. The values of the secondary measure of RNA purity A_{260}/A_{230} are also around the expected ~1.8.

Entry	Sample name	A_{260}(10mM)	$A_{260/280}$	$A_{260/230}$	Concentration(ng/ml)
1	185D21	12.851	2.09	2.13	514
2	185D21	9.636	2.05	1.98	385
3	185D23	5.839	2.08	1.95	233
4	235D21	9.833	2.03	2.22	393
5	235D22	6.3	2.07	1.82	252
6	235D23	12.14	2.05	2.02	485
7	CoD21	10.005	2.02	2.2	400
8	CoD22	12.039	2.04	2.09	481
9	CoD23	11.127	1.97	2.2	445
10	185D51	8.344	2.06	2.11	333
11	185D52	10.453	1.94	1.98	418
12	185D53	11.641	2.03	1.99	465
13	235D51	12.334	2.05	2.2	493
14	235D52	10.832	2	2.15	433
15	235D53	12.221	1.99	2.05	489
16	CoD51	12.187	2.05	2.05	487
17	CoD52	12.631	2.04	2.13	505
18	CoD53	11.802	1.98	2.04	472

Table S2: DiLeu reporter tags used to label each sample in the multiplex for Day 5 and Day 8 protein samples before analysis on LC-MS/MS for quantitative proteomics.

DiLeu Tag	Day 5 or Day 8 Sample	Reporter ion
115a	WMMB235 Mono rep1	115.12476m/z
116a	WMMB235 with WMMA185 extract rep1	116.12812m/z
117a	WMMB235 Mono rep2	117.13147m/z
118a	WMMB235 with WMMA185 extract rep2	118.13483m/z
117c	WMMB235 Mono rep3	117.14363m/z
118c	WMMB235 with WMMA185 extract rep3	118.14699m/z

Table S3: Comprehensive summary of putative functions for all *kyc orfs* based on NCBI database. *kyc orf* products identified by proteomics indicated with green shading; repressive regulatory *orfs* identified in transcriptomics highlighted with yellow.

ID	NCBI_Prot_Accession	*kyc orf*	Protein ID	putative function based on NCBI database
BFV98_00170	OHX01510.1	**kyc4**	WMMB235_BFV98_00170	deaminase
BFV98_00175	0	kyc5	WMMB235_BFV98_00175	DNA-binding response regulator
BFV98_00180	OHX01511.1	kyc6	WMMB235_BFV98_00180	methionine adenosyltransferase
BFV98_00185	OHX07111.1	**kyc7**	WMMB235_BFV98_00185	dTDP-glucose 4,6-dehydratase
BFV98_00190	OHX01512.1	kyc8	WMMB235_BFV98_00190	glucose-1-phosphate thymidylyltransferase
BFV98_00195	OHX01513.1	kyc9	WMMB235_BFV98_00195	cyclase
BFV98_00200	OHX07112.1	kyc10	WMMB235_BFV98_00200	hydroxylacyl-CoA dehydrogenase
BFV98_00205	OHX01514.1	kyc11	WMMB235_BFV98_00205	short-chain dehydrogenase
BFV98_00210	OHX01515.1	**kyc12**	WMMB235_BFV98_00210	glycosyl transferase family 28
BFV98_00215	OHX01516.1	kyc13	WMMB235_BFV98_00215	beta-ACP synthase
BFV98_00220	OHX01517.1	**kyc14**	WMMB235_BFV98_00220	beta-ketoacyl synthase
BFV98_00225	OHX01518.1	kyc15	WMMB235_BFV98_00225	actinorhodin polyketide synthase
BFV98_00230	OHX01519.1	**kyc16**	WMMB235_BFV98_00230	epimerase
BFV98_00235	OHX01520.1	kyc17	WMMB235_BFV98_00235	phytanoyl-CoA dioxygenase
BFV98_00240	OHX01521.1	kyc18	WMMB235_BFV98_00240	diguanylate cyclase
BFV98_00245	OHX01522.1	kyc19	WMMB235_BFV98_00245	SAM-dependent methyltransferase
BFV98_00250	OHX01523.1	**kyc20**	WMMB235_BFV98_00250	glycosyl transferase family 28
BFV98_00255	OHX01524.1	kyc21	WMMB235_BFV98_00255	cyclase
BFV98_00260	OHX01525.1	kyc22	WMMB235_BFV98_00260	ketoacyl reductase
BFV98_00265	OHX07113.1	kyc24	WMMB235_BFV98_00265	glycosyl transferase family 28
BFV98_00270	OHX01526.1	**kyc25**	WMMB235_BFV98_00270	glycosyl transferase family 28
BFV98_00275	OHX01527.1	**kyc26**	WMMB235_BFV98_00275	cytochrome
BFV98_00280	OHX01528.1	kyc27	WMMB235_BFV98_00280	NAD-dependent epimerase
BFV98_00285	OHX01529.1	**kyc28**	WMMB235_BFV98_00285	dTDP-4-dehydrorhamnose 3,5-epimerase
BFV98_00290	OHX01530.1	kyc29	WMMB235_BFV98_00290	glycosyl transferase family 28
BFV98_00295	OHX01531.1	kyc30	WMMB235_BFV98_00295	TetR family transcriptional regulator
BFV98_00300	OHX01532.1	kyc31	WMMB235_BFV98_00300	NmrA family transcriptional regulator
BFV98_00305	OHX01533.1	**kyc32**	WMMB235_BFV98_00305	glycosyl transferase family 28
BFV98_00310	OHX01534.1	kyc33	WMMB235_BFV98_00310	ketoreductase
BFV98_00315	OHX01535.1	**kyc34**	WMMB235_BFV98_00315	ester cyclase
BFV98_00320	OHX07114.1	kyc35	WMMB235_BFV98_00320	methyltransferase
BFV98_00325	OHX01536.1	kyc36	WMMB235_BFV98_00325	NmrA family transcriptional regulator
BFV98_00330	OHX01537.1	kyc37	WMMB235_BFV98_00330	FAD-linked oxidase
BFV98_00335	OHX01538.1	kyc38	WMMB235_BFV98_00335	antibiotic biosynthesis monooxygenase
BFV98_00340	OHX01539.1	kyc39	WMMB235_BFV98_00340	multidrug ABC transporter permease
BFV98_00345	OHX01540.1	kyc40	WMMB235_BFV98_00345	ABC transporter
BFV98_00350	OHX07115.1	kyc41	WMMB235_BFV98_00350	DNA-binding response regulator
BFV98_00355	OHX01541.1	kyc42	WMMB235_BFV98_00355	transcriptional regulator
BFV98_00360	OHX01542.1	kyc43	WMMB235_BFV98_00360	oxidoreductase
BFV98_00365	OHX01543.1	kyc44	WMMB235_BFV98_00365	SAM-dependent methyltransferase
BFV98_00370	OHX01544.1	**kyc45**	WMMB235_BFV98_00370	daunorubicin biosynthesis sensory transduction protein DnrJ
BFV98_00375	OHX01545.1	kyc46	WMMB235_BFV98_00375	NDP-hexose 2,3-dehydratase
BFV98_00380	OHX01546.1	kyc48	WMMB235_BFV98_00380	hypothetical protein
BFV98_00385	OHX01547.1	kyc49	WMMB235_BFV98_00385	NADPH-dependent FMN reductase
BFV98_00390	OHX01548.1	kyc50	WMMB235_BFV98_00390	SAM-dependent methyltransferase
BFV98_00395	OHX01549.1	kyc51	WMMB235_BFV98_00395	DUF4440 domain-containing protein
BFV98_00400	OHX01550.1	kyc52	WMMB235_BFV98_00400	hypothetical protein
BFV98_00405	OHX01551.1	kyc53	WMMB235_BFV98_00405	acyl-CoA dehydrogenase
BFV98_00410	OHX01552.1	kyc54	WMMB235_BFV98_00410	transcriptional regulator

Table S4: WMMB235-embedded BGC2 genes downregulated upon coculturing with *Rhodococcus* sp. WMMA185.

ID	NCBI accession number	Log_2FC Day 5 (Co/Mono)	NCBI Annotation (putative function of the gene product)	Homologue	MiBIG cluster
BFV98_00820	OHX01627.1	-1.48600927	hypothetical protein	AAN85518.1	BGC0001101
BFV98_00815	OHX01626.1	-1.41694817	enoyl-CoA hydratase	AAN85519.1	BGC0001101
BFV98_00940	OHX01649.1	-1.24897815	ABC transporter	AAN85531.1	BGC0001101
BFV98_00860	OHX01634.1	-1.20094237	ABC transporter	AAP85349.1	BGC0000233
BFV98_00865	OHX01635.1	-1.18908439	ABC transporter	AAG32068.1	BGC0000026
BFV98_00925	OHX01646.1	-1.1046561	ABC transporter substrate-binding protein	AAN85534.1	BGC0001101
BFV98_00915	OHX01644.1	-1.07446882	polyketide β-ketoacyl:ACP synthase	ADI59526.1	BGC0001091
BFV98_00910	OHX01643.1	-1.07396388	phosphopantetheine-binding protein	AAW33973.1	BGC0000255

Table S5: WMMB235-embedded BGC9 genes downregulated upon coculturing with *Rhodococcus* sp. WMMA185.

ID	NCBI accession number	Log_2FC Day 5 (Co/Mono)	NCBI Annotation (putative function of the gene product)	Homologue	MiBIG Cluster
BFV98_26800	OHX06338.1	-2.14562736	acyl carrier protein (ACP)	AHA12082.1	BGC0001172
BFV98_26795	OHX06337.1	-1.92477769	β-ketoacyl-ACP synthase II	CUI25675.1	BGC0001353
BFV98_26855	OHX06349.1	-1.89956036	pyruvate dehydrogenase (acetyl-transferring)-homodimeric type	A0R0B0.1	-
BFV98_26790	OHX06336.1	-1.70957738	hypothetical protein	A3Q339.1	-
BFV98_26830	OHX06344.1	-1.44829961	copper resistance protein CopC	Q56797.1	-
BFV98_26805	OHX06339.1	-1.27058731	3-oxoacyl-ACP synthase	CBW54661.1	BGC0000971
BFV98_26835	OHX06345.1	-1.24636114	tRNA-synthetase class I	CAG14955.2	BGC0000253
BFV98_26810	OHX06340.1	-1.10536628	biotin attachment protein	AJO72737.1	BGC0001381
BFV98_26775	OHX06333.1	-1.07607101	hypothetical protein	-	-

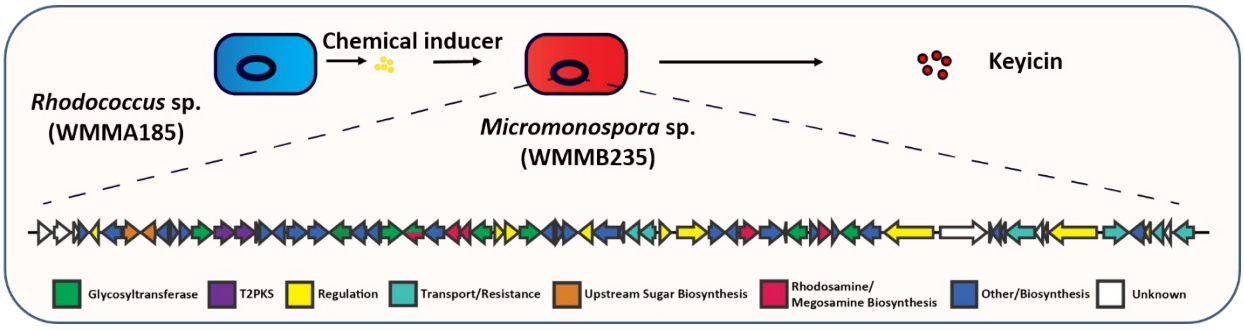


Figure S1. Schematic of coculture interactions and subsequent activation of silent BGC *kyc* cluster, by a *Rhodococcus*-derived small molecule.

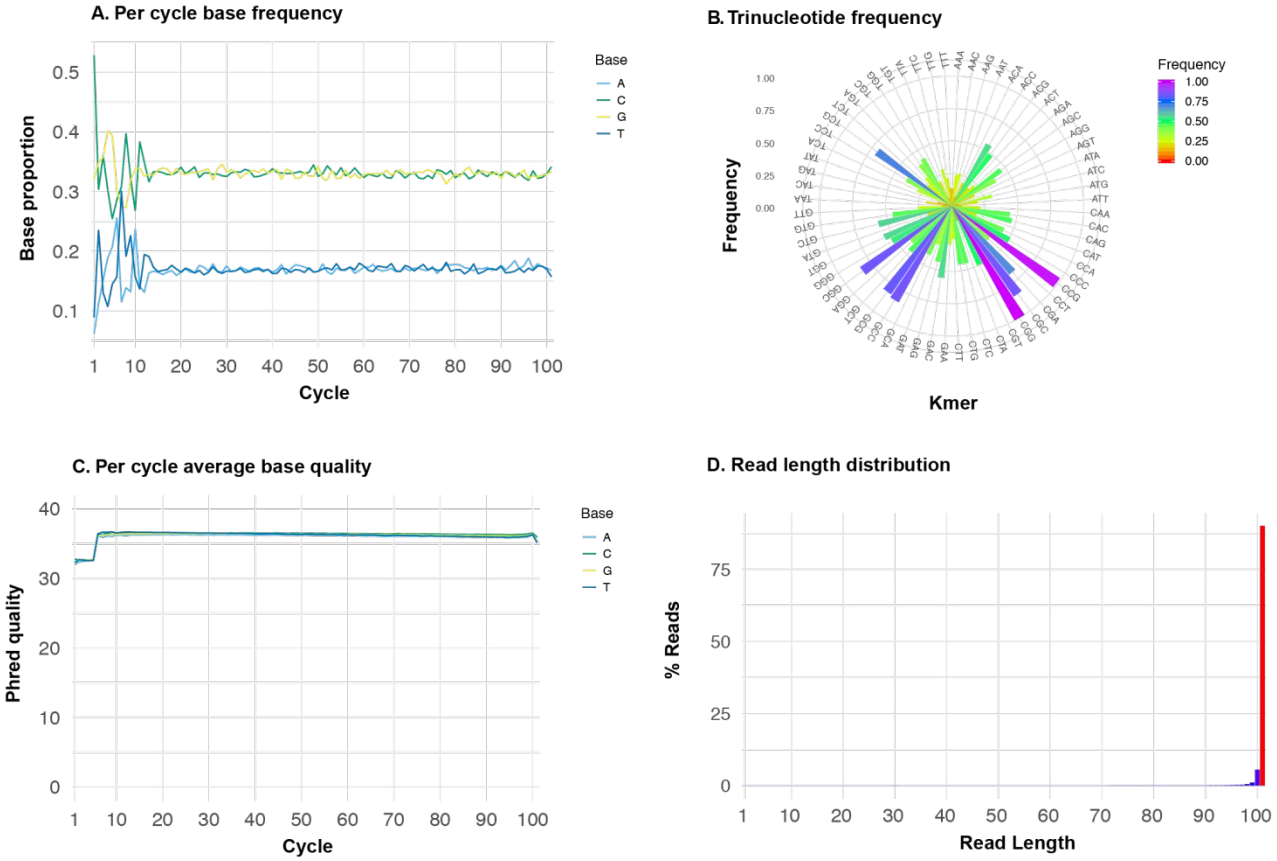


Figure S2. Sample quality control report of the RNA-Seq reads for sample – coculture day 2, Replicate 2. Similar results were obtained for all the nine samples, and suggested high quality reads. **A**) Per cycle base frequency. The frequency of each nucleotide base position across the length of the read shows the absence of any sequence bias. The effect on the base composition in the early sequences (cycles 1–10) is known and does not affect the base content of the read. **B**) Trinucleotide frequency. The relative frequency of trinucleotides (3-mer) shows relatively uniform representation. If one or more 3-mers dominate the remaining 3-mers, it may indicate the presence of a contaminating sequence, which is not the case here. **C**) Per cycle average base quality. The Phred scaled quality scores are > 30 which is equivalent to 99.9% base call accuracy. **D**) Read length distribution. Expected value of 100 bp (target read length) observed for majority of reads during RNASeq experiment.

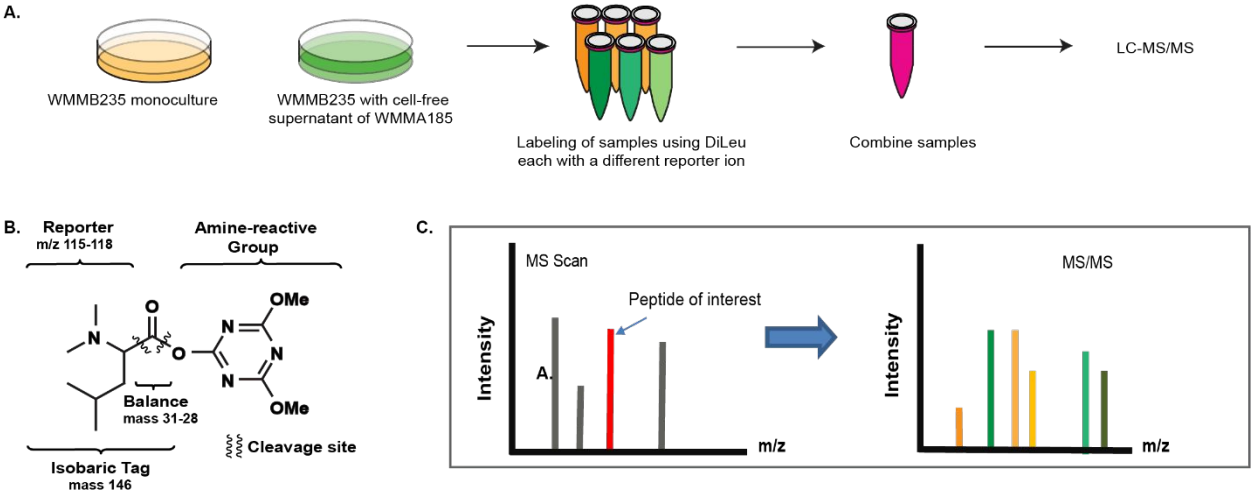


Figure S3. Schematic for labeled proteomics using DiLeu. **A**) Triplicates of WMMB235 monoculture and WMMB235 cultured in the supernatant of WMMA185 were labeled individually and combined according day of sampling (day 5 or day 8). **B**) DiLeu reporter showing the range in reporter ion mass and the amine reactive group, allowing the same mass of the modified peptide in MS1 scan but different MS2 ions after fragmentation at the cleavage site. **C**) Schematic showing a common peptide of interest in the MS1 scan fragmenting to 6 different reporter ions, each from a different sample on a particular day. This allows for quantification in the peptide present in each sample.

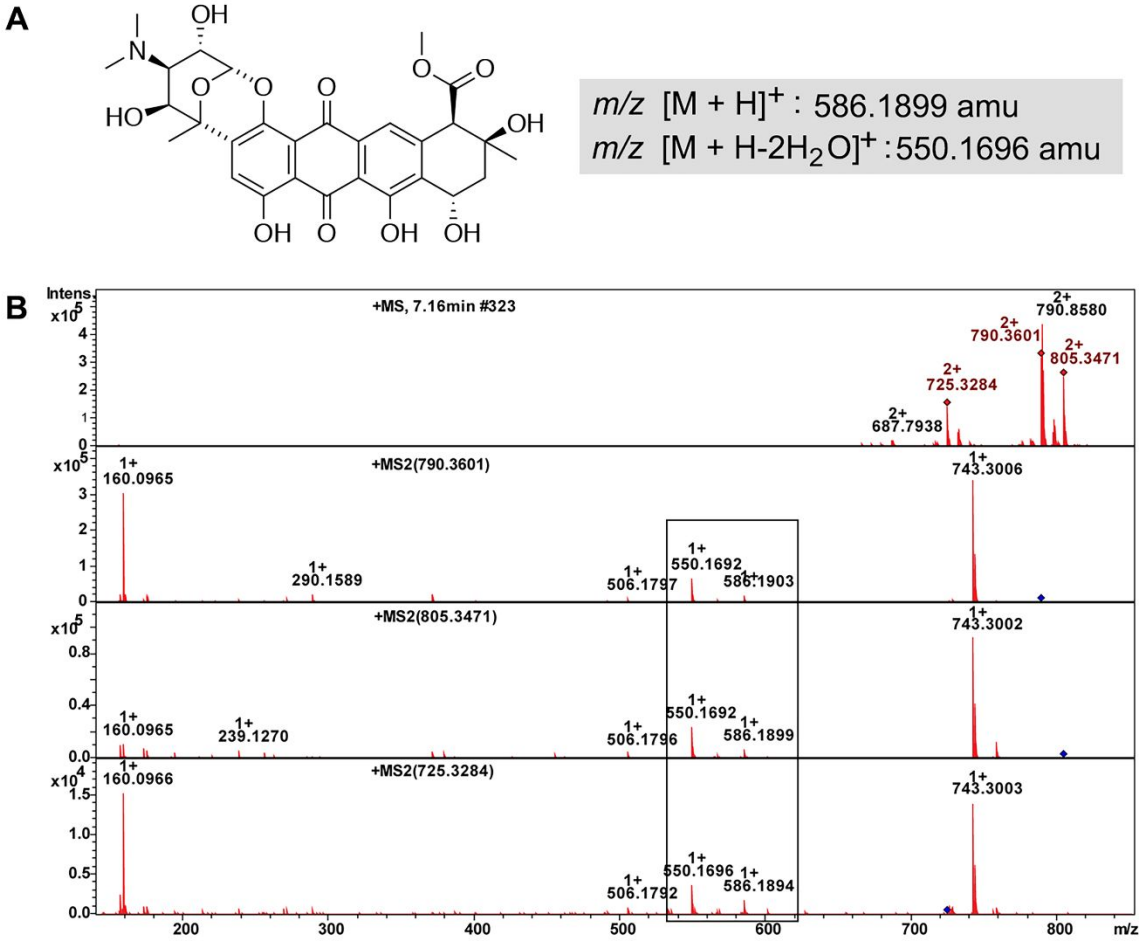


Figure S4. Structural analysis of keyicin-like compounds. **A**) Aglycone core of keyicin. **B**) Analogs of keyicin all share the anthracycline core as reflected by *m/z* signals at 550.1696, as well as *m/z* = 790.3601, 805.3471 (keyicin) and 725.3284 amu.

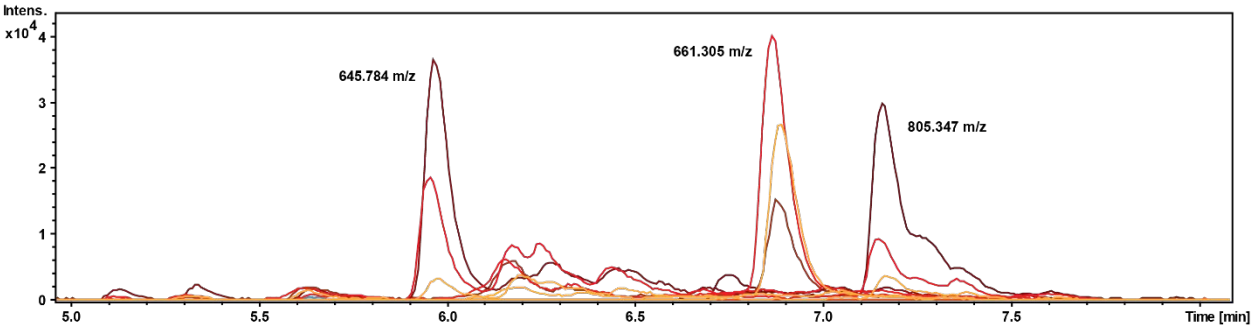


Figure S5. Keyicin and two of its analogs with temporal changes in intensity over time. Yellow: day 2, orange: day 5, Red: day 8 and Maroon: day 14. Keyicin (*m/z* = 805.347) increases in intensity over time; *m/z* = 645.784 follows a similar pattern, suggesting it is an analog; *m/z* = 661.305 first increases in intensity, but then is reduced by day 14, suggesting its intermediacy *en route* to keyicin or related analogs.

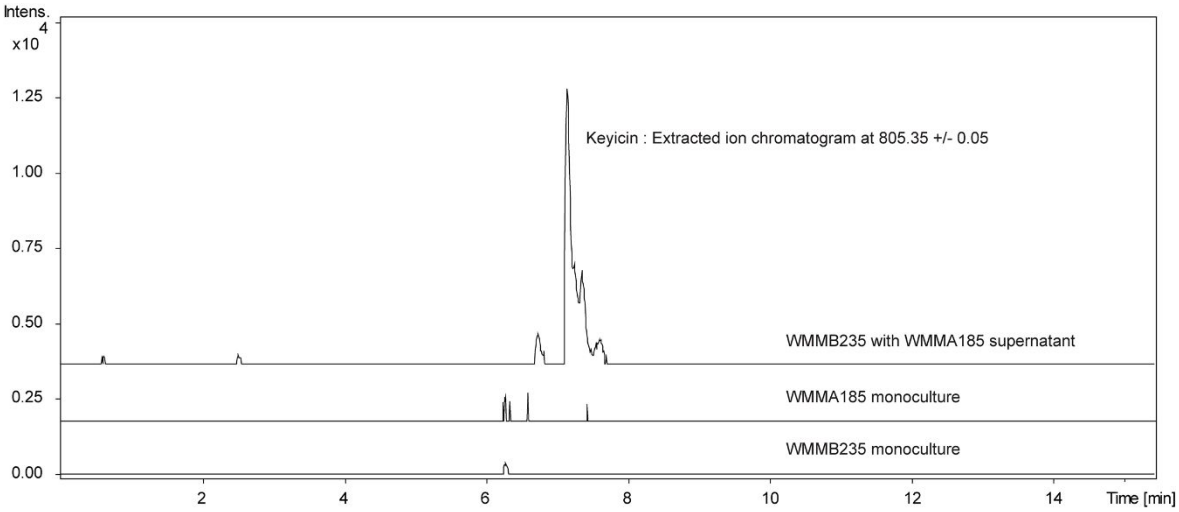


Figure S6. UPLC-MS extracted ion chromatogram (EIC) for *m/z* at 805.35 +/- 0.05, representative of keyicin. The EIC of monocultures, WMMA185 and WMMB235 show the absence of keyicin, whereas WMMB235 in the presence of WMMA185 supernatant shows keyicin production, indicating that the coculture interaction can be recapitulated by small molecule induction from WMMA185.

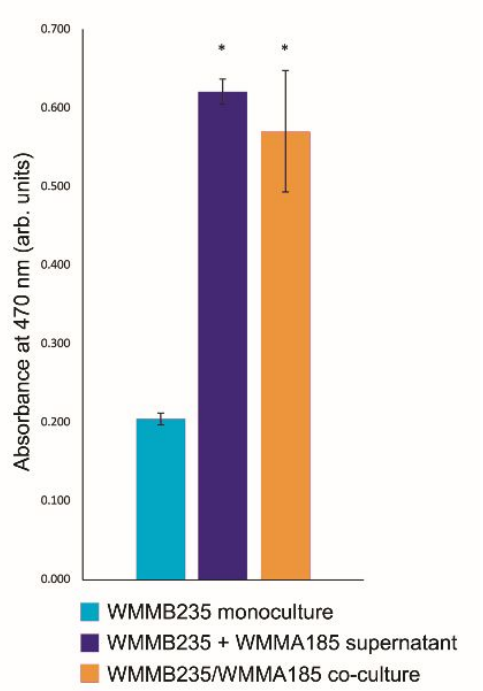


Figure S7. Comparisons of absorbance at 470 nm for WMMB235 monoculture versus WMMB235 + WMMA185 supernatant versus WMMB235 + WMMA185 coculture. Notably, absorbance at 470 nm is recorded in WMMB235 monoculture likely due to media components and cellular contents; however, this absorbance constitutes a baseline value used to generate fold-change (FC) values used to generate Figure 2 and other graphics in the manuscript. N=3, statistical significance was estimated for supernatant treated monoculture and the coculture by comparison with the monoculture absorbance; * indicates $p<0.05$

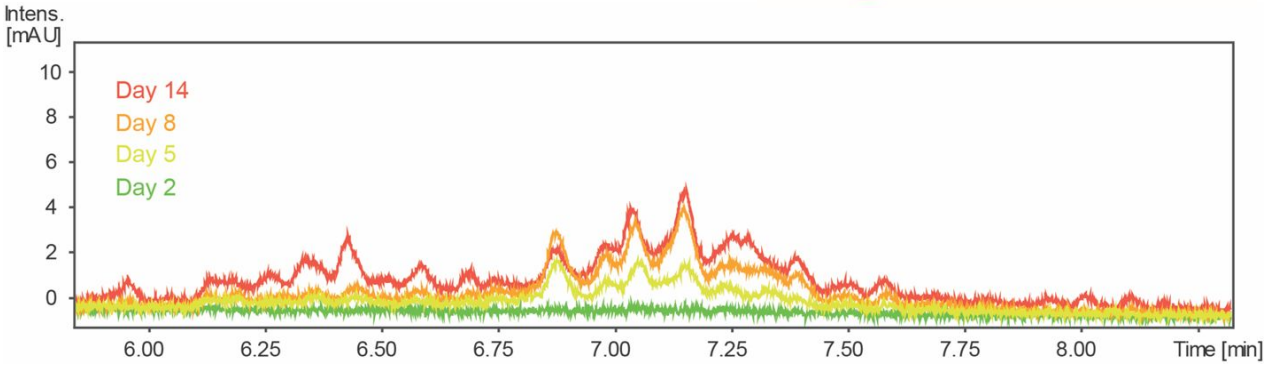
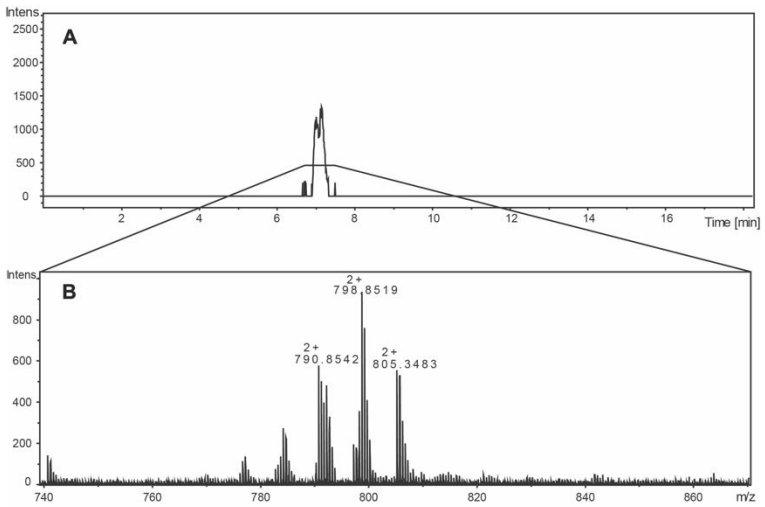


Figure S8. Extracted UV chromatogram at 470nm of the co-culture over time. Peaks corresponding to keyicin and its analogs can be visualized using LCMS-PDA, which shows these compounds increasing in intensity from day 2 to day 14. Each of these peaks were analyzed using MS/MS and confirmed to be analogs of keyicin from the fragmentation pattern, examples of which are shown in Figure S4.

Figure S9. LCMS chromatogram for AHL (odDHL) treated WMMB235. **A**) EIC for *m/z* 805.34 +/- 0.05 shows the LC peak corresponding to keyicin. **B**) The mass spectrum of the keyicin peak shows the doubly charged MS peak for keyicin, showing that odDHL can trigger keyicin production by WMMB235, albeit at lower intensity.



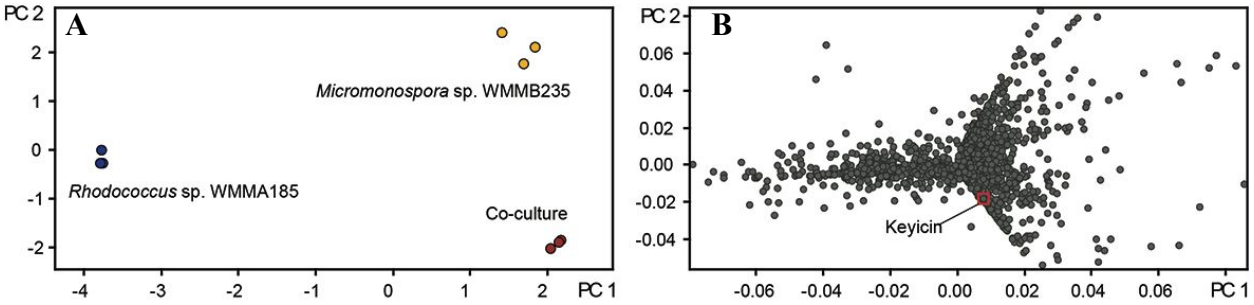


Figure S10. Principal component analysis (PCA) of LCMS data of monoculture WMMB235, WMMA185 and their coculture extracts from samples that were subsequently used to isolate RNA for transcriptomics. **A**) Scores plot showing the clustering of the replicates of monocultures and cocultures and spatial separation between groups which indicates significant differences in the metabolites produced in coculture. **B**) Loadings plot showing *m/z* ions representative of compounds produced in each condition. Keyicin ion is only found in the co-culture.