Supplementary Information

Generation of new glycoanalogues of polyene antibiotics by synthetic biology - testing current technical boundaries

Mark Hogan, Yuhao Song, Jimmy Muldoon and Patrick Caffrey*

School of Biomolecular and Biomedical Science, School of Chemistry and Centre for Synthesis and Chemical Biology, University College Dublin, Belfield, Dublin 4, Ireland.

*Correspondence: patrick.caffrey@ucd.ie

pIAGO-pegA1

```
            ermEp2 promoter
GCGGTCGATCTTGACGGCTGGCGAGAGGTGCGGGGAGGATCTGACCGACGCGGTCCACACGT


                                Hybrid BamHI-BglII site
GGCACCGCGATGCTGTTGTGGGCACAATCGTGCCGGTTGGTAGGATCTGCTAAGGAAATTCG


                M   R   V   S   L   Q   G   T   G   K   R   P   S
GCACCGATCGGTCGCAGGATCGGTGCGGGTGTCCTTGCAGGGCACCGGCAAACGCCCGAGCG


  V   G   D   E   M   T   S   H   R   P   I   L   F   C   C
TAGGAGACGAGATGACATCGCATCGTCCCATCCTGTTCTGTTGCAC
```

pIAGO-pegA2

```
            ermEp2 promoter
GCGGTCGATCTTGACGGCTGGCGAGAGGTGCGGGGAGGATCTGACCGACGCGGTCCACACGT


                                Hybrid BamHI-BglII site
GGCACCGCGATGCTGTTGTGGGCACAATCGTGCCGGTTGGTAGGATCTTGCAGGGCACCGGC


                  M   T   S   H   R   P   I   L   F   C   C
AAACGCCCGAGCGTAGGAGACGAGATGACATCGCATCGTCCCATCCTGTTCTGTTGCAC
```

Fig. S1 Difference between pIAGO-pegA1 and pIAGO-pegA2. pIAGO-pegA1 has 36 nucleotides upstream from the first possible start codon. pIAGO-pegA2 has 38 nucleotides upstream from the second methionine codon.
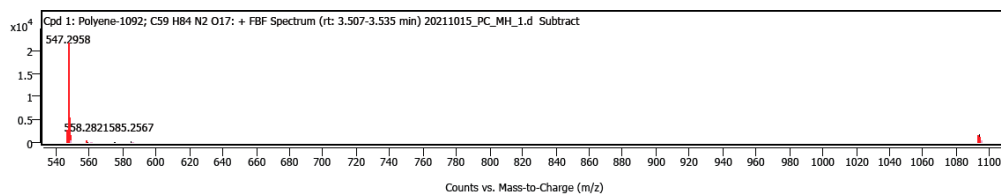
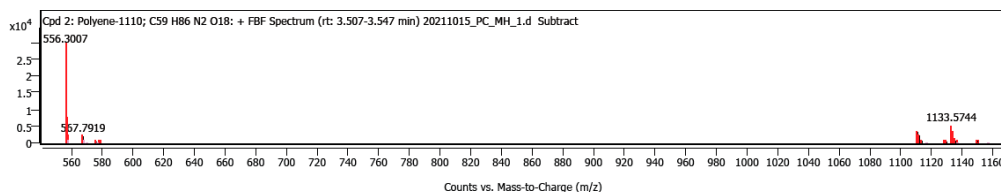Fig. S2 Detection of candicidin A1, M = 1092.5770, $(M + 2H)^{2+}/2$ = 547.2958

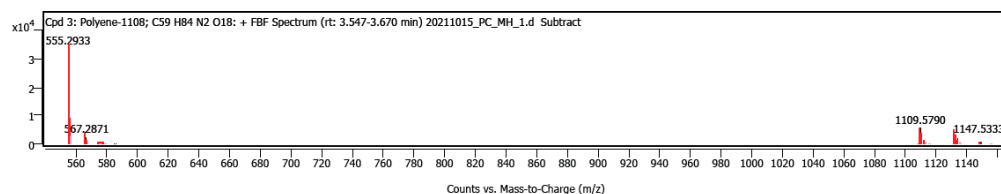Fig. S3 Detection of candicidin A3, M = 1110.5876, $(M + 2H)^{2+}/2$ = 556.3007; $(M + Na)^+$ = 1133.5744

Fig. S4 Detection of candicidin D, M = 1108.5719, $(M + 2H)^{2+}/2$ = 555.2933; $(M + H)^+$ = 1109.5790; $(M + K)^+$ = 1147.5333.
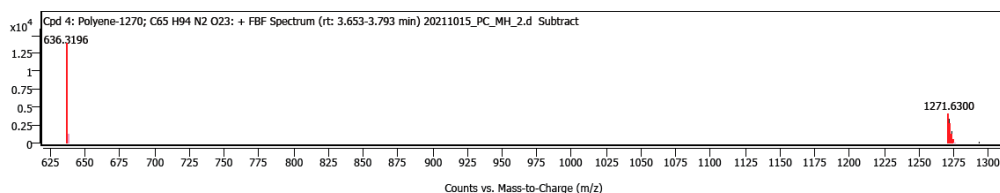

Fig. S5 Detection of mannosyl candicidin D, M = 1270.6247, $(M + 2H)^{2+}/2$ = 636.3196, $(M + H)^+$ = 1271.6300.
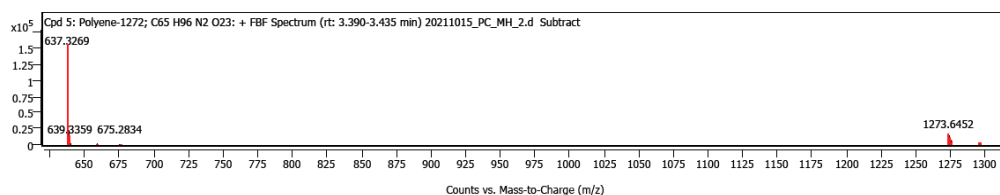

Fig. S6 Detection of mannosyl candicidin A3, M = 1272.6404, $(M + 2H)^{2+}/2$ = 637.3269, $(M + H)^+$ = 1273.6452.
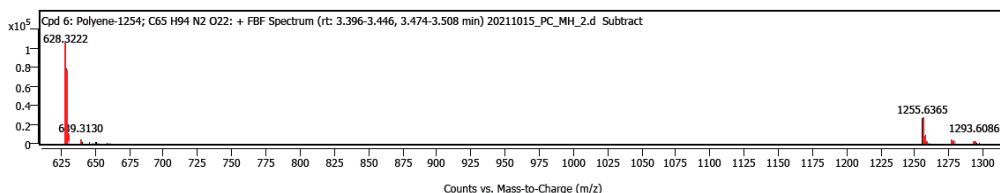
Fig. S7 Detection of mannosyl candicidin A1, M = 1254.6298, $(M + 2H)^{2+}/2$ = 628.3222, $(M + H)^+$ = 1255.6365, $(M + K)^+$ = 1293.6086
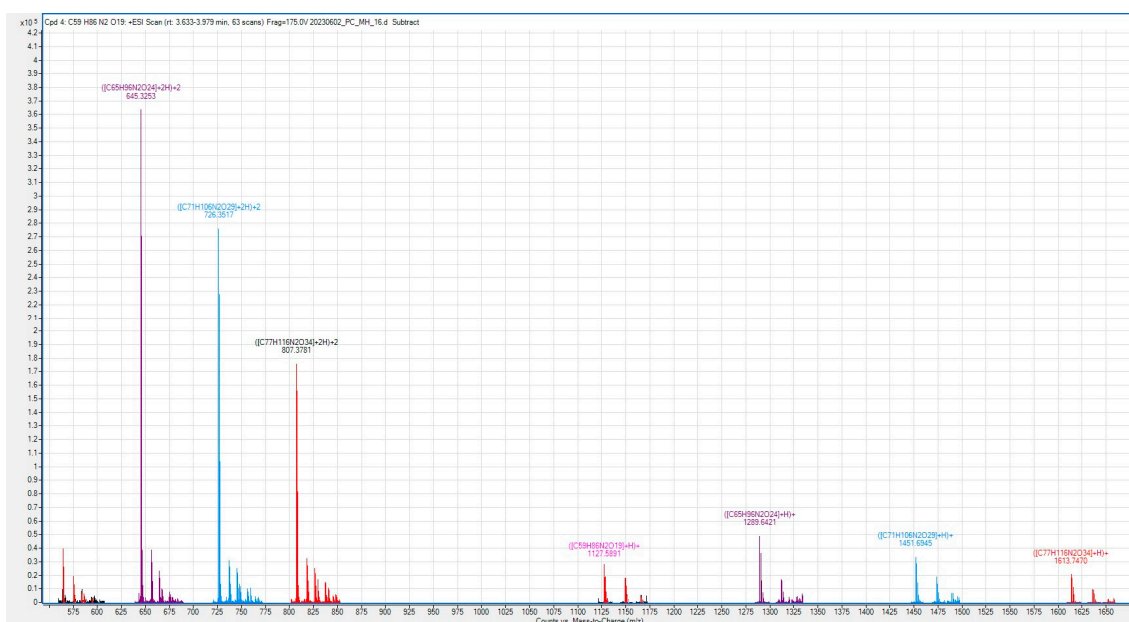
Fig. S8. Detection of tetrasaccharide-, trisaccharide-, disaccharide- and monosaccharide-containing aromatic heptaenes after modification of the 67-121C/67-121A complex with lactose. Molecular ions were identified as follows: lactose-modified 67-121C (tetrasaccharide), $[M + H]^+$ = 1613.7470, $[M + 2H]^{2+}/2$ = 807.3781; lactose-modified 67-121A (trisaccharide), $[M + H]^+$ = 1451.6945, $[M + 2H]^{2+}/2$ = 726.3517; unmodified 67-121C (disaccharide), $[M + H]^+$ = 1289.6421, $[M + 2H]^{2+}/2$ = 645.3253; unmodified 67-121A (monosaccharide), $[M + H]^+$ = 1127.5891.

```
PenSV    MRILFSVSSWTGHYYAMFPLGWALRAAGHEVRVLCRPGDRDDVVRAGLVPVPVLDGPDML    60
SelSV    MRVLFAISSWTGHYFPMVPLAWAMRAAGHDVRVLCRPSDQADVTAAGLIPVPALDGLDLL    60
         **:**:*:*******: *.**.**:*****:*******.*: **. ***:***.*** *:*

PenSV    TGARLLNVLSSFSGTWPYPTPPPHPDSGEPVDD-AFDLQSWLAGYWARATESSRAGTDAA    119
SelSV    RGARLLNVMSLLQGTWPYPEPPPHPDTGEAMDPAGFDIAAWHAENMPAMVASSRAGTDAA    120
          *******:* :.****** *******:** :*  .**: :* *      . *********

PenSV    VRYARDWQPDLVVHDQLSLDGPLVAAVLGVPQVMHLWGPVGPADSFGPVGGEEAGRPADP    179
SelSV    VAFGRSWAPDLVVHDQLSLEGPLVSAVTGAPSVLHLWGPAGTADAFAPVGGEQAGLPQDL    180
         * :.*.* ***********:****:** *.*.*:*****.* **:*.*****:** * *

PenSV    SDAFGRHGAGEFDHGRADLILDPCPGPLAAGLGKRVLPIRYVPYNGPGDDPGPGALPAPT    239
SelSV    SDAFTRYGAGTLSHDLADHVLDPCPPPLRTAVAGRDAGIRYVPYNGPGAAPL--DLPEPD    238
         **** *:*** ::.*. ** :***** ** :.:. *    ********** *    ** *

PenSV    GR-RRVCVVWGRSATRTFGPVANKLPEVVRAATDHGADVLLLARPDDAAAAGELPENVTA    298
SelSV    GRRPRVCVIWGRSVTRTFGPVVNRLPQAVRAAADLGAEVLLLARPEDARDAGPLPDGVRP    298
         **   ****:****.*******.*:**:.****:* **:********:**   ** **:.*

PenSV    MCEVPMSLVLDGCDAVVHYGGAGSAMTALVAGLPQLSVPIGLDQDLVAARLAATGAACTV    358
SelSV    FHEVPLSLVLPGCEAVVHYAGAGSVMTALTAGVPQLSVPCGFDQPMVAERLSATGAGLHV    358
         : ***:**** **:*****.****.****:**:****** *:** :** **:****. *

PenSV    PGPIADVESIGNALAAVLEGPQHAEAAEKLAAAVAAMPPPAAVVADLESLVASRARAA    416
SelSV    HNLDADAATLGGALEKLIGQPSYADAARDLARRCAAMPSPAEVVADLEALAAR-----    411
         .  **. ::*.** ::  *.:*:**..**    **** ** ******:*.*
```

Fig. S9 Alignment of PenSV and SelSV glycosyltransferase sequences.

SelSV synthetic gene sequence

```
>SelSV
CATATGCGTGTGCTCTTCGCCATCTCCTCGTGGACCGGCCACTACTTCCCGATGGTCCCG
CTGGCCTGGGCGATGCGGGCCGCCGGACACGACGTCCGCGTGCTCTGCCGTCCCTCCGAC
CAGGCCGACGTGACCGCGGCAGGACTGATCCCGGTGCCCGCGCTGGACGGCCTCGACCTG
CTGCGCGGCGCGCGGCTGCTCAACGTGATGAGCCTGCTGCAGGGGACGTGGCCCTACCCA
GAGCCGCCGCCCCACCCCGACACCGGGGAGGCGATGGATCCGGCCGGGTTCGACATCGCG
GCGTGGCACGCGGAGAACATGCCCGCCATGGTGGCGTCGTCGCGGGCCGGTACCGACGCC
GCCGTCGCGTTCGGCCGGTCCTGGGCCCCGGACCTGGTCGTGCACGACCAGCTGAGCCTG
GAGGGCCCGCTGGTCTCCGCGGTGACGGGCGCGCCGTCAGTGCTGCACCTGTGGGGGCCT
GCCGGCACCGCCGACGCGTTCGCCCCCGGTCGGTGGCGAGCAGGCGGGTCTGCCGCAGGAC
CTCTCGACGCCTTCACCCGGTACGGCGCCGGGACGCTCAGCCACGACCTCGCCGACCAC
GTACTCGACCCCTGCCCGCCCCCGCTGCGGACCGCCGTGGCCGGGCGCGACGCCGGCATC
CGTTACGTCCCCTACAACGGCCCCGGGGCTGCCCCGCTCGACCTGCCCGAGCCGGACGGG
CGACGGCCGCGGGTCTGCGTGATCTGGGGCCGGTCGGTGACCAGGACGTTCGGACCGGTC
GTGAACCGGCTGCCACAGGCCGTCCGGGCCGCCGCCGACCTCGGTGCCGAGGTGCTGCTG
CTGGCCCGGCCGGAGGACGCCCGCGACGCCGGCCCGCTCCCCGACGGCGTGCGTCCGTTC
CACGAGGTCCCGCTGAGCCTGGTGCTGCCCGGGCTGCGAAGCGGTCGTGCACTACGCCGGT
GCTGGTTCGGTGATGACCGCGCTGACCGCGGGCGTGCCCGGCTGTGGTGCCCTGCCGGG
TTCGACCAGCCGATGGTCGCCGAGCGCCTGAGCGCGACCGGGGCCGGTCTGCACGTGCAC
AACCTCGACGCGGATGCCGCGACTCTGGGCGGCGCCCTGGAGAAGCTGATCGGACAGCCG
TCCTACGCCGACGCGGCCCGCGACCTCGCCCGGCGGTGCGCGCCATGCCCAGCCCGGCC
GAGGTCGTCGCGGACCTGGAGGCGCTGGCCGCCCGTTAGAAGCTT
```

Fig. S10 Synthetic SelSV sequence

4

```
>PenSV
CATATGCGCATCCTGTTCTCCGTGTCCTCGTGGACCGGCCACTACTACGCGATGTTCCCG
CTGGGCTGGGCGCTGCGCGCGCCGCCGGCCACGAGGTCCGCGTGCTCTGCCGCCCCGGCGAC
CGCGACGACGTCGTGCGGGCCGGCCTGGTCCCGGTGCCCGTGCTGGACGGGCCGGACATG
CTGACCGGGGCCCGCCTGCTGAACGTGCTCAGCTCGTTCTCCGGCACATGGCCCTACCCG
ACCCCGCCGCCGCACCCGGACAGCGGCGAGCCCGTCGACGACGCGTTCGACCTCCAGTCC
TGGCTGGCCGGCTACTGGGCGCGGGCGACCGAGTCCTCGCGCGCCGGCACCGACGCGGCG
GTGCGGTACGCGCGGGACTGGCAGCCGGACCTCGTGGTGCACGACCAGCTCAGCCTCGAC
GGCCCGCTCGTCGCCGCGGTTCTCGGGGTCCCGCAGGTGATGCACCTGTGGGGTCCGGTC
GGCCCCGCGGACTCGTTCGGTCCCGTCGGCGGCGAGGAGGCCGGCCGTCCCGCCGACCCG
AGCGACGCGTTCGGCCGCCACGGCGCCGGCGAGTTCGACCACGGTCGCGCCGACCTGATC
CTCGACCCGTGCCCCGGCCCGCTGGCCGCCGGGCTCGGCAAGCGCGTCCTGCCGATCCGC
TACGTGCCCTACAACGGGCCGGGCGACGACCCGGGCCCCGGCGCGCTGCCCGCACCGACC
GGGCGCAGGCGGGTCTGCGTCGTCTGGGGCCGCTCCGCGACGCGGACGTTCGGTCCGGTG
GCGAACAAGCTGCCCGAGGTGGTGCGGGCGGCCACCGACCACGGCGCCGACGTGCTGCTG
CTGGCCCGCCCGGACGACGCCGCCGCGGCAGGCGAGCTGCCGGAGAACGTGACCGCGATG
TGCGAGGTGCCGATGAGTCTCGTGCTCGACGGGTGCGACGCCGTCGTGCACTACGGCGGT
GCCGGGTCGGCGATGACGGCGCTCGTGGCCGGCCTGCCGCAGCTGTCCGTGCCGATCGGG
CTCGACCAGGATCTGGTGGCCGCGCGACTCGCCGCCACCGGCGCCGCCTGTACCGTGCCG
GGCCCGATCGCCGACGTGGAGTCGATCGGCAACGCGCTGGCCGCCGTGCTGGAGGGGCCG
CAGCACGCGGAGGCCGCCGAGAAGCTCGCCGCCGCCGTCGCCGCGATGCCGCCGCCCGCG
GCCGTCGTCGCCGATCTGGAGTCGCTCGTGGCCAGTCGCGCCCGCGCTGCCTGAAGCTT
```

Fig. S11. Synthetic PenSV coding sequence ordered from BaseClear. The GTG start codon was changed to ATG and three upstream nucleotides were added to create an NdeI restriction site (5′ CATATG 3′). Nucleotides were added after the stop codon to create a HindIII site (5′ AAGCTT 3′). Silent changes were made to eliminate internal PstI and NotI sites and to change the GGA glycine codon to GGT. This sequence was cloned between the NdeI and HindIII sites of the pET28a(+) expression plasmid.
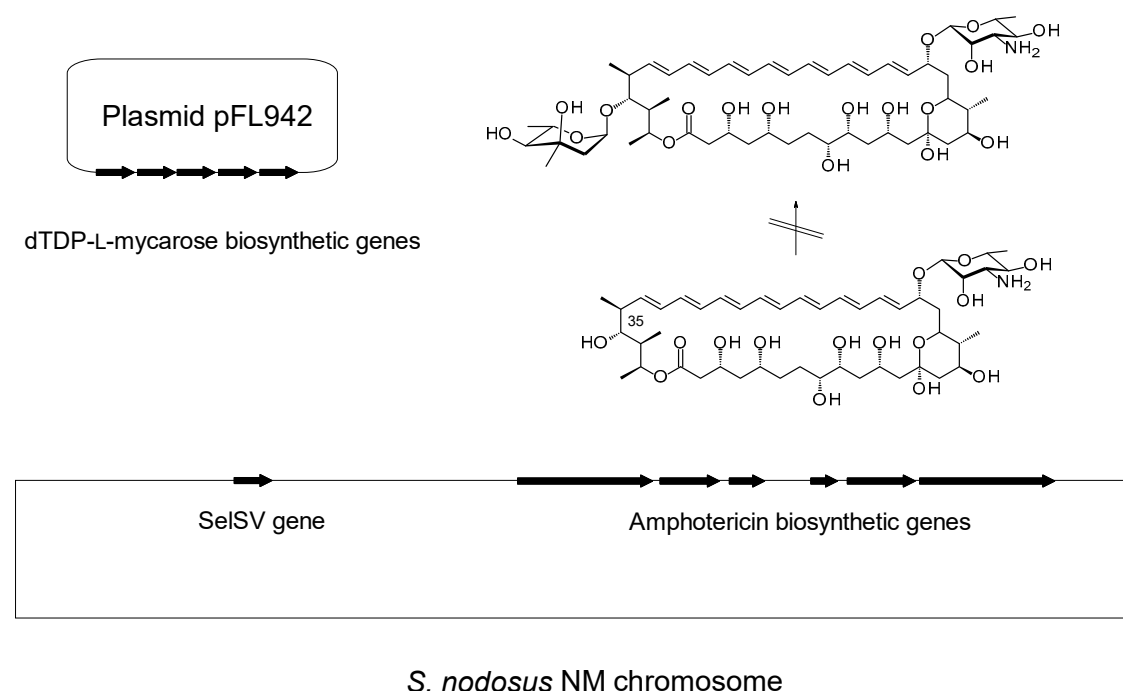

Fig. S12 Strategy for engineering biosynthesis of a mycarosylated or digitoxosylated amphotericin analogue. The *S. nodosus ΔamphNM* strain produces 16-methyl-amphotericin B and 8-deoxy-16-methyl-amphotericin A. This strain was transformed with a pIJ10257-based integrating plasmid containing a synthetic *selSV* gene. The resulting strain was then transformed with the Salas pFL942 plasmid that gives biosynthesis of dTDP-β-L-mycarose and dTDP-β-L-digitoxose. However, no mycarosylated or digitoxosylated analogues were detected by HPLC or mass spectrometry.
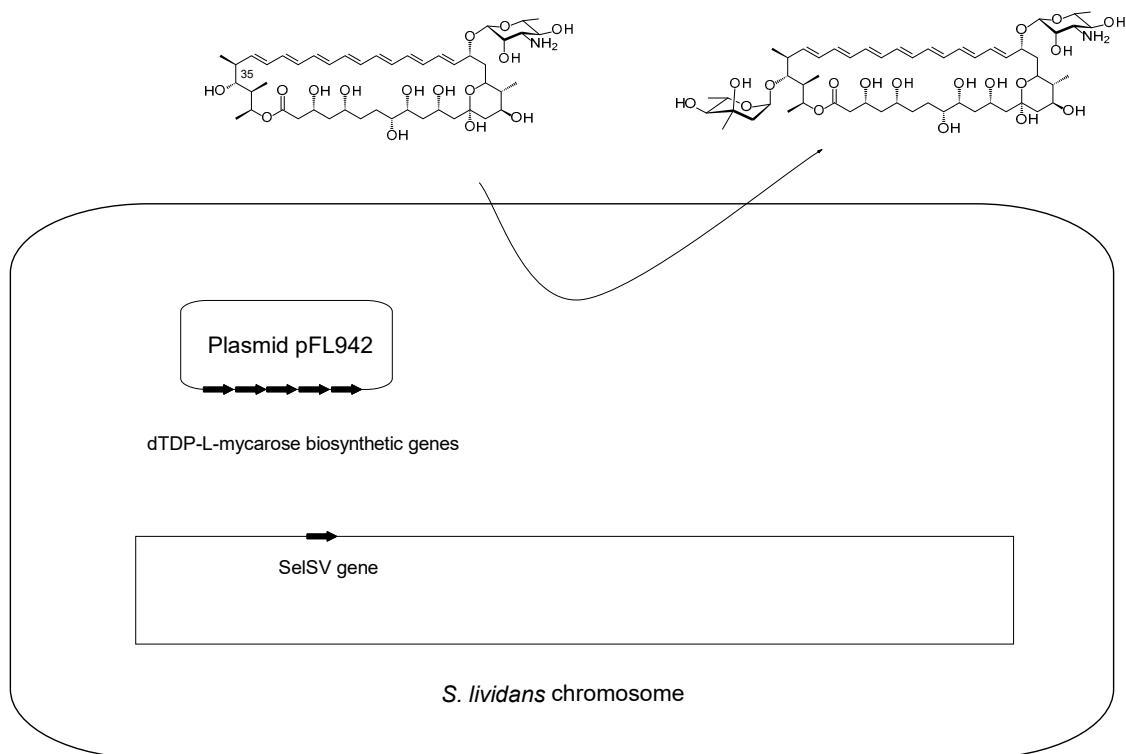
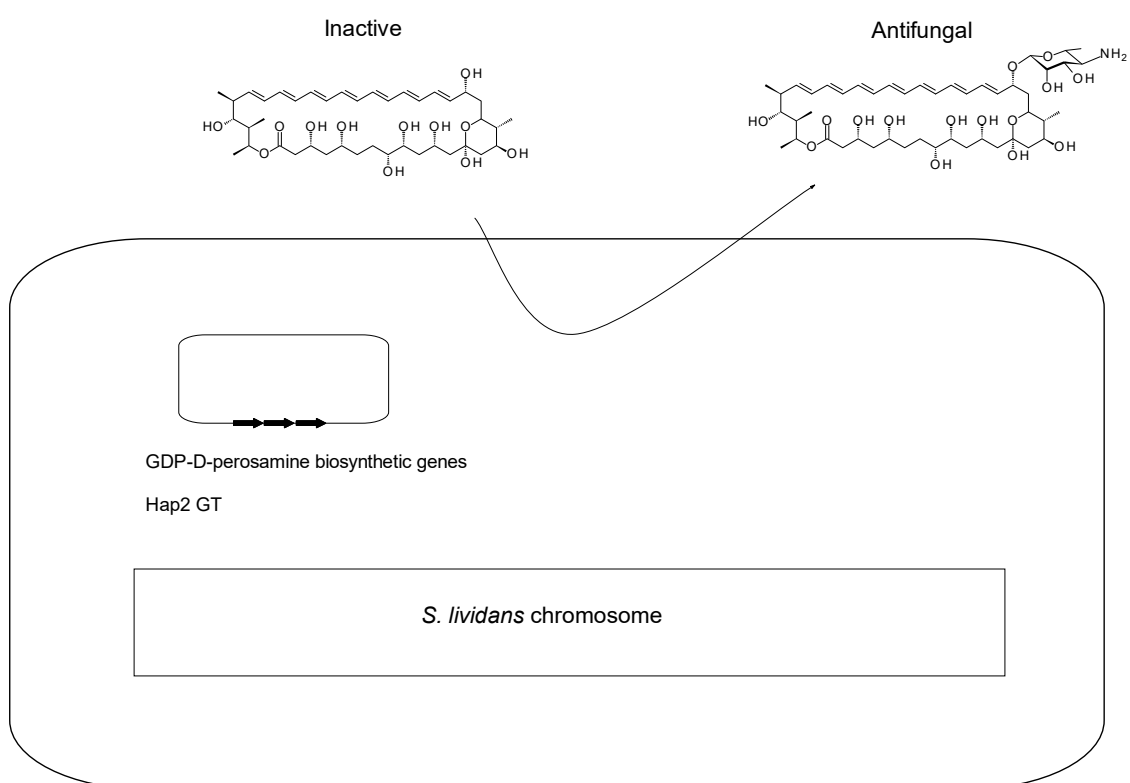Fig. S13 Schematic diagram illustrating strategy for aglycone feeding.

Fig. S14 Schematic diagram illustrating strategy for biotransformation of inactive amphoteronolides to active perosaminyl-amphoteronolides.

Table S1. *S. nodosus* transformants undergoing testing for production of new glycoanalogues.

Strain	Expected products
S. nodosus pIJ-*penSV* pIAGO	Amphotericins A and B
S. nodosus pIJ-*penSV* pFL942	Mycarosylated/digitoxosylated amphotericins A and B
S. nodosus NM pIJ-*penSV* pIAGO	16-Methyl-amphotericin B, 8-Deoxy-16-methyl amphotericin A
S. nodosus NM pIJ-*penSV* pFL942	Mycarosylated/digitoxosylated 16-methyl-amphotericin B and 8-deoxy-16-methyl amphotericin A
S. nodosus DI-DII-NM pIJ-*penSV* pIAGO	8-Deoxy-16-methyl-amphoteronolides A and B
S. nodosus DI-DII-NM pIJ-*penSV* pFL942	Mycarosylated/digitoxosylated 8-deoxy-16-methyl-amphoteronolides A and B

Table S2. Molecular masses of hypothetical 35-*O*-glycosylated amphotericins.

Polyene	Name	Exact mass	Molecular formula
1	Amphotericin B	923.4879	$C_{47}H_{73}NO_{17}$
2	Mycarosyl-amphotericin B	1067.5665	$C_{54}H_{85}NO_{20}$
3	Digitoxosyl-amphotericin B	1053.5508	$C_{53}H_{83}NO_{20}$
4	Amphotericin A	925.5035	$C_{47}H_{75}NO_{17}$
5	Mycarosyl-amphotericin A	1069.5821	$C_{54}H_{87}NO_{20}$
6	Digitoxosyl-amphotericin A	1055.5665	$C_{54}H_{87}NO_{20}$
7	16-Methyl-amphotericin B	893.5137	$C_{47}H_{75}NO_{15}$
8	Mycarosyl-16-methyl-amphotericin B	1037.5923	$C_{54}H_{87}NO_{18}$
9	Digitoxosyl-16-methyl-amphotericin B	1023.5767	$C_{53}H_{85}NO_{18}$
10	8-Deoxy-16-methyl-amphotericin A	879.5344	$C_{47}H_{77}NO_{14}$
11	Mycarosyl-8-Deoxy-16-methyl-amphotericin A	1023.6131	$C_{54}H_{89}NO_{17}$
12	Digitoxosyl-8-Deoxy-16-methyl-amphotericin A	1009.5974	$C_{53}H_{87}NO_{17}$
13	8-Deoxy-16-methyl-amphoteronolide B	732.4449	$C_{41}H_{64}O_{11}$
14	Mycarosyl-8-deoxy-16-methyl-amphoteronolide B	876.5235	$C_{48}H_{76}O_{14}$
15	Digitoxosyl-8-deoxy-16-methyl-amphoteronolide B	862.5079	$C_{47}H_{74}O_{14}$
16	8-Deoxy-16-methyl-amphoteronolide A	734.4605	$C_{41}H_{66}O_{11}$
17	Mycarosyl-8-deoxy-16-methyl-amphoteronolide A	878.5392	$C_{48}H_{78}O_{14}$
18	Digitoxosyl-8-deoxy-16-methyl-amphoteronolide A	864.5235	$C_{47}H_{76}O_{14}$

Table S3. Microbial strains

Micro-organism	Properties
Escherichia coli DH5α	General cloning host
Cryptosporangium arvum DSM44712	Predicted to synthesise disaccharide-containing octaene
S. nodosus	Synthesises amphotericins B and A
S. nodosus ΔamphNM	Synthesises 16-descarboxyl-16-methyl-amphotericin B and 8-deoxy-16-descarboxyl-16-methyl-amphotericin A
S. nodosus ΔamphDI-NM	Synthesises aglycones 8-deoxy-16-descarboxyl-16-methyl-amphoteronolides B and A, proficient in GDP-mycosamine biosynthesis
S. nodosus ΔamphDI-DII-NM	Synthesises aglycones 8-deoxy-16-descarboxyl-16-methyl-amphoteronolides B and A, deficient in GDP-mycosamine biosynthesis
S. nodosus ΔamphDIII	Synthesises aglycones 8-deoxy-amphoteronolides B and A
S. nodosus M57	Synthesises pentaene analogue of amphotericin B
Streptomyces albidoflavus DSM40624	Synthesises candicidins
Couchioplanes caeruleus DSM43634	Synthesises monosaccharide-containing 67-121A and disaccharide-containing 67-121C
Pseudonocardia endophytica DSM44969	Predicted to synthesise nystatin A3, which contains two unlinked monosaccharide sugar residues
Streptomyces cacaoi DSM40057	Synthesises perimycin
Streptomyces lividans	Host for biotransformation strain
Streptomyces eurocidicus DSM40604	Synthesises eurocidins
Saccharopolyspora gloriosae DSM45582	Predicted to synthesise selvamicin-related pentaene containing two unlinked monosaccharide sugar residues
Candida albicans	Indicator organism for antifungal assays

Table S4. Oligonucleotides

Oligonucleotide	Sequence
NGTF3	5′ AGCTAGATCTTGCAGGGCACCGGCAAACGCCCG 3′
NGTR1	5′ GATCAAGCTTAAGGCACGTCAGCCGGACGGAAG 3′
PerD3F	5′ GATCAAGCTTGAGGAGACGGGATGTCCAAGCGCGCG 3′
PerD3R	5′ GATCAAGCTTCCCTAGTGATGCCGCGGGCTAC 3′
CrypF	5′ ACTGGGATCCGGCGACATCGGAGGGCTGGCGGCATG 3′
CrypR	5′ GATCAAGCTTCATCGGTTCGCTCCGGTCATCAG 3′
SGDF1	5′ ATGCAGATCTCGCGAAGGGCGTGTTCATG 3′
SGDR	5′ GTCAAAGCTTCGAAACGGGGTAGCGCATCAG 3′
EurDI-F	5′ GTACGGATCCACTGAAAGGGAATCCATGGACTC 3′
EurM-R	5′ GATCAAGCTTCTCAGGTGCTTTCATCGATG 3′