

Supplementary information

Chevrette et al.

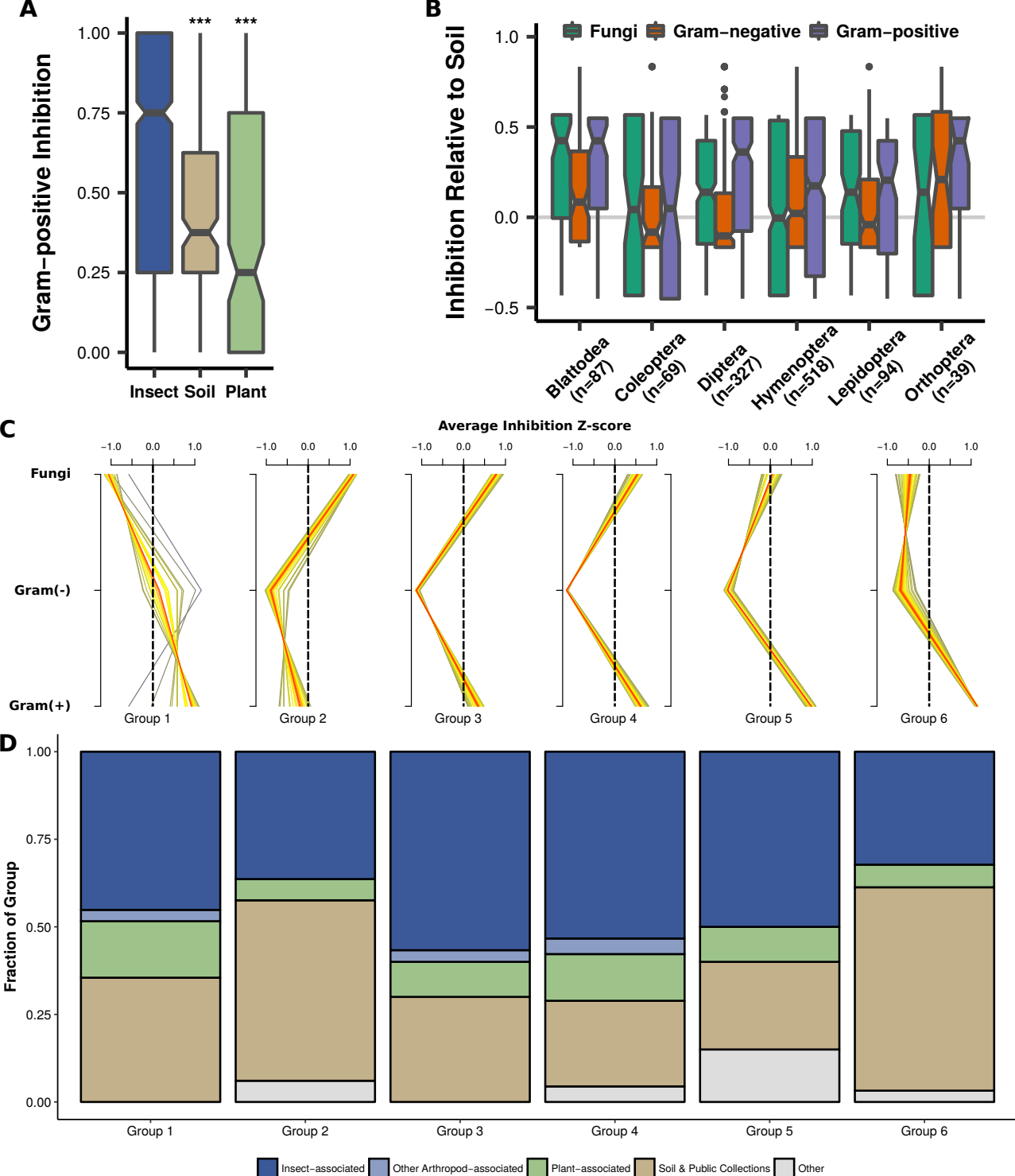
The antimicrobial potential of *Streptomyces* from insect microbiomes

Supplementary Figure 1: Inhibition bioassays.

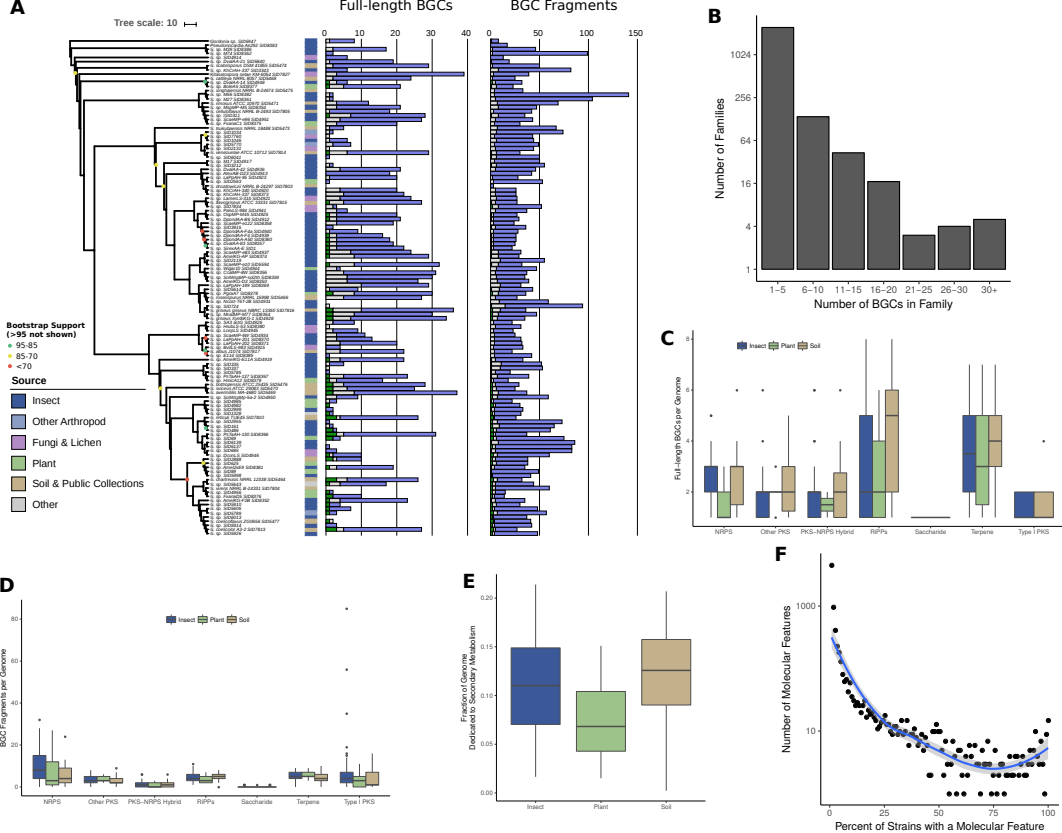
Supplementary Figure 2: Genomics and metabolomics.

Supplementary Figure 3: Structural determination of cyphomycin.

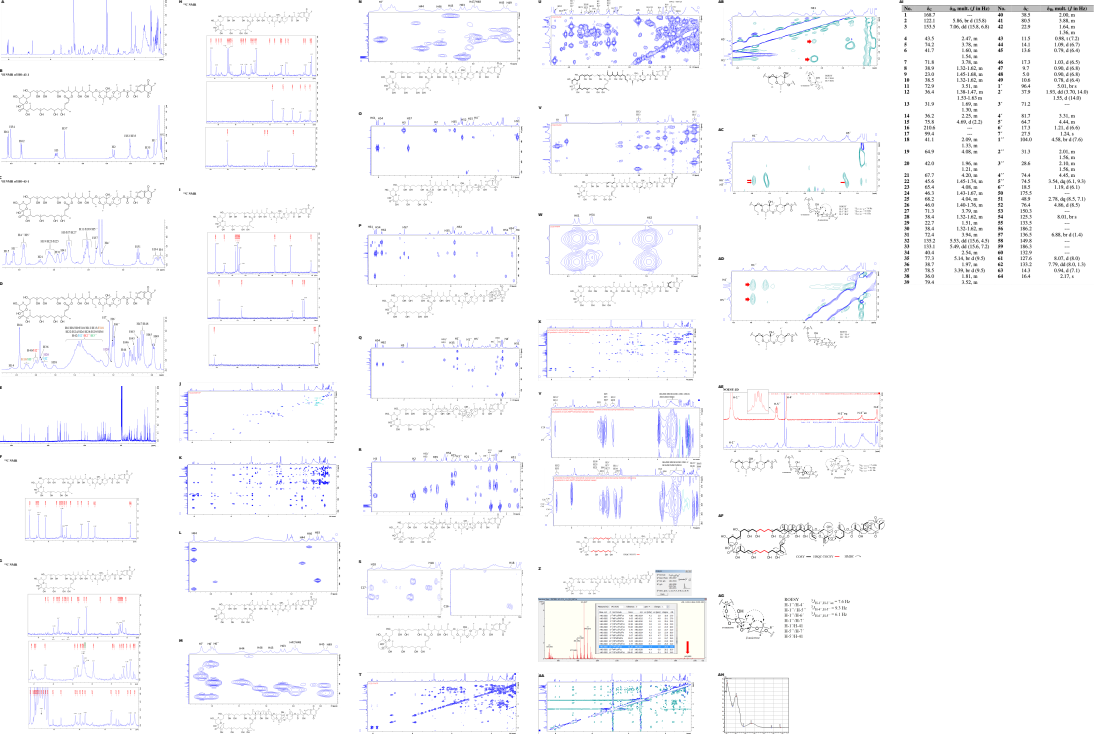
Supplementary Figure 4: Proposed biosynthetic logic of cyphomycin.



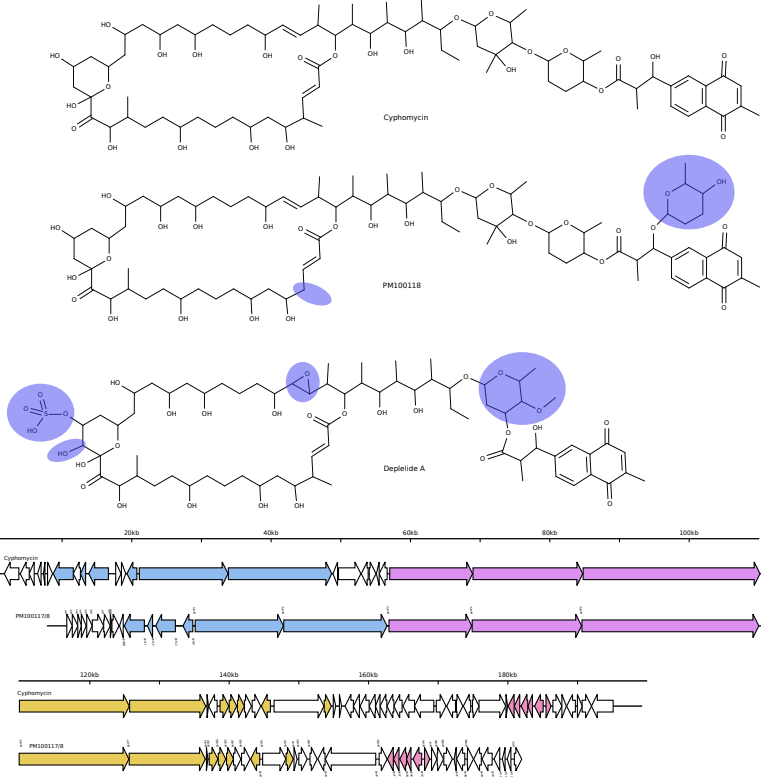
Supplementary Figure 1. Inhibition bioassays. (A) Gram-positive inhibition by *Streptomyces* strains from insect, soil, or plant sources. Insect microbiome *Streptomyces* show significantly higher inhibition against Gram-positive bacteria than do soil- or plant-sourced *Streptomyces*. (n=1162, 186, and 178 for insect, soil, and plant, respectively; ***p<1e-3; t-test, BY correction) (B) *Streptomyces* from insect hosts of different orders have differential activity in inhibition bioassays against Gram-positive bacteria, Gram-negative bacteria, and fungi. On average, *Streptomyces* from these hosts show higher activity compared to soil *Streptomyces*. (n=87, 69, 327, 518, 94, and 39 for Blattodea, Coleoptera, Diptera, Hymenoptera, Lepidoptera, and Orthoptera, respectively). A, B: center, median; box, upper and lower quartiles; notches, 95% confidence; whiskers, 1.5x interquartile range; points, outliers. (C) Fuzzy clustering of *Streptomyces* isolates identified groups of strains by their patterns of inhibitory activity against Gram-positive bacteria, Gram-negative bacteria, and fungi. Z-scores of inhibitory activities (deviance from the mean, 0) are shown for fungi, Gram-negative bacteria, and Gram-positive bacteria. Each line corresponds to an isolate within a given group. The color of the line denotes how strongly a strain fits into the group. For example, group 1 shows much less than average fungal activity but relatively high Gram-positive activity. Some strains within this group have high and relatively specific Gram-negative activity. In group 2, there is average Gram-positive activity, low Gram-negative activity, and high fungal activity. (D) Membership of these groups shows patterns of inhibition correspond to patterns of community membership. For example, group 6 has a higher proportion of soil strains than does group 3.



Supplementary Figure 2. Genomics and metabolomics. (A) Core-genome phylogeny based on a concatenated 93-gene multilocus sequence of core, single copy bacterial genes (JCVI GenProp0799; see http://www.jcvi.org/cgi-bin/genome-properties/GenomePropDefinition.cgi?prop_acc=GenProp0799). *Gordonia* sp. SID5947 was used as an outgroup. Nodes with bootstrap support below 95 are marked. Source (e.g., insect-associated, other arthropod-associated, fungi-associated, plant-associated, or soil) is shown in the color strip to the right of the tree. Full-length BGCs (BGCs with at least 0.5kb of DNA on either side of a cluster boundary) and BGC fragments for each genome are shown at right. All were compared to the known BGCs in the MiBIG database. As in Figure 2A, green denotes a BiG-SCAPE distance of less than 0.5 (similar), grey is a distance between 0.5 and 0.75 (divergent), and blue is a distance over 0.75 (uncharacterized). (B) The distribution of the number of BGCs in each family, as defined by a BiG-SCAPE distance of 0.5 or less, in *Streptomyces*. Most BGC families have 5 or less examples in our dataset. (C) The number of BGC fragments per genome by type for each of insect-, plant-, or soil-sourced samples. (D) The number of full-length BGCs per genome by type for each of insect-, plant-, or soil-sourced samples. (E) The fraction of the genome dedicated to secondary metabolism for insect, plant, or soil isolates. C, D, E: $n=69$, 21, and 13 for insect, soil, and plant respectively; center, median; box, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers. (F) Distribution of LC/MS molecular features across *Streptomyces*. As with BGCs (panel B above) most molecular features are in a small percentage of strains within the dataset. A LOESS regression curve is shown in blue with 95% confidence intervals shown in grey.



Supplementary Figure 3. Structural determination of cyphomycin. (A) ¹H NMR (CD₃OD, 500 MHz). (B) ¹H NMR (CD₃OD, 500 MHz). Expansion 1. (C) ¹H NMR (CD₃OD, 500 MHz). Expansion 2. (D) ¹H NMR (CD₃OD, 500 MHz). Expansion 3. (E) ¹³C NMR (CD₃OD, 125 MHz). Expansion 1. (F) ¹³C NMR (CD₃OD, 125 MHz). Expansion 2. (G) ¹³C NMR (CD₃OD, 125 MHz). Expansion 3. (H) ¹³C NMR (CD₃OD, 125 MHz). Expansion 4. (I) ¹³C NMR (CD₃OD, 125 MHz). Expansion 5. (J) Multiplicity-edited gHSQC (CD₃OD, 500 MHz). (K) gHMBC (CD₃OD, 500 MHz). (L) gHMBC (CD₃OD, 500 MHz). Expansion 1. (M) gHMBC (CD₃OD, 500 MHz). Expansion 2. (N) gHMBC (CD₃OD, 500 MHz). Expansion 3. (O) gHMBC (CD₃OD, 500 MHz). Expansion 4. (P) gHMBC (CD₃OD, 500 MHz). Expansion 5. (Q) gHMBC (CD₃OD, 500 MHz). Expansion 6. (R) gHMBC (CD₃OD, 500 MHz). Expansion 7. (S) gHMBC (CD₃OD, 500 MHz). Expansion 8. (T) gCOSY (CD₃OD, 500 MHz). (U) gCOSY (CD₃OD, 500 MHz). Expansion 1. (V) gCOSY (CD₃OD, 500 MHz). Expansion 2. (W) gCOSY (CD₃OD, 500 MHz). Expansion 3. (X) gHSQC-TOCSY (CD₃OD, 500 MHz). (Y) gHSQC-TOCSY (CD₃OD, 500 MHz). Expansion 1. (Z) HRESIMS spectrum. (AA) gROESY (CD₃OD, 500 MHz). (AB) gROESY (CD₃OD, 500 MHz). Expansion 1. (AC) gROESY (CD₃OD, 500 MHz). Expansion 2. (AD) gROESY (CD₃OD, 500 MHz). Expansion 3. (AE) NOESY-1D (CD₃OD, 500 MHz). (AF) Key COSY, HSQC-TOCSY and HMBC correlations. (AG) Relative configuration proposed for disaccharide moiety. (AH) UV and optical rotation [α]_D²³ = -96.4 (c 0.21, MeOH); UV (MeOH) λ_{max} (log ϵ) 204 (4.8), 254 (4.7), 334 (3.8); ¹H and ¹³C NMR (CD₃OD, 500/125 MHz); HRESIMS m/z 1463.8282 [M + H]⁺ (calculated for C₇₇H₁₂₃O₂₆, 1463.8297, err 1.0 ppm). (AI) ¹H and ¹³C NMR (CD₃OD, 500/125 MHz).



Supplementary Figure 4. Proposed biosynthetic logic of cyphomycin. The structures of cyphomycin, PM100118, and deplelide A are shown with key differences marked in blue. Specifically, PM100117 and PM100118 do not have a methyl on carbon 4. The second sugar in the PM100118 tail, L-2-deoxifucose, alternatively be hydroxylated and become L-rhodinose in PM100117. In contrast to cyphomycin, both have an L-rhodinose as the third sugar in the tail, the loss of which has been shown to greatly effect antitumor activity. Deplelide A varies from cyphomycin in many places, including an extra sulfonic acid, epoxide, and hydroxyl attached to the core macrolide, and variance in both the number and type of sugars in the tail. Cyphomycin has a β -amictose in second sugar unit compared to PM100117/8. Also shown are the biosynthetic gene clusters for cyphomycin and PM100117/8. Here we propose a similar T1PKS biosynthetic logic for the macrolide core as that described previously for PM100117/8, from an isolate associated with a marine polychaete worm of the genus *Filograna*. Open reading frames are depicted as arrows and colors correspond to groups shared gene content between the two clusters (proteins have over 75% coverage and 75% identity). Genes that are not shared between the two are colored white. There are no differences in domain or predicted elongation unit specificity within the T1PKS itself. Interestingly, the acyltransferase of module 19 in PM100117/8 (first module of gonP7) is predicted to encode specificity for methylmalonyl-CoA, but the final structure is missing this methyl. This suggests either a malonyl-CoA is incorporated at this position or the methyl is cleaved in post-PKS tailoring. In contrast, cyphomycin has this methyl moiety.