

# PNAS

[www.pnas.org](http://www.pnas.org)

Supplementary Information for

GRINS: genetic elements that recode assembly-line polyketide synthases and accelerate their diversification

Aleksandra Nivina<sup>1,2,3</sup>, Sur Herrera Paredes<sup>1,4</sup>, Hunter Fraser<sup>4</sup>, Chaitan Khosla<sup>2,3,5</sup>

<sup>1</sup>Equal contribution

<sup>2</sup>Department of Chemistry, Stanford University, Stanford, California 94305, United States

<sup>3</sup>Stanford ChEM-H, Stanford University, Stanford, California 94305, United States

<sup>4</sup>Department of Biology, Stanford University, Stanford, California 94305, United States

<sup>5</sup>Department of Chemical Engineering Stanford University, Stanford, California 94305, United States

Chaitan Khosla

Email: [khosla@stanford.edu](mailto:khosla@stanford.edu)

**This PDF file includes:**

Table S1

Figure S2

Figure S3

Figure S4

Table S5

Table S6

Figure S7

Figure S8

**Table S1.** List of known assembly-line PKS clusters, their NCBI accessions, and results of duplicated region and GRINS detection.

| <b>Known PKS clusters</b>                      | <b>NCBI accession and cluster number</b> | <b>Duplications</b> | <b>GRINS</b> |
|--|--|---------------------|--------------|
| Oligomycin synthase                            | AB070940-1                               | Yes                 | Yes          |
| Vicenistatin synthase                          | AB086653-1                               | Yes                 | Yes          |
| Mycinamicin synthase                           | AB089954-1                               | Yes                 | Yes          |
| Neocarzililn synthase                          | AB097904-1                               | Yes                 | No           |
| Halstoctacosanolide synthase                   | AB241068-1                               | Yes                 | Yes          |
| Nemadectin synthase                            | AB363939-1                               | Yes                 | Yes          |
| Pladienolide synthase                          | AB435553-1                               | Yes                 | Yes          |
| Lasalocid synthase                             | FM173265-1                               | Yes                 | Yes          |
| FD-891 synthase                                | AB469193-1                               | Yes                 | Yes          |
| Reveromycin synthase                           | AB568601-1                               | Yes                 | Yes          |
| Incednine synthase                             | AB767280-1                               | Yes                 | Yes          |
| Cremimycin synthase                            | AB818354-1                               | Yes                 | Yes          |
| Niddamycin synthase                            | AF016585-1                               | Yes                 | Yes          |
| Rifamycin synthase                             | AF040570-1                               | Yes                 | Yes          |
| Methymycin/pikromycin/narbonolide synthase     | AF079138-1                               | Yes                 | No           |
| Oleandolide and 8,8a-deoxyoleandolide synthase | AR159871-1                               | Yes                 | Yes          |
| Megalomicin synthase                           | AF263245-1                               | Yes                 | Yes          |
| Simocyclinone D8 synthase                      | AF324838-1                               | Yes                 | Yes          |
| Amphotericin synthase                          | AF357202-1                               | Yes                 | Yes          |
| Monensin synthase                              | AF440781-1                               | Yes                 | Yes          |
| Ansamitocin synthase                           | AF453501-1                               | Yes                 | No           |
| Nanchangmycin synthase                         | AF521085-1                               | Yes                 | Yes          |
| Natamycin/pimaricin synthase                   | HQ386234-1                               | Yes                 | Yes          |
| Stigmolone synthase                            | AJ421825-1                               | Yes                 | No           |
| Spirangien synthase                            | AM407731-1                               | Yes                 | No           |

|  |                |     |     |
|--|----------------|-----|-----|
| Aureothin synthase                     | AJ575648-1     | Yes | No  |
| Borrelin synthase                      | KT362046-1     | Yes | Yes |
| Stambomycin a synthase                 | AM238664-4     | Yes | Yes |
| DEBS (Erythromycin synthase)           | AM420293-3     | Yes | Yes |
| Neoaurerthrin synthase                 | AM778535-1     | Yes | No  |
| Kendomycin synthase                    | AM992894-1     | Yes | Yes |
| Zincophorin synthase                   | KT345957-1     | Yes | Yes |
| Spinosyn synthase                      | AY007564-1     | Yes | Yes |
| Midecamycin synthase                   | BD420675-1     | Yes | Yes |
| Nystatin synthase                      | AF263912-1     | Yes | Yes |
| Dihydrochalcone synthase               | AY118081-1     | Yes | Yes |
| Geldanamycin synthase                  | AY179507-1     | Yes | Yes |
| Hedamycin synthase                     | AY196994-1     | Yes | Yes |
| Candididin (FR-008) synthase           | AY310323-1     | Yes | Yes |
| Lactomycin and phoslactomycin synthase | AY354515-1     | Yes | Yes |
| Pyoluteorin synthase                   | AY394844-1     | No  | No  |
| Obscurin synthase                      | AY466441-1     | Yes | Yes |
| Chalcone synthase                      | AY509120-1     | Yes | Yes |
| Erythromycin synthase                  | AY623658-1     | Yes | Yes |
| ECO-02301 synthase                     | AY899214-1     | Yes | Yes |
| Filipin (polyene) synthase             | BA000030-3     | Yes | Yes |
| Avermectin synthase                    | BA000030-6     | Yes | Yes |
| Lorneic acid a synthase                | BBOM01000004-1 | Yes | Yes |
| Akaiolide synthase                     | BBOM01000011-4 | Yes | Yes |
| Pyoluteorin synthase                   | AF081920-1     | No  | No  |
| Aculeximycin synthase                  | CP007155-35    | Yes | Yes |
| Chlorothricin synthase                 | DQ116941-1     | Yes | Yes |
| Concanamycin synthase                  | DQ149987-1     | Yes | Yes |
| E-837 synthase                         | DQ272520-1     | Yes | Yes |
| Nigericin synthase                     | DQ354110-1     | Yes | Yes |

|                                     |                |     |     |
|-------------------------------------|----------------|-----|-----|
| Ambruticin synthase                 | DQ897667-1     | Yes | No  |
| Jerangolid a synthase               | DQ897668-1     | Yes | Yes |
| Tautomycetin synthase               | EU035755-1     | Yes | Yes |
| Tautomycin synthase                 | EF990140-1     | Yes | Yes |
| Nystatin-like polyene synthase      | EU108007-1     | Yes | No  |
| Angolamycin and anglomycin synthase | EU220288-1     | Yes | Yes |
| Kijanamicin synthase                | EU301739-1     | Yes | Yes |
| Tetrocarcin synthase                | EU443633-1     | Yes | Yes |
| Macbecin synthase                   | EU827593-1     | Yes | Yes |
| Indanomycin synthase                | FJ545274-1     | Yes | Yes |
| BE-14106 synthase                   | FJ872523-1     | Yes | Yes |
| Heronamide synthase                 | LT548273-1     | Yes | No  |
| Meilingmycin synthase               | FJ952082-1and2 | Yes | Yes |
| Herboxidiene synthase               | JN671974-1     | Yes | Yes |
| Elaiophylin synthase                | GP697151-1     | Yes | Yes |
| Cylindrospermopsin synthase         | GQ385961-1     | Yes | No  |
| Naphthomycin synthase               | GQ452266-1     | Yes | Yes |
| Bafilomycin synthase                | GU390405-1     | Yes | Yes |
| Salinomycin synthase                | HE586118-1     | Yes | Yes |
| L-155, L-175 synthase               | HE648167-1     | Yes | Yes |
| Divergolide synthase                | HF563079-1     | Yes | No  |
| Calcimycin synthase                 | HM452329-1     | Yes | Yes |
| Tiacumicin synthase                 | HQ011923-1     | Yes | Yes |
| Fostriecin synthase                 | HQ434551-1     | Yes | Yes |
| Ansatrienin (mycotrienin) synthase  | JF803483-1     | Yes | Yes |
| Apoptolidin synthase                | JF819834-1     | Yes | Yes |
| Quartromycin synthase               | JF970188-1     | Yes | No  |
| Tetramycin synthase                 | JN688154-1     | Yes | Yes |
| Laidlomycin synthase                | JQ793783-1     | Yes | Yes |
| Marinopyrrole synthase              | JX157625-1     | No  | No  |

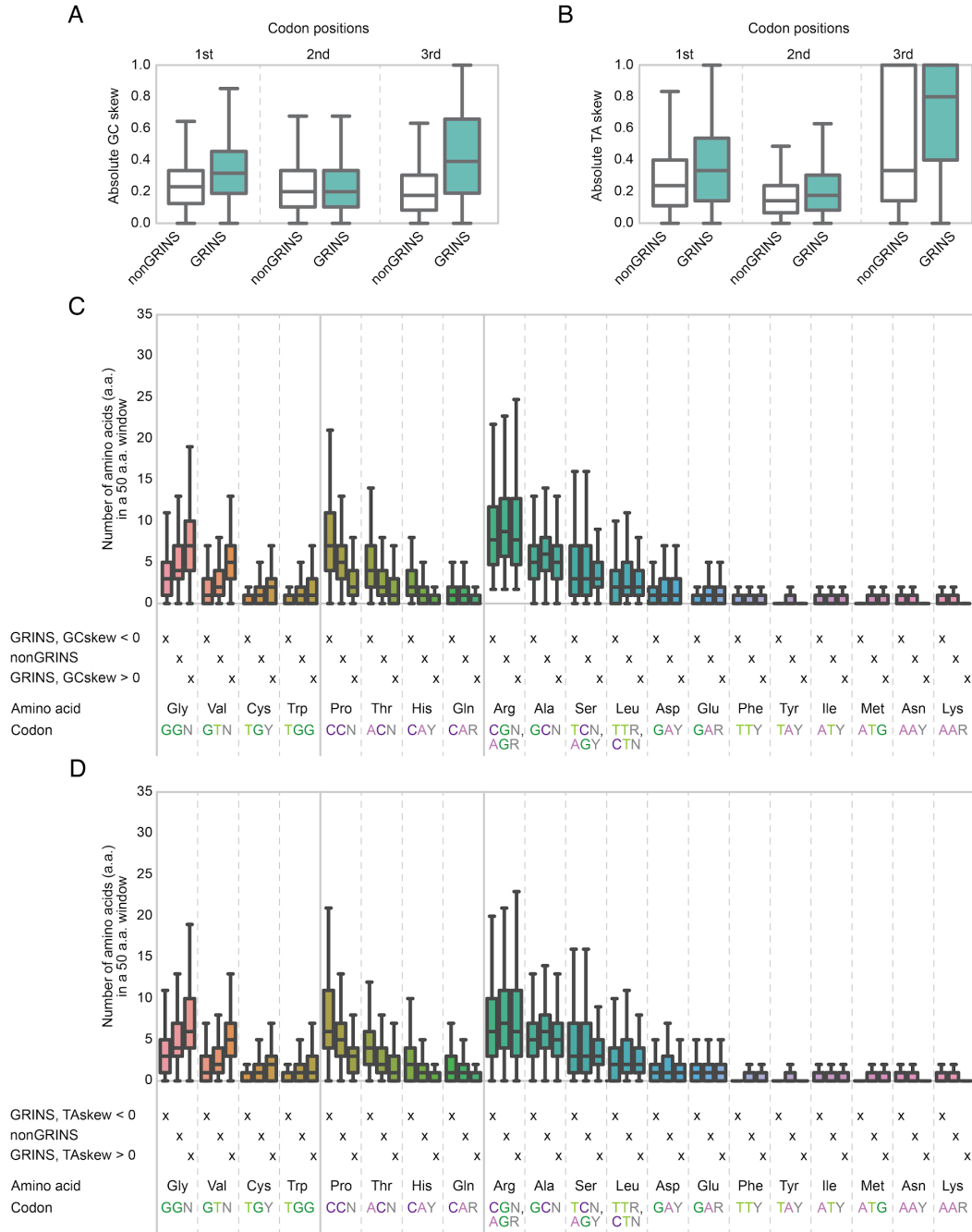
|  |            |     |     |
|--|------------|-----|-----|
| Hygrocin synthase                            | JX504844-1 | Yes | Yes |
| Lobophorin synthase                          | KC013978-1 | Yes | Yes |
| Cuevaene synthase                            | KF017602-1 | Yes | Yes |
| Gephyronic acid synthase                     | KF479198-1 | Yes | No  |
| 4-Z-annimycin synthase                       | KF683117-1 | Yes | Yes |
| Piericidin synthase                          | KF874660-1 | Yes | Yes |
| FK506 synthase                               | HM116536-1 | Yes | Yes |
| Ibomycin synthase                            | KJ159185-1 | Yes | Yes |
| Neoansamycin synthase                        | KJ590158-1 | Yes | Yes |
| Anatoxin synthase                            | KM245023-1 | Yes | No  |
| Gulmirecin synthase                          | KM361622-1 | Yes | No  |
| Brasilinolide synthase                       | KP161205-1 | Yes | Yes |
| Rosamicin/salinopyrone/pacificanone synthase | KP997155-1 | Yes | Yes |
| Ansalactam synthase                          | KR270805-1 | Yes | Yes |
| Lobosamide synthase                          | KT209587-1 | Yes | Yes |
| Cyclizidine synthase                         | KT327068-1 | Yes | Yes |
| Aldgamycin j synthase                        | KU568466-1 | Yes | Yes |
| Sceliphrolactam synthase                     | KX230849-1 | Yes | No  |
| Abyssomicin synthase                         | KY432814-1 | Yes | Yes |
| Azalomycin synthase                          | KY484834-1 | Yes | Yes |
| Streptovaricin synthase                      | KY593296-1 | Yes | Yes |
| Versipelostatin synthase                     | LC006086-1 | Yes | Yes |
| Hitachimycin synthase                        | LC008143-1 | Yes | Yes |
| Maklamicin synthase                          | LC021382-1 | Yes | Yes |
| Streptazone synthase                         | LC061217-1 | Yes | Yes |
| Fluvirucin synthase                          | LC095592-1 | Yes | Yes |
| Neomediomycin synthase                       | LC208004-1 | Yes | Yes |
| Mediomycin synthase                          | LC208005-1 | Yes | No  |
| Tetrafibricin synthase                       | LC208006-1 | Yes | Yes |
| Trichostatin synthase                        | LC217606-1 | Yes | Yes |

|  |                         |     |     |
|--|-------------------------|-----|-----|
| Actinoallolide synthase                          | LC326402-1              | Yes | Yes |
| Lavendiol synthase                               | LC330869-1              | Yes | No  |
| Chaxamycin synthase                              | LN831790-9              | Yes | Yes |
| Conglobatin synthase                             | LN849060-1              | Yes | No  |
| PM100117 and PM100118 synthase                   | LN997801,<br>LN997802-1 | Yes | Yes |
| Ebelactone synthase                              | KC894072-1              | Yes | Yes |
| Argimycin synthase                               | LT615255-1              | Yes | Yes |
| 67-121C synthase                                 | MEIA01000001-1          | Yes | Yes |
| Juvenimicin/rosamicin/M-4365/izenamicin synthase | MF033535-1              | Yes | Yes |
| Niphimycin synthase                              | MF671979-1              | Yes | Yes |
| Rifamorpholine synthase                          | MF773686-1              | Yes | No  |
| Tetrapetalone synthase                           | MG603754-1              | Yes | Yes |
| Streptoseomycin synthase                         | MG891745-1              | Yes | Yes |
| Epothilone synthase                              | AF210843-1              | Yes | No  |
| FK520 synthase                                   | AF235504-1              | Yes | Yes |
| Milbemycin synthase                              | CP002047-5and6          | Yes | Yes |
| Rapamycin synthase                               | X86780-1                | Yes | No  |
| Rubradirin synthase                              | AJ871581-1              | Yes | No  |
| Spiramycin synthase                              | CP012382-13             | Yes | Yes |
| Lankamycin synthase                              | AB088224-1              | Yes | Yes |
| Jamaicamide synthase                             | AY522504-1              | Yes | No  |
| Curacin synthase                                 | AY652953-1              | Yes | Yes |
| Myxothiazol synthase                             | AF188287-1              | Yes | Yes |
| Butyrolactol synthase                            | BBOK01000014-1          | Yes | Yes |
| Linfuranone synthase                             | BCBX01000035-1          | Yes | Yes |
| Ajudazol synthase                                | AM946600-1              | Yes | No  |
| Chondrochloren synthase                          | AM988861-1              | Yes | Yes |
| Herbimycin synthase                              | AY947889-1              | Yes | Yes |
| Meridamycin synthase                             | DQ351275-1              | Yes | Yes |

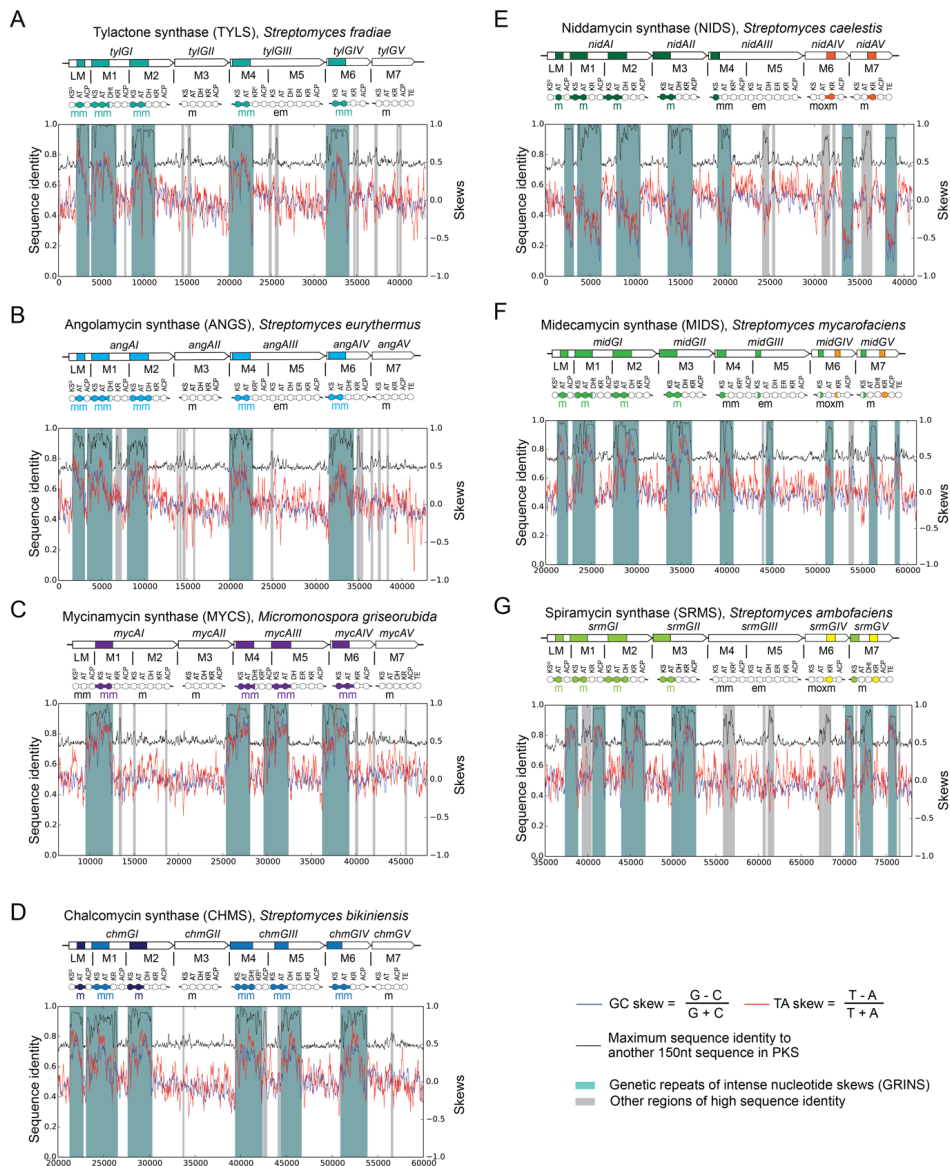
|                       |                |     |     |
|-----------------------|----------------|-----|-----|
| Mycolactone synthase  | BX649209-1     | Yes | Yes |
| Tylactone synthase    | U78289-1       | Yes | Yes |
| Soraphen synthase     | U24241-1       | Yes | No  |
| Salinilactam synthase | CP000667-12    | Yes | Yes |
| Mirilactam synthase   | CP001630-5     | Yes | No  |
| ML-449 synthase       | FJ872525-1     | Yes | Yes |
| NOCAP synthase        | BAFR01000107-1 | Yes | No  |
| Ossamycin synthase    | MH763624-1     | Yes | Yes |
| Lydicamycin synthase  | MG459168-1     | Yes | Yes |
| Thailandin synthase   | MKQR01000009-2 | Yes | No  |
| E-492, E-975 synthase | AH015386-1     | Yes | Yes |
| Leptomycin            | EA357094-1     | Yes | Yes |
| Nargenicin synthase   | MH544245-1     | Yes | Yes |
| Albicidin             | AJ586576-1     | Yes | No  |
| Bacillaene            | AJ634060-1     | Yes | No  |
| Basiliskamide         | AXBT01000013-1 | Yes | Yes |
| Batumin               | GU479979-1     | Yes | No  |
| Bongkrelic acid       | JX173632-1     | Yes | No  |
| Bryostatins           | DQ889941-1     | Yes | No  |
| Calyculin             | AB933566-x     | Yes | No  |
| Chivosazol            | DQ065771-1     | Yes | No  |
| Corallopyronin        | HM071004-1     | Yes | No  |
| Cycloheximide         | JX014302-1     | No  | No  |
| Diaphorin             | CP003468-1     | Yes | No  |
| Difficidin            | AJ634062-1     | Yes | No  |
| Disorazole            | DQ013294-1     | Yes | No  |
| Elansolid             | HQ680975-1     | Yes | No  |
| Enacyloxin            | CP000442-1     | Yes | No  |
| Etnangien             | AM746676-3     | Yes | No  |
| Grieseoviridin        | JX508597-1     | Yes | Yes |

|                                |             |     |     |
|--------------------------------|-------------|-----|-----|
| Spliceostatin/FR901464         | HM047288-1  | Yes | No  |
| Dorrigocin/migrastatin         | GQ274953-1  | Yes | No  |
| Kirromycin                     | AM746336-1  | Yes | No  |
| Lactimidomycin                 | GQ274954-1  | Yes | Yes |
| Lankacidin synthase            | AB088224-2  | Yes | Yes |
| Legioliulin                    | KM222819-1  | No  | No  |
| Leinamycin                     | AF484556-1  | Yes | No  |
| Luminaolide                    | KR857272-1  | Yes | No  |
| Macrolactin                    | AJ634061-1  | No  | No  |
| Malleilactone                  | NC_007650-1 | No  | No  |
| 9-methylstreptim/Streptimidone | FR878059-1  | Yes | No  |
| Misakinolide                   | KR857273-1  | Yes | No  |
| Mupirocin                      | AF318063-1  | Yes | No  |
| Myxopyronin                    | KF356280-1  | No  | No  |
| Myxovirescin                   | CP000113-7  | Yes | No  |
| Nosperin                       | GQ979609-1  | No  | No  |
| Onnamide                       | AY688304-1  | No  | No  |
| Oocydin                        | JX315603-1  | No  | No  |
| Oxazolomycin                   | EF552687-1  | Yes | No  |
| Pederin                        | AH013687-1  | Yes | No  |
| Psymberin                      | FJ823461-1  | Yes | No  |
| Rhizopodin                     | FR854394-1  | No  | No  |
| Rhizoxin                       | AM411073-1  | No  | No  |
| SIA7248                        | JQ269659-1  | Yes | Yes |
| Sorangicin                     | HM584908-1  | Yes | No  |
| Tartrolon                      | CP001614-1  | No  | No  |
| Thailandamide                  | CP000085-2  | Yes | No  |
| Thiomarinol                    | FN689524-1  | No  | No  |
| Virginiamycin                  | AB283030-1  | Yes | No  |

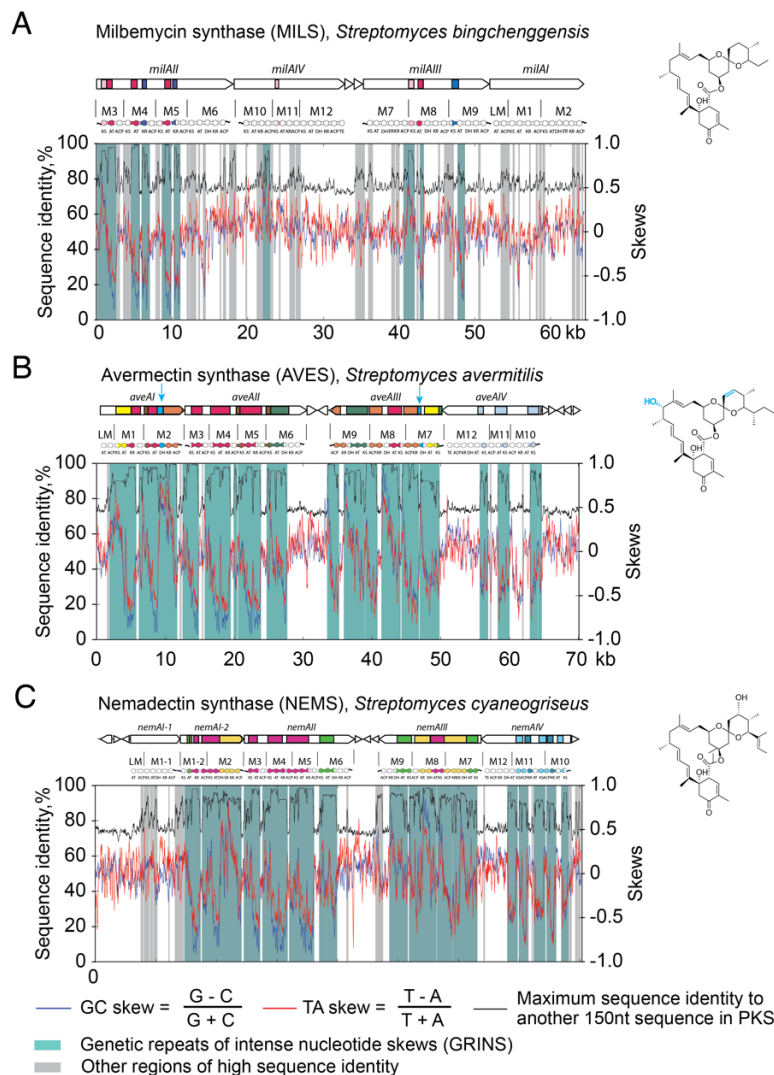




**Fig. S2. GRINS recode regions of assembly-line PKSs. A-B.** Absolute GC skews (A) and TA skews (B) in first, second and third codon positions of non-GRINS and GRINS regions of PKS genes. Even though all differences between non-GRINS and GRINS regions are significant (t-test,  $p < 0.0001$ ), the differences in median values are the most pronounced for third codon positions. Skews were measured for first, second and third codon positions in 150nt sliding windows of CDSs with a 30nt step size, for each cluster among 203 known assembly-line PKSs. **C.** Frequency of each amino acid (a.a.) within sliding windows of 50a.a., with a step size of 10a.a. in GRINS regions with a negative GC skew; in non-GRINS regions; in GRINS regions with a positive GC skew. **D.** Same as **C**, but for GRINS regions with a negative TA skew; non-GRINS regions; GRINS regions with a positive TA skew.



**Fig. S3. Seven homologous PKSs encoding 16-membered macrolide synthases.** For each PKS (A-G), the PKS genes within the cluster and encoded protein modules and domains are shown. Domains: KS, ketosynthase; KS<sup>Q</sup>, decarboxylative ketosynthase; AT, acyltransferase; DH, dehydratase; DhT, inactive dehydratase; ER, enoyl-reductase; KR, ketoreductase; KR<sup>o</sup>, ketoreductase-inactive epimerase; ACP, acyl carrier protein; TE, thioesterase. For each module, the extender unit specificity of its AT domain is specified: mm, methylmalonyl extender; m, malonyl; em, ethylmalonyl; moxm, methoxymalonyl. Each color shows a different group of GRINS within a PKS (>80% DNA sequence identity within the group) and their locations. Molecules biosynthesized by these PKSs are shown to the right. Each color shows the differences in chemical structure, all of which can be attributed to the different extender unit incorporated by the GRINS-encoded AT domains. The bottom graph shows the GC and TA skews, as well as the maximum sequence similarity to another window in the same PKS, calculated for a sliding window of 150nt with a step of 30nt. Duplicated regions were identified by high sequence similarity (black line) and are shaded. Several regions also have intense and correlated nucleotide skews (red and blue lines), and therefore are annotated as GRINS (shaded in green).



**Fig. S4. GRINS in homologous 16-membered macrocyclic lactone synthase gene clusters.** Gene clusters encoding: (A) Milbemycin synthase, (B) avermectin synthase and (C) nemadectin synthases. These are homologous PKSs (overall DNA sequence identity >65%) that produce structurally similar molecules. Interestingly, GRINS present in modules 3, 4 and 5 of milbemycin synthase (red color, top) share >95% DNA sequence identity not only among each other, but also with GRINS present in modules 1, 2, 3, 4, 5 and 8 of avermectin synthase (red color, middle). Sequence identity between these GRINS in homologous modules of milbemycin and avermectin (e.g., between respective modules 3) reach 100% over regions over 1kb, suggesting that gene conversion of GRINS can occur not only between paralogous modules within one PKS, but also between orthologous modules of different PKSs. Although the polyketide products are similar, they have key differences (blue color, middle). While avermectins can have a single or a double bond between C22-C23, milbemycins are found exclusively with a single bond at that position (leading to greater stability in the latter family of compounds), which corresponds to a greater safety of these compounds. The hydroxyl substituent at C13 of avermectins is further glycosylated, which also affects the anthelmintic activity of these compounds. Both of these differences can be traced back to reductive loops of modules 2 and 7, which harbor GRINS in avermectin synthase but not in milbemycin synthase (blue arrows, middle), suggesting that by re-coding reductive domains of PKSs, GRINS can introduce structural changes that are important from the perspective of their biological activity.

**Table S5. GRINS found in PKS domains.** Domain names correspond to antiSMASH5.0 predictions.

| Domain type       | Total | GRINS | GRINS, | Domain type    | Total | GRINS | GRINS, % |
|-------------------|-------|-------|--------|----------------|-------|-------|----------|
| PKS_AT            | 17787 | 3555  | 20%    | Thioesterase   | 3372  | 3     | <1%      |
| PKS_DH            | 9108  | 1417  | 16%    | Aminotran_5    | 287   | 0     | 0%       |
| PKS_KS            | 24771 | 2636  | 11%    | GNAT           | 93    | 0     | 0%       |
| PKS_ER            | 3313  | 323   | 10%    | nMT            | 0     | 0     | 0%       |
| PKS_KR            | 19494 | 1491  | 8%     | Hal            | 5     | 0     | 0%       |
| PKS_Docking_Nterm | 3455  | 248   | 7%     | F              | 2     | 0     | 0%       |
| ACP               | 23328 | 1185  | 5%     | TD             | 265   | 0     | 0%       |
| PKS_DHt           | 811   | 31    | 4%     | oMT            | 0     | 0     | 0%       |
| AMP-binding       | 7190  | 206   | 3%     | B              | 22    | 0     | 0%       |
| Condensation      | 5744  | 147   | 3%     | ER             | 0     | 0     | 0%       |
| PKS_Docking_Cterm | 4200  | 95    | 2%     | PKS_ACP        | 0     | 0     | 0%       |
| ACP_beta          | 906   | 20    | 2%     | PS             | 53    | 0     | 0%       |
| PCP               | 7664  | 166   | 2%     | Polyketide_cyc | 299   | 0     | 0%       |
| Epimerization     | 883   | 17    | 2%     | cMT            | 0     | 0     | 0%       |
| PKS_DH2           | 3089  | 35    | 1%     | ACPS           | 631   | 0     | 0%       |
| NRPS-COM_Nterm    | 292   | 3     | 1%     | Aminotran_3    | 532   | 0     | 0%       |
| MT                | 2175  | 19    | 1%     | ECH            | 768   | 0     | 0%       |
| Trans-AT_docking  | 4378  | 35    | 1%     | NRPS-          | 82    | 0     | 0%       |
| Heterocyclization | 591   | 3     | <1%    | NAD_binding_4  | 481   | 0     | 0%       |
| FkbH              | 435   | 2     | <1%    | Aminotran_4    | 63    | 0     | 0%       |
| CAL_domain        | 781   | 2     | <1%    | Aminotran_1_2  | 647   | 0     | 0%       |

**Table S6. List of 300 high quality bacterial genomes.** Table with the list of the 300 genomes representing bacterial genera across five phyla. Column accession indicates the assembly accession from RefSeq. Column “organism\_name” contains the RefSeq official name. Column “phylum” represents the phylum to which the corresponding genome belongs.

| accession       | organism name   | phylum            |
|-----------------|---|-------------------|
| GCF_014646155.1 | Mangrovihabitans endophyticus                                   | p__Actinobacteria |
| GCF_013410525.1 | Galbitalea soli   | p__Actinobacteria |
| GCF_000428945.1 | Hamadaea tsunoensis DSM 44101                                   | p__Actinobacteria |
| GCF_001552435.1 | Kribbia dieselivorans NBRC 106261                               | p__Actinobacteria |
| GCF_003751265.1 | Pseudokineococcus lusitanus                                     | p__Actinobacteria |
| GCF_001552335.1 | Granulicoccus phenolivorans DSM 17626 = NBRC 107789 = JCM 15570 | p__Actinobacteria |
| GCF_900188375.1 | Geodermatophilus pulveris                                       | p__Actinobacteria |
| GCF_009758175.1 | Nesterenkonia alkaliphila                                       | p__Actinobacteria |
| GCF_014202315.1 | Thermocatellispora tengchongensis                               | p__Actinobacteria |
| GCF_008124835.1 | Blastococcus xanthinilyticus                                    | p__Actinobacteria |
| GCF_001652275.1 | Peptidiphaga gingivicola  | p__Actinobacteria |
| GCF_004345645.1 | Actinocrispum wychmicini  | p__Actinobacteria |
| GCF_900128965.1 | Ferrithrix thermotolerans DSM 19514                             | p__Actinobacteria |
| GCF_008017545.1 | Homoserinibacter sp. GY 40078                                   | p__Actinobacteria |
| GCF_007828725.1 | Marihabitans asiaticum  | p__Actinobacteria |
| GCF_000420025.1 | Patulibacter americanus DSM 16676                               | p__Actinobacteria |
| GCF_003148865.1 | Lechevalieria deserti   | p__Actinobacteria |
| GCF_002082605.1 | Kitasatospora aureofaciens                                      | p__Actinobacteria |
| GCF_009739655.1 | Clavibacter michiganensis subsp. michiganensis                  | p__Actinobacteria |
| GCF_000527155.1 | Haloglycomyces albus DSM 45210                                  | p__Actinobacteria |
| GCF_002563675.1 | Propionicimonas paludicola                                      | p__Actinobacteria |
| GCF_000214175.1 | Hoyosella subflava DQS3-9A1                                     | p__Actinobacteria |
| GCF_001515525.1 | Piscicoccus intestinalis NBRC 104926                            | p__Actinobacteria |
| GCF_003721155.3 | Janibacter melonis  | p__Actinobacteria |
| GCF_900115565.1 | Yuhushiella deserti   | p__Actinobacteria |
| GCF_006385015.1 | Brevibacterium linens   | p__Actinobacteria |
| GCF_007827045.1 | Trebonia kvetii   | p__Actinobacteria |
| GCF_000969705.1 | Nitriliruptor alkaliphilus DSM 45188                            | p__Actinobacteria |

|                 |  |                   |
|-----------------|--|-------------------|
| GCF_009687865.1 | <i>Acrocarpospora macrocephala</i>         | p__Actinobacteria |
| GCF_004564355.1 | <i>Microbacterium wangchenii</i>           | p__Actinobacteria |
| GCF_003336425.1 | <i>Marinitenerispora sediminis</i>         | p__Actinobacteria |
| GCF_003751805.1 | <i>Salana multivorans</i>                  | p__Actinobacteria |
| GCF_002355535.1 | <i>Aurantimicrobium minutum</i>            | p__Actinobacteria |
| GCF_004332295.1 | <i>Alloscardovia theropitheci</i>          | p__Actinobacteria |
| GCF_000755585.2 | <i>Arthrobacter alpinus</i>                | p__Actinobacteria |
| GCF_000612055.1 | <i>Trueperella pyogenes</i>                | p__Actinobacteria |
| GCF_004216535.1 | <i>Xylanibacterium ulmi</i>                | p__Actinobacteria |
| GCF_011758685.1 | <i>Auritidibacter ignavus</i>              | p__Actinobacteria |
| GCF_000024345.1 | <i>Kribbella flavida</i> DSM 17836         | p__Actinobacteria |
| GCF_004526345.1 | <i>Ornithinimicrobium flavum</i>           | p__Actinobacteria |
| GCF_007970665.1 | <i>Baekduia soli</i>                       | p__Actinobacteria |
| GCF_900119495.1 | <i>Arcanobacterium urinimassiliense</i>    | p__Actinobacteria |
| GCF_004078655.1 | <i>Labedella populi</i>                    | p__Actinobacteria |
| GCF_900637725.1 | <i>Pseudopropionibacterium propionicum</i> | p__Actinobacteria |
| GCF_003755125.1 | <i>Plantibacter flavus</i>                 | p__Actinobacteria |
| GCF_006716095.1 | <i>Barrientosiimonas humi</i>              | p__Actinobacteria |
| GCF_003634695.1 | <i>Motilibacter peucedani</i>              | p__Actinobacteria |
| GCF_007483665.1 | <i>Catellatospora sichuanensis</i>         | p__Actinobacteria |
| GCF_001552615.1 | <i>Millisia brevis</i> NBRC 105863         | p__Actinobacteria |
| GCF_900129375.1 | <i>Streptoalloteichus hindustanus</i>      | p__Actinobacteria |
| GCF_000974805.1 | <i>Demequina salsinemoris</i>              | p__Actinobacteria |
| GCF_000380485.1 | <i>Gordonia kroppenstedtii</i> DSM 45133   | p__Actinobacteria |
| GCF_001647635.1 | <i>Leifsonia xyli</i>                      | p__Actinobacteria |
| GCF_004801885.1 | <i>Naasia lichenicola</i>                  | p__Actinobacteria |
| GCF_900078545.1 | <i>Olegusella massiliensis</i>             | p__Actinobacteria |
| GCF_900104895.1 | <i>Microterricola viridarii</i>            | p__Actinobacteria |
| GCF_000312125.1 | <i>Timonella senegalensis</i> JC301        | p__Actinobacteria |
| GCF_004362825.1 | <i>Labedaea rhizosphaerae</i>              | p__Actinobacteria |
| GCF_000195955.2 | <i>Mycobacterium tuberculosis</i> H37Rv    | p__Actinobacteria |
| GCF_003253775.1 | <i>Mycobacterium leprae</i>                | p__Actinobacteria |
| GCF_900178525.1 | <i>Massilibacteroides vaginae</i>          | p__Bacteroidetes  |

|                 |  |                  |
|-----------------|--|------------------|
| GCF_001306415.1 | <i>Croceitalea dokdonensis</i> DOKDO 023       | p__Bacteroidetes |
| GCF_003991355.1 | <i>Pseudoflavitalea rhizosphaerae</i>          | p__Bacteroidetes |
| GCF_900112795.1 | <i>Thermophagus xiamenensis</i>                | p__Bacteroidetes |
| GCF_900100765.1 | <i>Catalinimonas alkaloidigena</i>             | p__Bacteroidetes |
| GCF_003204205.1 | <i>Sanguibacteroides justesenii</i>            | p__Bacteroidetes |
| GCF_000688335.1 | <i>Sediminibacter</i> sp. Hel_I_10             | p__Bacteroidetes |
| GCF_003353425.1 | <i>Marinirhabdus gelatinilytica</i>            | p__Bacteroidetes |
| GCF_001483135.1 | <i>Solirubrum puertoriconensis</i>             | p__Bacteroidetes |
| GCF_014596935.1 | <i>Aestuariibaculum suncheonense</i>           | p__Bacteroidetes |
| GCF_900101815.1 | <i>Pricia antarctica</i>                       | p__Bacteroidetes |
| GCF_007489275.1 | <i>Aquimarina algiphila</i>                    | p__Bacteroidetes |
| GCF_003413745.1 | <i>Marixanthomonas ophiurae</i>                | p__Bacteroidetes |
| GCF_900113045.1 | <i>Thermoflexibacter ruber</i>                 | p__Bacteroidetes |
| GCF_003667275.1 | <i>Mesonina aquimarina</i>                     | p__Bacteroidetes |
| GCF_900112255.1 | <i>Flexibacter flexilis</i> DSM 6793           | p__Bacteroidetes |
| GCF_003347595.2 | <i>Botryobacter ruber</i>                      | p__Bacteroidetes |
| GCF_007993035.1 | <i>Vicingus serpentipes</i>                    | p__Bacteroidetes |
| GCF_002886045.1 | <i>Flavivirga eckloniae</i>                    | p__Bacteroidetes |
| GCF_002557795.1 | <i>Chitinophaga caeni</i>                      | p__Bacteroidetes |
| GCF_010686655.1 | <i>Cryomorpha ignava</i>                       | p__Bacteroidetes |
| GCF_000250635.1 | <i>Saprospira grandis</i> str. Lewin           | p__Bacteroidetes |
| GCF_001653755.1 | <i>Dokdonia donghaensis</i> DSW-1              | p__Bacteroidetes |
| GCF_900292045.1 | <i>Parabacteroides pacaensis</i>               | p__Bacteroidetes |
| GCF_002943105.1 | <i>Cloacibacterium normanense</i>              | p__Bacteroidetes |
| GCF_900096565.1 | <i>Williamwhitmania taraxaci</i>               | p__Bacteroidetes |
| GCF_000265405.1 | <i>Belliella baltica</i> DSM 15883             | p__Bacteroidetes |
| GCF_003347495.1 | <i>Negadavirga</i> sp. SW125                   | p__Bacteroidetes |
| GCF_011762485.1 | <i>Yeosuana</i> sp. JLT21                      | p__Bacteroidetes |
| GCF_010671605.1 | <i>Leptobacterium flavescens</i>               | p__Bacteroidetes |
| GCF_000260835.1 | <i>Imtechella halotolerans</i> K1              | p__Bacteroidetes |
| GCF_007097485.1 | <i>Carboxylicivirga</i> sp. M1479              | p__Bacteroidetes |
| GCF_002807015.1 | <i>Avrilella dinanensis</i>                    | p__Bacteroidetes |
| GCF_000511175.1 | <i>Sediminibacterium salmoneum</i> NBRC 103935 | p__Bacteroidetes |

|                 |   |                  |
|-----------------|---|------------------|
| GCF_900141715.1 | <i>Arenitalea lutea</i>                             | p__Bacteroidetes |
| GCF_002998435.1 | <i>Bacteroides zoogloiformans</i>                   | p__Bacteroidetes |
| GCF_000224085.1 | <i>Muricauda ruestringensis</i> DSM 13258           | p__Bacteroidetes |
| GCF_000422585.1 | <i>Hugenholtzia roseola</i> DSM 9546                | p__Bacteroidetes |
| GCF_000325705.1 | <i>Echinicola vietnamensis</i> DSM 17526            | p__Bacteroidetes |
| GCF_002770595.1 | <i>Chryseobacterium camelliae</i>                   | p__Bacteroidetes |
| GCF_013409575.1 | <i>Macellibacteroides fermentans</i>                | p__Bacteroidetes |
| GCF_003858635.1 | <i>Larkinella rosea</i>                             | p__Bacteroidetes |
| GCF_004118155.1 | <i>Edaphocola flava</i>                             | p__Bacteroidetes |
| GCF_002210225.1 | <i>Geofilum rhodophaeum</i>                         | p__Bacteroidetes |
| GCF_003003005.1 | <i>Mongoliibacter ruber</i>                         | p__Bacteroidetes |
| GCF_007970805.1 | <i>Flavisolibacter ginsenosidimutans</i>            | p__Bacteroidetes |
| GCF_001543325.1 | <i>Lutibacter profundus</i>                         | p__Bacteroidetes |
| GCF_003337435.1 | <i>Schleiferia thermophila</i>                      | p__Bacteroidetes |
| GCF_900167975.1 | <i>Ohtaekwangia koreensis</i>                       | p__Bacteroidetes |
| GCF_003313335.1 | <i>Nubsella zeaxanthinifaciens</i>                  | p__Bacteroidetes |
| GCF_014652135.1 | <i>Mongoliitalea lutea</i>                          | p__Bacteroidetes |
| GCF_900115015.1 | <i>Algoriella xinjiangensis</i>                     | p__Bacteroidetes |
| GCF_006542645.1 | <i>Alistipes onderdonkii</i> subsp. <i>vulgaris</i> | p__Bacteroidetes |
| GCF_007971025.1 | <i>Mucilaginibacter ginsenosidivorans</i>           | p__Bacteroidetes |
| GCF_001870735.1 | <i>Arsenicibacter rosenii</i>                       | p__Bacteroidetes |
| GCF_014642875.1 | <i>Puia dinghuensis</i>                             | p__Bacteroidetes |
| GCF_000473765.1 | <i>Rhodonellum psychrophilum</i> GCM71 = DSM 17998  | p__Bacteroidetes |
| GCF_000714815.1 | <i>Anditalea andensis</i>                           | p__Bacteroidetes |
| GCF_004115975.1 | <i>Gelidibacter gilvus</i>                          | p__Bacteroidetes |
| GCF_900115775.1 | <i>Siccationidurans arizonensis</i>                 | p__Bacteroidetes |
| GCF_014698965.1 | <i>Iningainema</i> sp. BLCCT55                      | p__Cyanobacteria |
| GCF_002252705.1 | <i>Vulcanococcus limneticus</i> LL                  | p__Cyanobacteria |
| GCF_000214075.1 | <i>Microcoleus vaginatus</i> FGP-2                  | p__Cyanobacteria |
| GCF_000314005.1 | <i>Spirulina subsalsa</i> PCC 9445                  | p__Cyanobacteria |
| GCF_002964865.1 | <i>Gloeocapsopsis</i> sp. AAB1 = 1H9                | p__Cyanobacteria |
| GCF_000332235.1 | <i>Geminocystis herdmannii</i> PCC 6308             | p__Cyanobacteria |
| GCF_000317285.1 | <i>Chlorogloeopsis fritschii</i> PCC 6912           | p__Cyanobacteria |



|                 |  |                  |
|-----------------|--|------------------|
| GCF_900659865.1 | Hyella patelloides LEGE 07179                        | p__Cyanobacteria |
| GCF_000021825.1 | Gloeotheca citriformis PCC 7424                      | p__Cyanobacteria |
| GCF_000317615.1 | Dactylococcopsis salina PCC 8305                     | p__Cyanobacteria |
| GCF_000317555.1 | Gloeocapsa sp. PCC 7428                              | p__Cyanobacteria |
| GCF_000817745.2 | Aphanocapsa montana BDHKU210001                      | p__Cyanobacteria |
| GCF_000332055.1 | Xenococcus sp. PCC 7305                              | p__Cyanobacteria |
| GCF_000973065.1 | Limnospira indica PCC 8005                           | p__Cyanobacteria |
| GCF_000775285.1 | Neosynechococcus sphagnicola sy1                     | p__Cyanobacteria |
| GCF_000317475.1 | Oscillatoria nigro-viridis PCC 7112                  | p__Cyanobacteria |
| GCF_001746915.1 | Desertifilum sp. IPPAS B-1220                        | p__Cyanobacteria |
| GCF_000175835.1 | Cylindrospermopsis raciborskii CS-505                | p__Cyanobacteria |
| GCF_900010725.2 | Planktothrixserta PCC 8927                           | p__Cyanobacteria |
| GCF_001904615.1 | Limnothrix rosea IAM M-220                           | p__Cyanobacteria |
| GCF_002368275.1 | Fremyella diplosiphon NIES-3275                      | p__Cyanobacteria |
| GCF_014696015.1 | Oculatella sp. FACHB-28                              | p__Cyanobacteria |
| GCF_000316515.1 | Cyanobium gracile PCC 6307                           | p__Cyanobacteria |
| GCF_000317205.1 | Fischerella muscicola PCC 7414                       | p__Cyanobacteria |
| GCF_014698035.1 | Nostoc muscorum FACHB-395                            | p__Cyanobacteria |
| GCF_003004015.1 | Cyanosarcina cf. burmensis CICALA 770                | p__Cyanobacteria |
| GCF_000756305.1 | Myxosarcina sp. GI1                                  | p__Cyanobacteria |
| GCF_001767235.1 | Moorea producens PAL-8-15-08-1                       | p__Cyanobacteria |
| GCF_012295525.1 | Oxynema sp. AP17                                     | p__Cyanobacteria |
| GCF_001275395.1 | Hapalosiphon sp. MRB220                              | p__Cyanobacteria |
| GCF_000332155.1 | Kamptonema formosum PCC 6407                         | p__Cyanobacteria |
| GCF_000007925.1 | Prochlorococcus marinus subsp. marinus str. CCMP1375 | p__Cyanobacteria |
| GCF_002368355.1 | Chondrocystis sp. NIES-4102                          | p__Cyanobacteria |
| GCF_000011385.1 | Gloeobacter violaceus PCC 7421                       | p__Cyanobacteria |
| GCF_014697025.1 | Alkalinema sp. FACHB-956                             | p__Cyanobacteria |
| GCF_000332315.1 | Prochlorothrix hollandica PCC 9006 = CALU 1027       | p__Cyanobacteria |
| GCF_003003695.1 | Phormidesmis priestleyi ULC007                       | p__Cyanobacteria |
| GCF_000175855.1 | Raphidiopsis brookii D9                              | p__Cyanobacteria |
| GCF_007904085.1 | Euhalothece natronophila Z-M001                      | p__Cyanobacteria |
| GCF_002367955.1 | Anabaena cylindrica PCC 7122                         | p__Cyanobacteria |

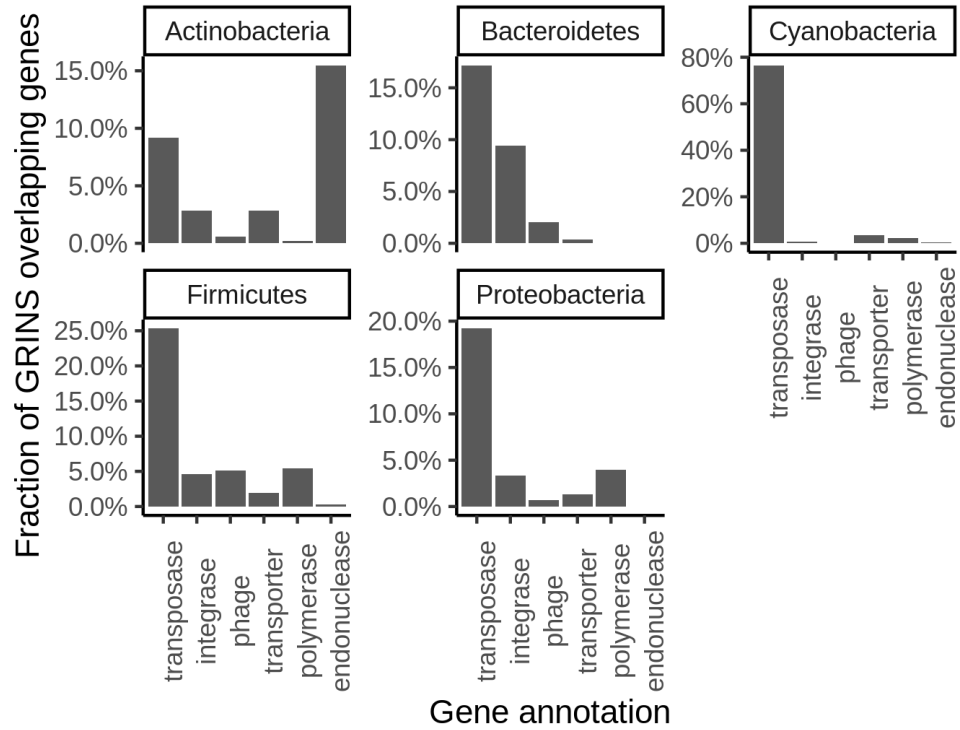
|                 |  |                  |
|-----------------|--|------------------|
| GCF_000317025.1 | Pleurocapsa sp. PCC 7327                                 | p__Cyanobacteria |
| GCF_002368255.1 | Leptolyngbya boryana NIES-2135                           | p__Cyanobacteria |
| GCF_013177315.1 | Thermoleptolyngbya sp. PKUAC-SCTA183                     | p__Cyanobacteria |
| GCF_005402885.1 | Sphaerospermopsis reniformis                             | p__Cyanobacteria |
| GCF_002368055.1 | Aulosira laxa NIES-50                                    | p__Cyanobacteria |
| GCF_000817785.2 | Hassallia byssoidea VB512170                             | p__Cyanobacteria |
| GCF_000521175.1 | Aphanizomenon flos-aquae NIES-81                         | p__Cyanobacteria |
| GCF_000022045.1 | Cyanothece sp. PCC 7425                                  | p__Cyanobacteria |
| GCF_003003795.1 | Stenomitros frigidus ULC18                               | p__Cyanobacteria |
| GCF_000024045.1 | Rippkaea orientalis PCC 8802                             | p__Cyanobacteria |
| GCF_000828075.3 | Tolypothrix campylonemoides VB511288                     | p__Cyanobacteria |
| GCF_000155555.1 | Coleofasciculus chthonoplastes PCC 7420                  | p__Cyanobacteria |
| GCF_004323185.1 | Westiellopsis prolifica IICB1                            | p__Cyanobacteria |
| GCF_003443655.1 | Nostoc sphaeroides                                       | p__Cyanobacteria |
| GCF_001264245.1 | Microcystis panniformis FACHB-1757                       | p__Cyanobacteria |
| GCF_000340565.2 | Nodularia spumigena CCY9414                              | p__Cyanobacteria |
| GCF_001870225.1 | Gloeomargarita lithophora Alchichica-D10                 | p__Cyanobacteria |
| GCF_014695385.1 | Trichocoleus sp. FACHB-46                                | p__Cyanobacteria |
| GCF_009176225.1 | Arthrospira platensis NIES-46                            | p__Cyanobacteria |
| GCF_010222815.1 | Synechocystis sp.  | p__Cyanobacteria |
| GCF_002844395.1 | Hungateiclostridium saccincola                           | p__Firmicutes    |
| GCF_003925875.1 | Intestinibaculum porci                                   | p__Firmicutes    |
| GCF_000305935.1 | Thermacetogenium phaeum DSM 12270                        | p__Firmicutes    |
| GCF_900112085.1 | Acetitomaculum ruminis DSM 5522                          | p__Firmicutes    |
| GCF_000014725.1 | Syntrophomonas wolfei subsp. wolfei str. Goettingen G311 | p__Firmicutes    |
| GCF_005280655.1 | Anaerostipes rhamnosivorans                              | p__Firmicutes    |
| GCF_006542275.1 | Saccharibacillus brassicae                               | p__Firmicutes    |
| GCF_000377705.1 | Salsuginibacillus kocurii DSM 18087                      | p__Firmicutes    |
| GCF_900605495.1 | Bacilliculturomica massiliensis                          | p__Firmicutes    |
| GCF_004363045.1 | Aureibacillus halotolerans                               | p__Firmicutes    |
| GCF_000948185.1 | Aeribacillus pallidus                                    | p__Firmicutes    |
| GCF_008033155.1 | Lactobacillus malefermentans                             | p__Firmicutes    |
| GCF_900167305.1 | Garciella nitratireducens DSM 15102                      | p__Firmicutes    |

|                 |  |               |
|-----------------|--|---------------|
| GCF_003350505.1 | <i>Falsibacillus pallidus</i>  | p__Firmicutes |
| GCF_904377855.1 | <i>Candidatus Avimonas merdigallarum</i>                             | p__Firmicutes |
| GCF_000376385.1 | <i>Desulfurispora thermophila</i> DSM 16022                          | p__Firmicutes |
| GCF_000143685.1 | <i>Clostridium Ijungdahlii</i> DSM 13528                             | p__Firmicutes |
| GCF_001280565.1 | <i>Sulfobacillus thermosulfidooxidans</i>                            | p__Firmicutes |
| GCF_902362575.1 | <i>Catenibacterium mitsuokai</i>                                     | p__Firmicutes |
| GCF_014637175.1 | <i>Compostibacillus humi</i>   | p__Firmicutes |
| GCF_000269565.1 | <i>Coprobacillus cateniformis</i>                                    | p__Firmicutes |
| GCF_000312505.2 | <i>Fenollaria massiliensis</i>                                       | p__Firmicutes |
| GCF_900142005.1 | <i>Dethiosulfatibacter aminovorans</i> DSM 17477                     | p__Firmicutes |
| GCF_000196455.1 | <i>Acetoanaerobium sticklandii</i>                                   | p__Firmicutes |
| GCF_001261775.1 | <i>Anaeromassilibacillus senegalensis</i>                            | p__Firmicutes |
| GCF_002160865.1 | <i>Massilimicrobiota timonensis</i>                                  | p__Firmicutes |
| GCF_009720735.1 | <i>Heliobacillus mobilis</i>   | p__Firmicutes |
| GCF_000213255.1 | <i>Mahella australiensis</i> 50-1 BON                                | p__Firmicutes |
| GCF_000007085.1 | <i>Caldanaerobacter subterraneus</i> subsp. <i>tengcongensis</i> MB4 | p__Firmicutes |
| GCF_003317055.1 | <i>Alkalibaculum bacchi</i>  | p__Firmicutes |
| GCF_900142995.1 | <i>Caldanaerovirga acetigignens</i>                                  | p__Firmicutes |
| GCF_900104675.1 | <i>Angelakisella massiliensis</i>                                    | p__Firmicutes |
| GCF_900015005.1 | <i>Lactobacillus oligofermentans</i> DSM 15707 = LMG 22743           | p__Firmicutes |
| GCF_002998925.1 | <i>Mogibacterium diversum</i>  | p__Firmicutes |
| GCF_002994005.1 | <i>Lactobacillus quenuiae</i>  | p__Firmicutes |
| GCF_009905255.1 | <i>Thermoanaerobacterium aotearoense</i>                             | p__Firmicutes |
| GCF_013127755.1 | <i>Pediococcus acidilactici</i>                                      | p__Firmicutes |
| GCF_004341685.1 | <i>Pectinatus cerevisiiphilus</i>                                    | p__Firmicutes |
| GCF_900243045.1 | <i>Eisenbergiella massiliensis</i>                                   | p__Firmicutes |
| GCF_008694105.1 | <i>Lactobacillus harbinensis</i>                                     | p__Firmicutes |
| GCF_000283575.1 | <i>Oscillibacter valericigenes</i> Sjm18-20                          | p__Firmicutes |
| GCF_014639255.1 | <i>Pullulanibacillus pueri</i>                                       | p__Firmicutes |
| GCF_900156305.1 | <i>Edaphobacillus lindanitolerans</i>                                | p__Firmicutes |
| GCF_001870205.1 | <i>Vagococcus teuberi</i>  | p__Firmicutes |
| GCF_007559355.1 | <i>Allobacillus</i> sp. SKP2-8                                       | p__Firmicutes |
| GCF_003001905.1 | <i>Planifilum fimeticola</i>   | p__Firmicutes |

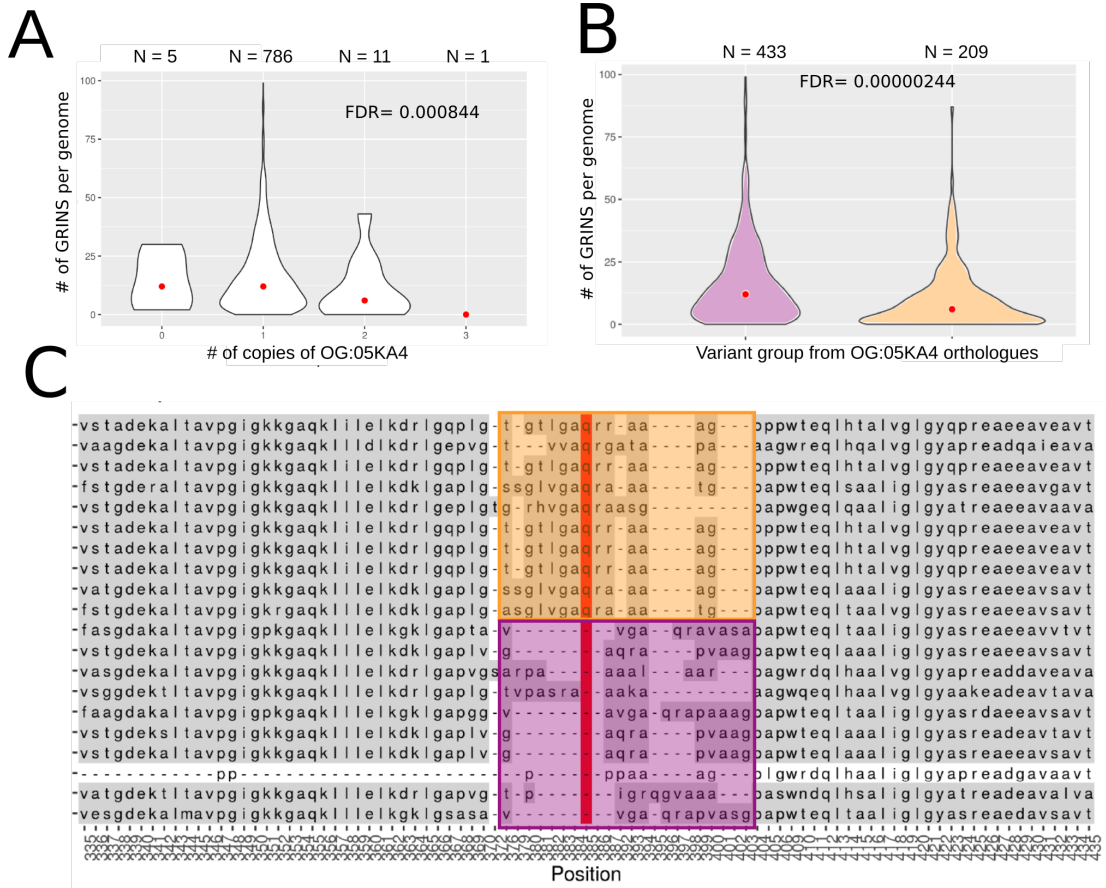
|                 |  |                   |
|-----------------|--|-------------------|
| GCF_000452245.2 | <i>Paraclostridium bifermentans</i> ATCC 638   | p__Firmicutes     |
| GCF_002207765.1 | <i>Calderihabitans maritimus</i>               | p__Firmicutes     |
| GCF_009734445.1 | Candidatus <i>Syntrophocurvum alkaliphilum</i> | p__Firmicutes     |
| GCF_900086705.1 | <i>Massilibacillus massiliensis</i>            | p__Firmicutes     |
| GCF_000420725.1 | <i>Marinococcus halotolerans</i> DSM 16375     | p__Firmicutes     |
| GCF_004364155.1 | <i>Fonticella tunisiensis</i>                  | p__Firmicutes     |
| GCF_900102435.1 | <i>Sporolituus thermophilus</i> DSM 23256      | p__Firmicutes     |
| GCF_000702205.1 | <i>Lachnospira multipara</i> LB2003            | p__Firmicutes     |
| GCF_000829905.1 | <i>Beduini massiliensis</i>                    | p__Firmicutes     |
| GCF_900079135.1 | <i>Trichococcus pasteurii</i>                  | p__Firmicutes     |
| GCF_000517385.1 | <i>Bacillus boroniphilus</i> JCM 21738         | p__Firmicutes     |
| GCF_001562415.1 | <i>Thermotalea metallivorans</i>               | p__Firmicutes     |
| GCF_000020005.1 | <i>Natranaerobius thermophilus</i> JW/NM-WN-LF | p__Firmicutes     |
| GCF_002871685.1 | <i>Dolosicoccus paucivorans</i>                | p__Firmicutes     |
| GCF_000350545.1 | <i>Desulfotignum phosphitoxidans</i> DSM 13687 | p__Proteobacteria |
| GCF_000376545.2 | <i>Epibacterium mobile</i> F1926               | p__Proteobacteria |
| GCF_000021885.1 | <i>Glaesserella parasuis</i> SH0165            | p__Proteobacteria |
| GCF_000055785.1 | <i>Chromohalobacter salexigens</i> DSM 3043    | p__Proteobacteria |
| GCF_009388985.1 | <i>Halioglobus maricola</i>                    | p__Proteobacteria |
| GCF_002356555.2 | <i>Halorhodospira halochloris</i>              | p__Proteobacteria |
| GCF_004343305.1 | <i>Pseudofulvimonas gallinarii</i>             | p__Proteobacteria |
| GCF_900044015.1 | Candidatus <i>Hoaglandella endobia</i>         | p__Proteobacteria |
| GCF_001720485.1 | <i>Cobetia marina</i>                          | p__Proteobacteria |
| GCF_009746125.1 | <i>Casimicrobium huifangae</i>                 | p__Proteobacteria |
| GCF_009761375.1 | <i>Ursidibacter arcticus</i>                   | p__Proteobacteria |
| GCF_004363315.1 | <i>Tepidicella xavieri</i>                     | p__Proteobacteria |
| GCF_900101365.1 | <i>Thiohalorhabdus denitrificans</i>           | p__Proteobacteria |
| GCF_003433515.1 | <i>Maritalea myrionectae</i>                   | p__Proteobacteria |
| GCF_000953015.1 | Candidatus <i>Methylopumilus turicensis</i>    | p__Proteobacteria |
| GCF_003852895.1 | <i>Geomonas soli</i>                           | p__Proteobacteria |
| GCF_001558695.1 | <i>Paramesorhizobium deserti</i>               | p__Proteobacteria |
| GCF_000054005.1 | <i>Brucella abortus</i> 2308                   | p__Proteobacteria |
| GCF_009792355.1 | <i>Terricaulis silvestris</i>                  | p__Proteobacteria |

|                 |   |                   |
|-----------------|---|-------------------|
| GCF_000219045.1 | <i>Francisella salina</i>                         | p__Proteobacteria |
| GCF_000019225.1 | <i>Cellvibrio japonicus</i> Ueda107               | p__Proteobacteria |
| GCF_000953695.1 | <i>Aliivibrio wodanis</i>                         | p__Proteobacteria |
| GCF_900089455.2 | <i>Orrella dioscoreae</i>                         | p__Proteobacteria |
| GCF_004346035.1 | <i>Acidomonas methanolica</i>                     | p__Proteobacteria |
| GCF_000968135.1 | <i>Magnetospira</i> sp. QH-2                      | p__Proteobacteria |
| GCF_002976435.1 | <i>Solimicrobium silvestre</i>                    | p__Proteobacteria |
| GCF_000384395.1 | <i>Geopsychrobacter electrodiphilus</i> DSM 16401 | p__Proteobacteria |
| GCF_000378465.1 | <i>Elioraea tepidiphila</i> DSM 17972             | p__Proteobacteria |
| GCF_000006925.2 | <i>Shigella flexneri</i> 2a str. 301              | p__Proteobacteria |
| GCF_012295595.1 | <i>Thalassobius gelatinovorus</i>                 | p__Proteobacteria |
| GCF_002877015.1 | <i>Enterovibrio norvegicus</i>                    | p__Proteobacteria |
| GCF_002117105.1 | <i>Oceanicoccus sagamiensis</i>                   | p__Proteobacteria |
| GCF_009900765.1 | <i>Rhodovarius lipocyclicus</i>                   | p__Proteobacteria |
| GCF_000420045.1 | <i>Perlucidibaca piscinae</i> DSM 21586           | p__Proteobacteria |
| GCF_001648175.1 | <i>Halotalea alkalilenta</i>                      | p__Proteobacteria |
| GCF_001543305.1 | <i>Liberibacter crescens</i>                      | p__Proteobacteria |
| GCF_003576595.1 | <i>Pusillimonas maritima</i>                      | p__Proteobacteria |
| GCF_002952735.2 | <i>Lelliottia nimipressuralis</i>                 | p__Proteobacteria |
| GCF_000813705.1 | <i>Morococcus cerebrosus</i>                      | p__Proteobacteria |
| GCF_900184825.1 | <i>Tropicibacter phthalicus</i>                   | p__Proteobacteria |
| GCF_000374525.1 | <i>Amorphus coralli</i> DSM 19760                 | p__Proteobacteria |
| GCF_000739695.1 | <i>Tepidicaulis marinus</i>                       | p__Proteobacteria |
| GCF_000314675.2 | <i>Afipia broomeae</i> ATCC 49717                 | p__Proteobacteria |
| GCF_000021905.1 | <i>Desulfatibacillum aliphaticivorans</i>         | p__Proteobacteria |
| GCF_000154705.2 | <i>Hoeflea phototrophica</i> DFL-43               | p__Proteobacteria |
| GCF_000299575.1 | <i>Oceaniovalibus guishaninsula</i> JLT2003       | p__Proteobacteria |
| GCF_000828475.1 | <i>Methyloceanibacter caenitepidi</i>             | p__Proteobacteria |
| GCF_000155735.2 | <i>Octadecabacter arcticus</i> 238                | p__Proteobacteria |
| GCF_900115595.1 | <i>Tranquillimonas alkanivorans</i>               | p__Proteobacteria |
| GCF_002105465.1 | <i>Aquidulcibacter paucihalophilus</i>            | p__Proteobacteria |
| GCF_000169415.1 | <i>Sagittula stellata</i> E-37                    | p__Proteobacteria |
| GCF_009903735.1 | <i>Agaribacter marinus</i>                        | p__Proteobacteria |

|                 |                                     |                   |
|-----------------|-------------------------------------|-------------------|
| GCF_900114705.1 | Rugamonas rubra                     | p__Proteobacteria |
| GCF_000176355.1 | Citromicrobium bathyomarinum JL354  | p__Proteobacteria |
| GCF_900174585.1 | Parendozoicomonas haliclona         | p__Proteobacteria |
| GCF_003350345.1 | Alkalilacustris brevis              | p__Proteobacteria |
| GCF_900100005.1 | Lutimaribacter saemankumensis       | p__Proteobacteria |
| GCF_000006765.1 | Pseudomonas aeruginosa PAO1         | p__Proteobacteria |
| GCF_000008865.2 | Escherichia coli O157:H7 str. Sakai | p__Proteobacteria |
| GCF_000767075.1 | Haemophilus influenzae              | p__Proteobacteria |



**Fig. S7. GRINS are widespread in transposases.** Each panel shows the proportion of GRINS within a phylum that overlaps with genes annotated to one of the categories in the x-axis. The annotations are from eggNOG v4.5 (Methods).



**Fig. S8. RuvA is inversely associated with the number of GRINS per genome. A.** Distributions of the number of GRINS per genome according to the number of copies of genes in orthologous group OG:05KA4 (eggNOG 4.5), which corresponds to the RuvA protein. The association is significant after controlling for phylogenetic relatedness and genome quality (Methods). Numbers at the top indicate the number of genomes in each violin, red dots show the median, and the width of the violin is proportional to the empirical density. **B.** After eliminating paralogues (Methods), two amino acid sequence variants around position 385 of the RuvA multiple sequence alignment are associated with significant differences in the number of GRINS between genomes. Numbers at the top indicate the number of genomes in each violin, red dots show the median, and the width of the violin is proportional to the empirical density. **C.** Representative sequences from multiple sequence alignment of the RuvA orthologues around position 385 (highlighted in red). The colors match the two groups in panel B and correspond to two general architectures of this region of the protein.