## **Supporting Information**

## Wang et al. 10.1073/pnas.1401734111

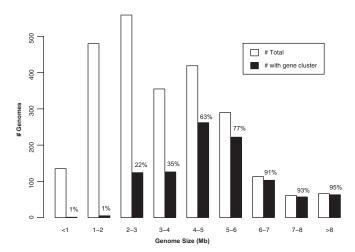


Fig. S1. Distribution of bacterial strains found with gene clusters among different genome-size ranges. The gene clusters are nearly absent from those strains with a genome size less than 2 Mb.

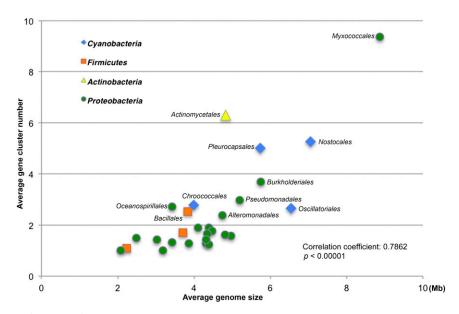


Fig. S2. Significant correlation (P < 0.00001) between average cluster number and average genome size in bacterial orders with enriched NRPSs and PKSs. The names are displayed for the orders whose average cluster number is over two per genome.

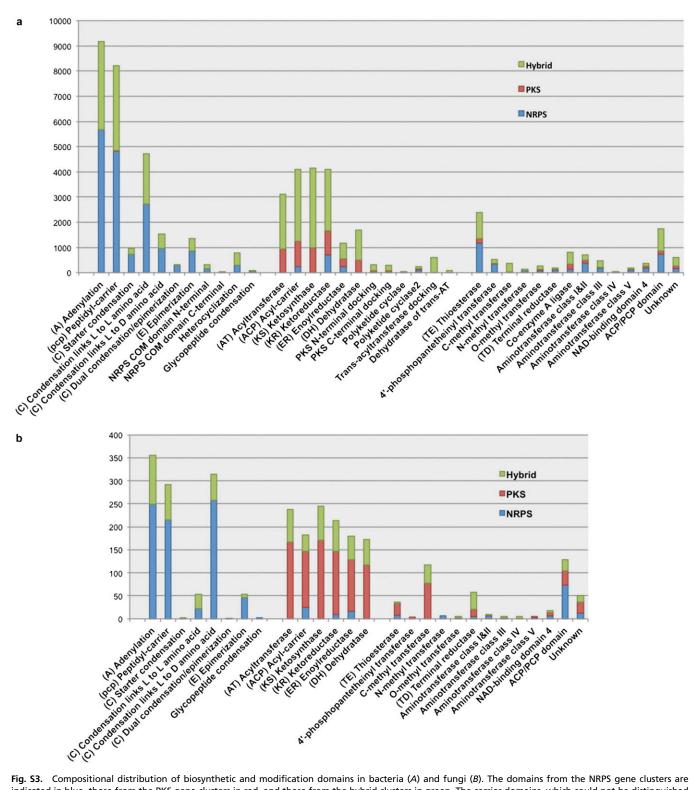


Fig. S3. Compositional distribution of biosynthetic and modification domains in bacteria (A) and fungi (B). The domains from the NRPS gene clusters are indicated in blue, those from the PKS gene clusters in red, and those from the hybrid clusters in green. The carrier domains, which could not be distinguished from linking peptide or acyl acid by the software packages, were named as the ACP/PCP type.

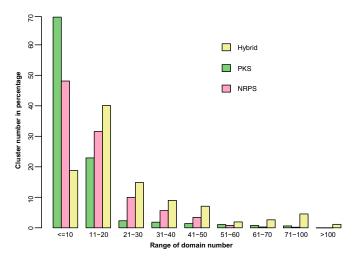


Fig. S4. Distribution of NRPS, PKS, and hybrid gene cluster in different ranges of domain numbers.

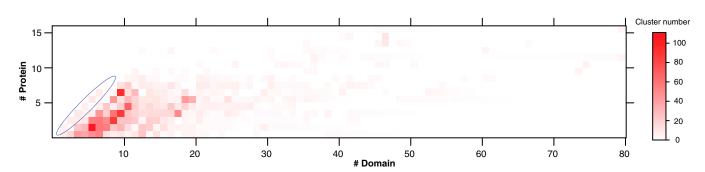


Fig. S5. Compositional distribution of gene clusters according to the protein and domain numbers. The blue circle labels the clusters comprised of discrete proteins.

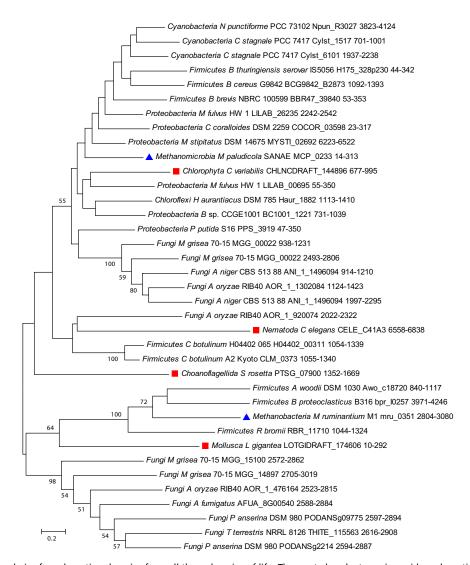


Fig. S6. Phylogenetic analysis of condensation domains from all three domains of life. The most abundant ι-amino acid condensation domains were used in making this tree. The sequences from archaea and eukarya (excluding fungi) were used as queries to search against bacterial and fungal sequences by blastP. Top hits in bacteria and fungi were used together with the query sequences for tree building (as described in *Materials and Methods*). The name of each taxon is constituted by lineage, organism, protein locus tag, and domain coordination in the protein. Branches from archaea are indicated by a blue triangle; those from eukarya (other than fungi) are indicated by a red square.

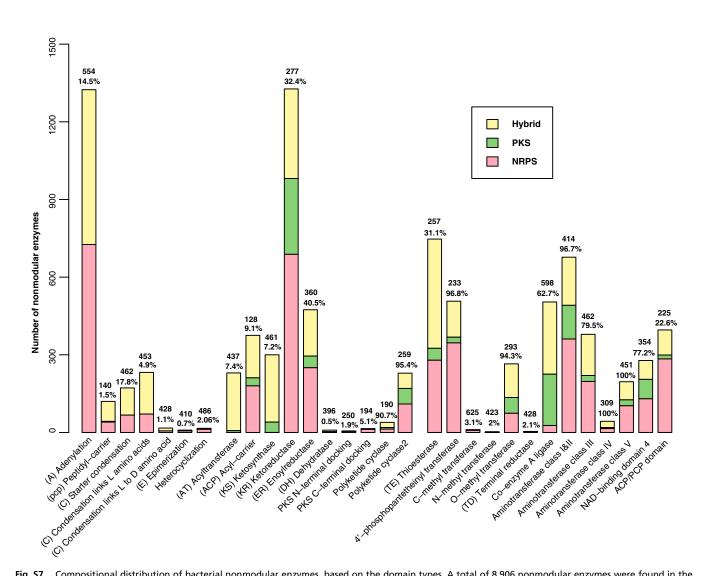


Fig. S7. Compositional distribution of bacterial nonmodular enzymes, based on the domain types. A total of 8,906 nonmodular enzymes were found in the bacterial gene clusters. They were counted according to their domain types. The average enzyme sizes (aa) and the domain frequencies (%) of the nonmodular enzymes are shown above the corresponding bars. The enzymes from NRPS gene clusters are indicated in red, those from PKS gene clusters in green, and those from the hybrid clusters in yellow. Nearly every domain type was found as a nonmodular enzyme. Some aminotransferase and polyketide cyclase domains were found almost exclusively as separate proteins. Many of the nonmodular enzymes are also present in hybrid clusters.

Table S1. Summary of bacterial NRPS and PKS gene clusters

	No. of strains		Avg genome	No. o	No. of	No. of			
Phylum	Total	With clusters (%)	size, Mb	Total (plasmid*)	NRPS	PKS	Hybrid	proteins	domains
Proteobacteria	1,098	519 (47%)	3.93	1,167 (59)	701	103	363	5,905	19,769
Firmicutes	573	178 (31%)	3.11	396 (14)	279	5	112	2,202	9,293
Actinobacteria	271	197 (73%)	4.44	1,225 (14)	381	319	525	6,742	23,627
Tenericutes	77	0 (0)	0.90	0	0	0	0	0	0
Cyanobacteria	72	37 (51%)	4.55	136 (14)	51	32	53	763	2,447
Bacteroidetes	96	13 (14%)	3.60	15 (1)	4	4	7	95	432
Chlamydiae	103	1 (1%)	1.23	1	1	0	0	3	14
Spirochaetes	57	0 (0)	2.31	0	0	0	0	0	0
Deinococcus-Thermus	19	1 (5%)	2.97	1	1	0	0	6	6
Chloroflexi	18	1 (6%)	3.37	9	3	1	5	67	284
Thermotogae	15	0 (0)	2.08	0	0	0	0	0	0
Chlorobi	11	1 (9%)	2.62	1	1	0	0	3	8
Aquificae	11	0 (0)	1.66	0	0	0	0	0	0
Acidobacteria	8	4 (50%)	5.69	7	2	2	3	33	126
Fusobacteria	8	0 (0)	2.31	0	0	0	0	0	0
Planctomycetes	7	5 (71%)	6.28	10	0	5	5	29	107
Synergistetes	5	0 (0)	2.14	0	0	0	0	0	0
Verrucomicrobia	4	1 (25%)	3.66	1	0	0	1	14	67
Nitrospirae	4	1 (25%)	2.82	2	1	0	1	12	51
Deferribacteres	4	0 (0)	2.63	0	0	0	0	0	0
Ignavibacteria	2	1 (50%)	3.48	1	0	0	1	4	12
Elusimicrobia	2	1 (50%)	1.97	2	1	1	0	6	16
Thermodesulfobacteria	2	0 (0)	1.98	0	0	0	0	0	0
Dictyoglomi	2	0 (0)	1.91	0	0	0	0	0	0
Gemmatimonadetes	1	1 (100%)	4.64	1	1	0	0	4	7
Fibrobacteres	1	1 (100%)	3.84	1	1	0	0	1	3
Armatimonadetes	1	0 (0)	3.44	0	0	0	0	0	0
Thermobaculum	1	0 (0)	3.10	0	0	0	0	0	0
Chrysiogenetes	1	0 (0)	2.93	0	0	0	0	0	0
Caldiserica	1	0 (0)	1.56	0	0	0	0	0	0
Saccharobacteria	1	0 (0)	1.01	0	0	0	0	0	0

<sup>\*</sup>Numbers of gene clusters found in plasmids.

Table S2. Distribution of NRPS and PKS gene clusters among classes within the phylum of Ascomycota in fungi

	No. of strains		Avg genome	No. of	No. of gene clusters			No. of	No. of
Class	Total	With clusters (%)	size, Mb	clusters	NRPS	PKS	Hybrid	proteins	domains
Saccharomycetes	18	0 (0)	11.38	0	0	0	0	0	0
Sordariomycetes	6	6 (100%)	38.07	127	33	64	30	215	1,143
Eurotiomycetes	5	5 (100%)	32.48	179	57	86	36	386	1,601
Schizosaccharomycetes	1	1 (100%)	12.59	1	1	0	0	1	12

Table S3. Organisms used for construction of the rRNA gene tree in Fig. 1

	Organisms used for constitut	g u.eg
Domain	Phylum/Class	Organism
Bacteria	Proteobacteria	Proteus mirabilis HI4320
Bacteria	Firmicutes	Acetobacterium woodii DSM 1030
Bacteria	Actinobacteria	Actinoplanes missouriensis 431
Bacteria	Tenericutes	Aster yellows witches'-broom phytoplasma AYWB
Bacteria	Cyanobacteria	Anabaena sp. 90
Bacteria	Bacteroidetes	Flavobacterium johnsoniae UW101
Bacteria	Chlamydiae	Parachlamydia acanthamoebae UV-7
Bacteria	Spirochaetes	Spirochaeta africana DSM 8902
Bacteria	Deinococcus-Thermus	Deinococcus maricopensis DSM 21211
Bacteria	Chloroflexi	Herpetosiphon aurantiacus DSM 785
Bacteria	Thermotogae	Marinitoga piezophila KA3
Bacteria	Chlorobi	Chlorobium phaeobacteroides DSM 266
Bacteria	Aquificae	Aguifex aeolicus VF5
Bacteria	Acidobacteria	Acidobacterium capsulatum ATCC 51196
Bacteria	Fusobacteria	Fusobacterium nucleatum subsp. nucleatum ATCC 25586
Bacteria	Planctomycetes	Planctomyces brasiliensis DSM 5305
Bacteria	Synergistetes	Thermanaerovibrio acidaminovorans DSM 6589
Bacteria	Verrucomicrobia	Opitutus terrae PB90-1
Bacteria	Nitrospirae	Candidatus Nitrospira defluvii
Bacteria	Deferribacteres	Deferribacter desulfuricans SSM1
Bacteria	Ignavibacteria	Ignavibacterium album JCM 16511
Bacteria	Elusimicrobia	Elusimicrobium minutum Pei191
Bacteria	Thermodesulfobacteria	Thermodesulfobacterium sp. OPB45
Bacteria	Dictyoglomi	Dictyoglomus turgidum DSM 6724
Bacteria	Gemmatimonadetes	Gemmatimonas aurantiaca T-27
Bacteria	Fibrobacteres	Fibrobacter succinogenes subsp. succinogenes S85
Archaea	Thermoprotei	Pyrolobus fumarii 1A
Archaea	Archaeoglobi	Archaeoglobus veneficus SNP6
Archaea	Halobacteria	Halobacterium salinarum R1
Archaea	Methanobacteria	Methanobrevibacter ruminantium M1
Archaea	Methanococci	Methanococcus aeolicus Nankai-3
Archaea	Methanococci	Methanocella paludicola SANAE
Archaea	Methanopyri	Methanopyrus kandleri AV19
Archaea	Thermococci	Thermococcus onnurineus NA1
Archaea	Thermoplasmata	Thermococcus of marmeds NAT  Thermoplasma volcanium GSS1
Eukarya	•	Chlorella variabilis
Eukarya	Chlorophyta Mollusca	Lottia gigantea
-	Rotifera	Adineta vaga
Eukarya Eukarya		Aspergillus oryzae RIB40
Eukarya Eukarya	Ascomycota Chordata	Mus musculus
-		Drosophila melanogaster
Eukarya	Arthropoda Nematoda	
Eukarya		Caenorhabditis elegans Bristol N2
Eukarya	Annelida Strontonbuta	Capitella teleta
Eukarya	Streptophyta	Arabidopsis thaliana
Eukarya	Apicomplexa	Cryptosporidium parvum Iowa II
Eukarya	Slime-molds	Dictyostelium discoideum AX4
Eukarya	Microsporidia	Encephalitozoon cuniculi GB-M1
Eukarya	Basidiomycota	Cryptococcus gattii WM276
Eukarya	Amoebozoa	Acanthamoeba castellanii Neff ATCC 50373
Eukarya	Dinoflagellata	Karenia brevis isolate CCMP718
Eukarya	Choanoflagellida -	Salpingoeca sp. ATCC 50818
Eukarya	Stramenopiles	Phytophthora infestans T30-4

## **Other Supporting Information Files**

Datasets S1–S4 (XLS)