Host microbial lifestyle shapes diversity in the vastly underrepresented global Porifera plasmidome

Başak Öztürk[1], Barbara Cania[2], Kristina Bayer[1], Tal Dagan[2], Ute Hentschel[1,3], David K. Ngugi[1*]

[1] GEOMAR Helmholtz Centre for Ocean Research Kiel, Division of Marine Ecology, Research Unit Marine Symbioses, Kiel, Germany
[2] Institute of General Microbiology, Evolutionary Microbiology Group, Kiel University, Kiel, Germany
[3] Christian-Albrechts-University (CAU) Kiel, Kiel, Germany

[*] Correspondence: dngugi@geomar.de

Supplementary Figures

Figure S1. The bioinformatics workflow used to recover plasmid sequences from 526 metagenomic samples covering 50 sponge genera produced 6,945 reference sponge microbiome plasmids (RefSMPs), representing reference plasmid contig sequences from the global sponge microbiome. RefSMPs are non-redundant plasmid nucleotide sequences clustered at 95% identity and 80% coverage.

Figure S2. The total abundance (RPKM) of reference sponge microbiome plasmids (RefSMPs) affiliated with different sponge genera across all sponge samples. N.D., sponge genus is not defined.

Figure S3. The total number of reference sponge microbiome plasmids (RefSMPs) affiliated with different sponge genera across all sponge samples. N.D., sponge genus is not defined.

Figure S4. The relationship between median plasmid abundance and median diversity of reference sponge microbiome plasmids (RefSMPs) in the plasmidome of HMA and LMA sponges. Each bubble is scaled by the median contig size of RefSMPs in individual sponge genera across the respective metagenomes.

Figure S5. Plasmid taxonomic units (PTUs) shared among *Aplysina aerophoba* individuals and tissues. The PTUs are derived from ANI-based clustering of plasmids from three *A. aerophoba* individuals from this study and the global reference sponge microbiome plasmids (RefSMPs) from published metagenomes. Nodes represent plasmids, colored by the two tissue types (mesohyl and pinacoderm) and shaped by tissue type. Edges indicate ANI ≥ 85%. Plasmids from different tissues and individuals cluster together, indicating they are not tissue-specific.

Figure S6. Circular plot illustrating the relationship between the reference sponge microbiome plasmids (RefSMPs) from one sponge genus and the number of different sponge genera in which the same RefSMP is found, based on read mapping coverage.

Each segment represents a reference sponge genus where the RefSMPs are found. The size of each segment corresponds to the number of plasmids associated with that genus.

Figure S7. Global map showing the locations of sampling sites for the most abundant and widespread RefSMPs organized by sponge genus.

Figure S8. Forest plot showing the enrichment of KEGG pathways within different functional plasmid clusters defined in **Figure 5c**. The primary KEGG pathways significantly distinguishing each functional cluster are identified by Fisher's exact test. The x-axis represents the log2-transformed odds ratio, with error bars indicating 95% confidence intervals. The red dotted line denotes an odds ratio of 1. Additional information is provided in **Supplementary Table S6**.

Figure S9. Distribution of COG categories (**a**) and the top 20 Pfam domains (**b**) encoded in the genes of the 185 most widespread and abundant RefSMPs.

Figure S10. The most widespread plasmids **carry rRNA genes.** Maps of GSP4368 and GSP0576 show complete or partial rRNA operons (*rrL*, *rrS*, and *rrF*) along with plasmid-specific genes, including the *trwC* relaxase and *tra* genes. Other rRNA-bearing RefSMPs (e.g., GSP6302, GSP0902, and GSP2965) also encode plasmid-associated genes related to replication (*repC, parA, parB, trnI*, *trnA*, and *trnfM*) and mobility (*trwC*, *vapC*, *virB2*, and *traE*), as well as genes conferring antibiotic or heavy metal resistance (*mobA* and *modE*/*mopA*) and other metabolic functions (e.g., cysteine synthase, crotonase). For brevity, the plasmid genomes are shown as circular plots.

Figure S11. Whole-genome alignments of rRNA-bearing RefSMPs to the closest sponge-derived MAGs (metagenome-assembled genomes) with the highest nucleotide sequence identity in the NCBI database. The sponge MAGs are *Theonella swinhoei* (Accession no. OZ220750.1, 61% coverage) and *Xestospongia bergquistia* (Accession no. OY763937.1).

Supplementary Tables

Table S1. General information on the sponge metagenomes used in this study.

Table S2. General information on plasmids and Bakta annotation statistics.

Table S3. CD-HIT-EST clustering results with the PLSDB database.

Table S4. Distribution of IMG/PR CD-HIT-EST plasmid hits across ecosystems.

Table S5. Plasmid genome sequences assembled from the mesohyl and pinacoderm of three independent *Aplysina aerophoba* sponge plasmidomes and validation of sequence presence through custom PCR-based gene analyses. Additional details are provided in Tables S17 and S18.

Table S6. Functional clusters of widespread RefSMPs based on KEGG pathways.

Table S7. Eukaryotic-like proteins (ELPs) detected in the global sponge plasmidome based on hmmsearch against the encoded protein sequences of each RefSMP.

Table S8. Antimicrobial resistance genes (ARGs) found in the global sponge plasmidome using DeepARG.

Table S9. Biosynthetic gene clusters (BGCs) found in the global sponge plasmidome with AntiSMASH.

Table S10. Homologs in the sponge plasmidome encoding additional putative enzymes involved in steroid degradation and the biosynthesis of sterols and ceramide, including the closest matches in UniProt and NCBI protein databases.

Table S11. Homologs encoding putative enzymes involved in ceramide biosynthesis in plasmids from the PLSDB, including gene and KEGG KO IDs, plasmid topology and mobility traits, and the taxonomy (including applicable ranks to species level) and lifestyle of the bacteria hosting these plasmids.

Table S12. Homologs encoding putative enzymes involved in sterol biosynthesis in plasmids from the PLSDB, including gene and KEGG KO IDs, plasmid topology and mobility traits, and the taxonomy (including applicable ranks to species level) and lifestyle of the bacteria hosting these plasmids.

Table S13. Homologs encoding putative enzymes involved in steroid degradation in plasmids from the PLSDB, including gene and KEGG KO IDs, plasmid topology and mobility traits, and the taxonomy (including applicable ranks to species level) and lifestyle of the bacteria hosting these plasmids.

Table S14. Pfam domains enriched in 185 widespread reference sponge plasmids (used as background) relative to others (used as foreground).

Table S15. Taxonomic affiliation of the MAGs sharing the highest rRNA nucleotide identity with the rRNA-bearing RefSMPs.

Table S16. Number of metagenomes with a high-scoring match to the rRNA genes of the two widespread plasmids (GSP0576 and GSP4368) as inferred using Pebblescout against public metagenomes in the NCBI database.

Table S17. PCR validation results of multiple gene amplifications directly from *Aplysina* replicate A tissue-specific DNA samples and alignment identities against the target plasmid GSP0576 sequence.

Table S18. PCR validation results of multiple gene amplifications directly from *Aplysina* replicate C tissue-specific DNA samples and alignment identities against the target plasmid GSP4368sequence.

Figure S1

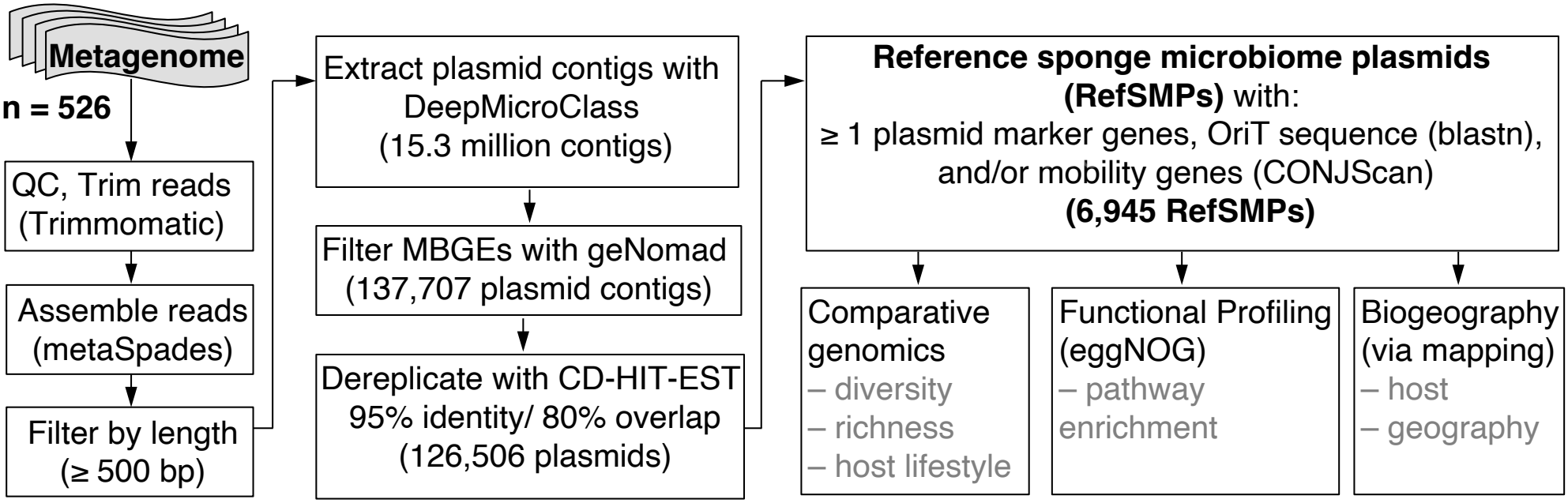

Metagenome
n = 526
QC, Trim reads (Trimmomatic)
Assemble reads (metaSpades)
Filter by length (≥ 500 bp)
Extract plasmid contigs with DeepMicroClass (15.3 million contigs)
Filter MBGEs with geNomad (137,707 plasmid contigs)
Dereplicate with CD-HIT-EST 95% identity/ 80% overlap (126,506 plasmids)
Reference sponge microbiome plasmids (RefSMPs) with:
≥ 1 plasmid marker genes, OriT sequence (blastn), and/or mobility genes (CONJScan)
(6,945 RefSMPs)
Comparative genomics
– diversity
– richness
– host lifestyle
Functional Profiling (eggNOG)
– pathway enrichment
Biogeography (via mapping)
– host
– geography

Figure S2

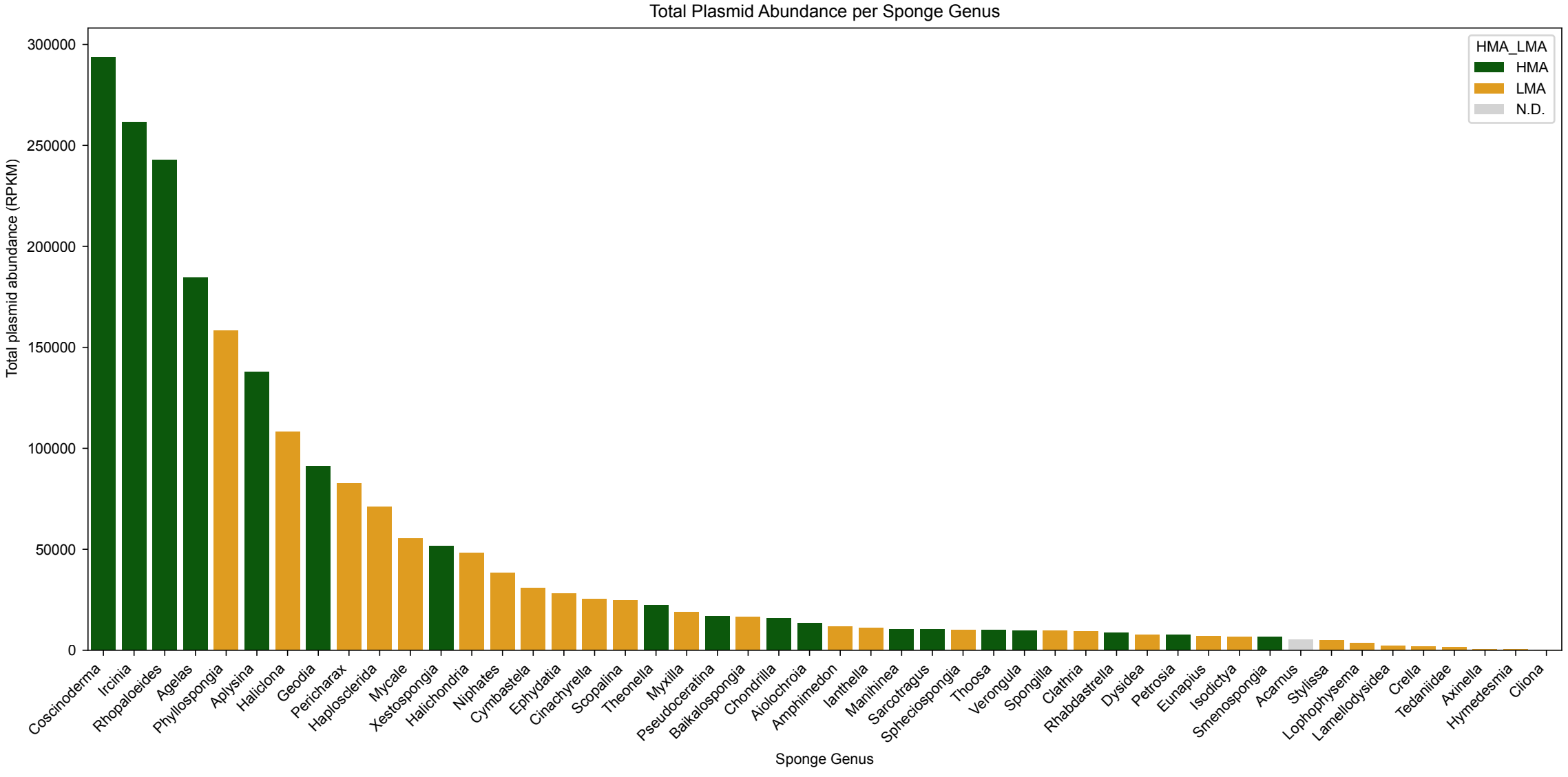

Total Plasmid Abundance per Sponge Genus
Total plasmid abundance (RPKM)
300000
250000
200000
150000
100000
50000
0
HMA_LMA
HMA
LMA
N.D.
Coscinoderma
Ircinia
Rhopaloeides
Agelas
Phyllospongia
Aplysina
Haliclona
Geodia
Pericharax
Haplosclerida
Mycale
Xestospongia
Halichondria
Niphates
Cymbastela
Ephydatia
Cinachyrella
Scopalina
Theonella
Myxilla
Pseudoceratina
Baikalospongia
Chondrilla
Aiolochroia
Amphimedon
Ianthella
Manihinea
Sarcotragus
Spheciospongia
Thoosa
Verongula
Spongilla
Clathria
Rhabdastrella
Dysidea
Petrosia
Eunapius
Isodictya
Smenospongia
Acarnus
Stylissa
Lophophysema
Lamellodysidea
Crella
Tedaniidae
Axinella
Hymedesmia
Cliona
Sponge Genus

Figure S3

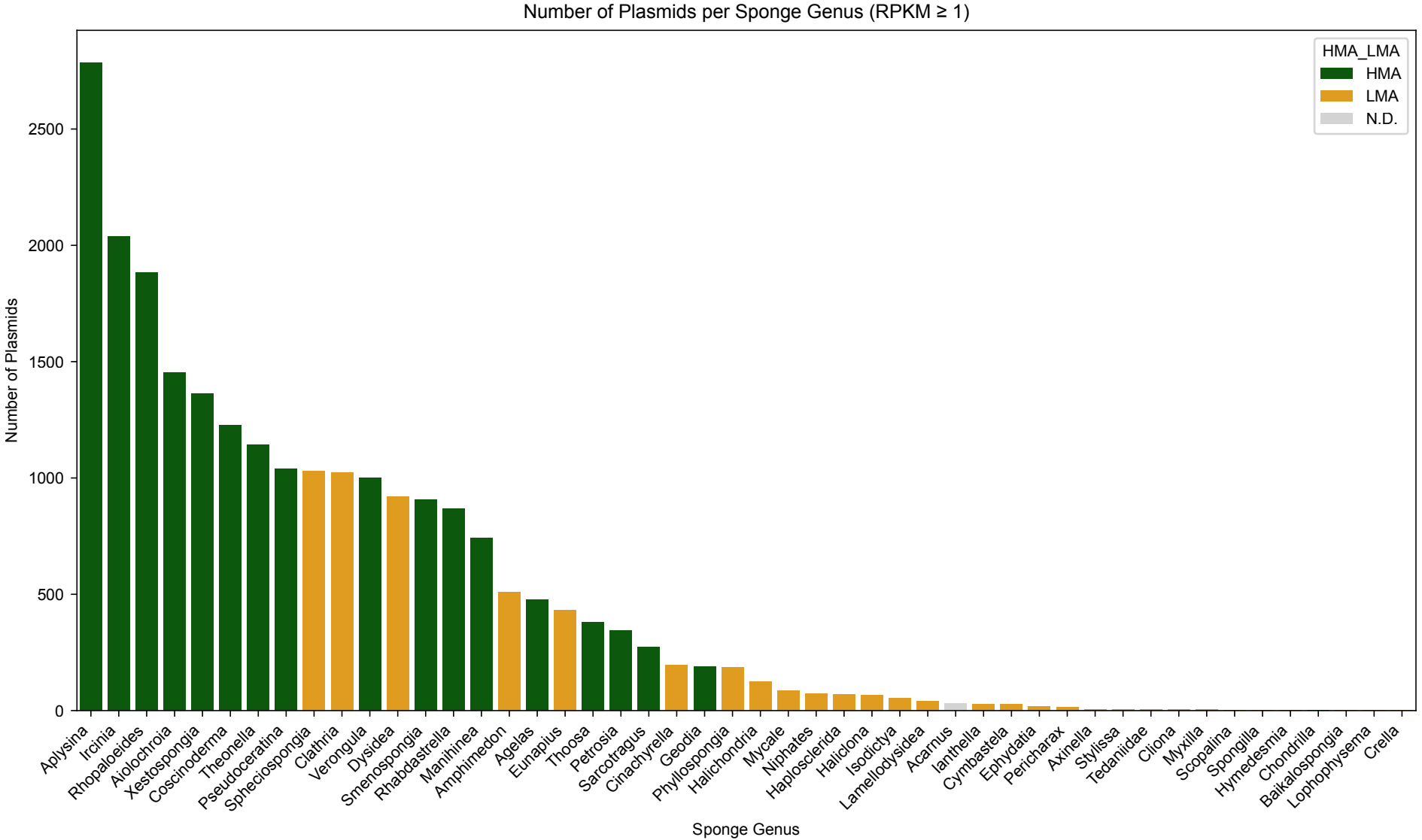

Number of Plasmids per Sponge Genus (RPKM ≥ 1)
HMA_LMA
HMA
LMA
N.D.
Number of Plasmids
0
500
1000
1500
2000
2500
Aplysina
Ircinia
Rhopaloeides
Aiolochroia
Xestospongia
Coscinoderma
Theonella
Pseudoceratina
Spheciospongia
Clathria
Verongula
Dysidea
Smenospongia
Rhabdastrella
Manihinea
Amphimedon
Agelas
Eunapius
Thoosa
Petrosia
Sarcotragus
Cinachyrella
Geodia
Phyllospongia
Halichondria
Mycale
Niphates
Haplosclerida
Haliclona
Isodictya
Lamellodysidea
Acarnus
Ianthella
Cymbastela
Ephydatia
Pericharax
Axinella
Stylissa
Tedaniidae
Cliona
Myxilla
Scopalina
Spongilla
Hymedesmia
Chondrilla
Baikalospongia
Lophophysema
Crella
Sponge Genus

Figure S4

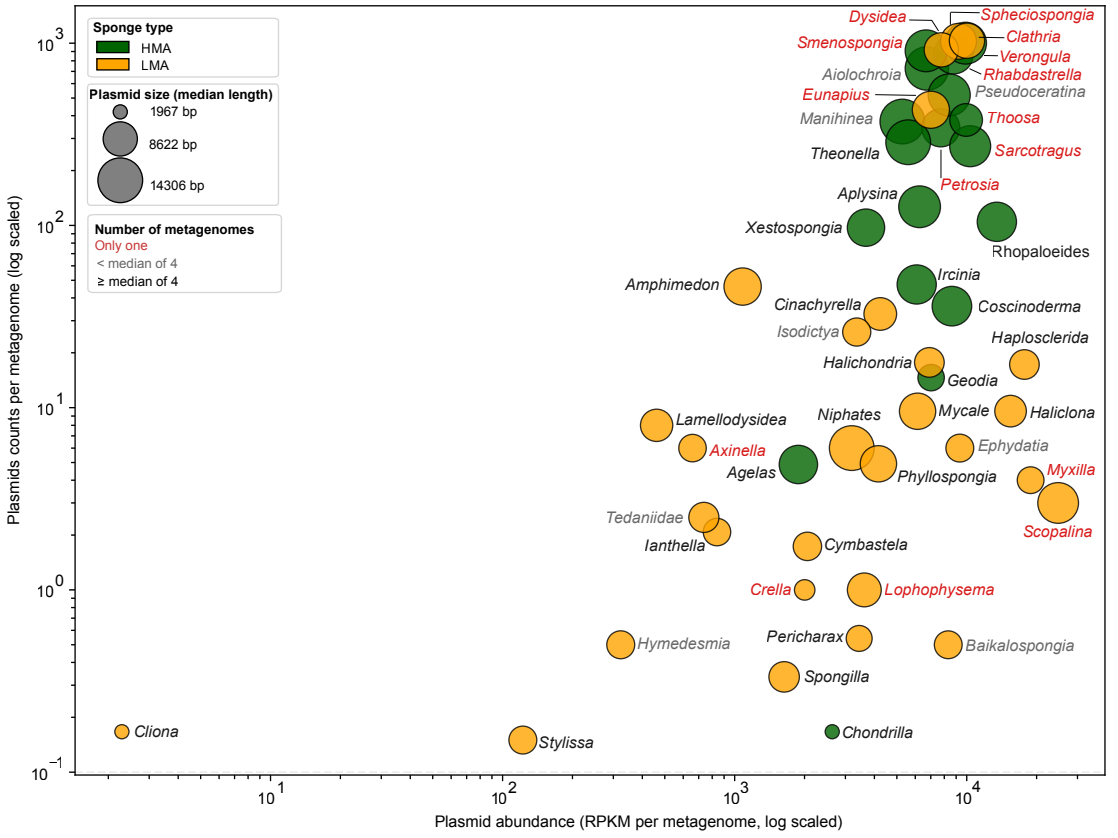

Sponge type
HMA
LMA
Plasmid size (median length)
1967 bp
8622 bp
14306 bp
Number of metagenomes
Only one
< median of 4
≥ median of 4
Dysidea
Spheciospongia
Smenospongia
Clathria
Verongula
Aiolochroia
Rhabdastrella
Eunapius
Pseudoceratina
Manihinea
Thoosa
Theonella
Sarcotragus
Petrosia
Aplysina
Xestospongia
Rhopaloeides
Amphimedon
Ircinia
Cinachyrella
Coscinoderma
Isodictya
Haplosclerida
Halichondria
Geodia
Lamellodysidea
Niphates
Mycale
Haliclona
Axinella
Ephydatia
Agelas
Phyllospongia
Myxilla
Tedaniidae
Scopalina
Ianthella
Cymbastela
Crella
Lophophysema
Hymedesmia
Pericharax
Baikalospongia
Spongilla
Cliona
Stylissa
Chondrilla
Plasmids counts per metagenome (log scaled)
Plasmid abundance (RPKM per metagenome, log scaled)

Figure S5

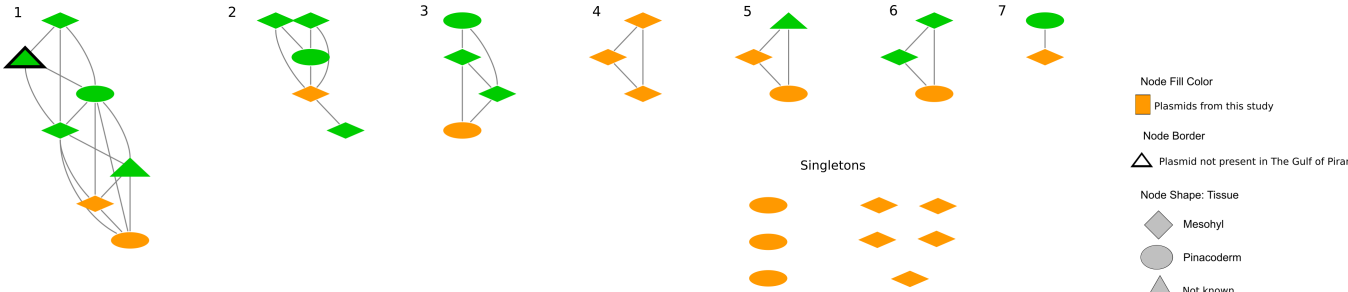

1
2
3
4
5
6
7
Singletons
Node Fill Color
Plasmids from this study
Node Border
Plasmid not present in The Gulf of Piran
Node Shape: Tissue
Mesohyl
Pinacoderm
Not known

Figure S6

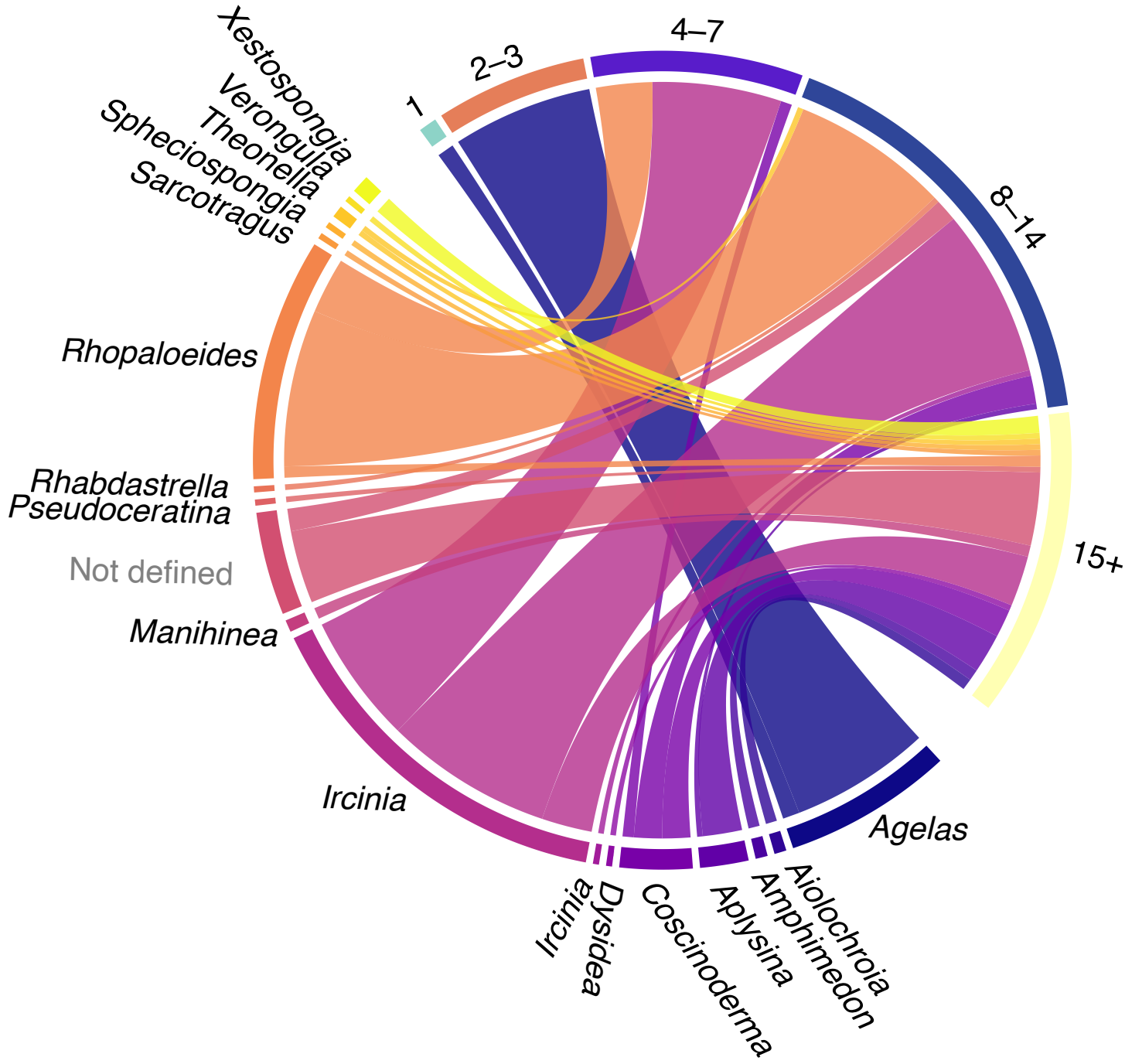

Xestospongia
Verongula
Theonella
Spheciospongia
Sarcotragus
Rhopaloeides
Rhabdastrella
Pseudoceratina
Not defined
Manihinea
Ircinia
Ircinia
Dysidea
Coscinoderma
Aplysina
Amphimedon
Aiolochroia
Agelas
15+
8–14
4–7
2–3
1

Figure S7

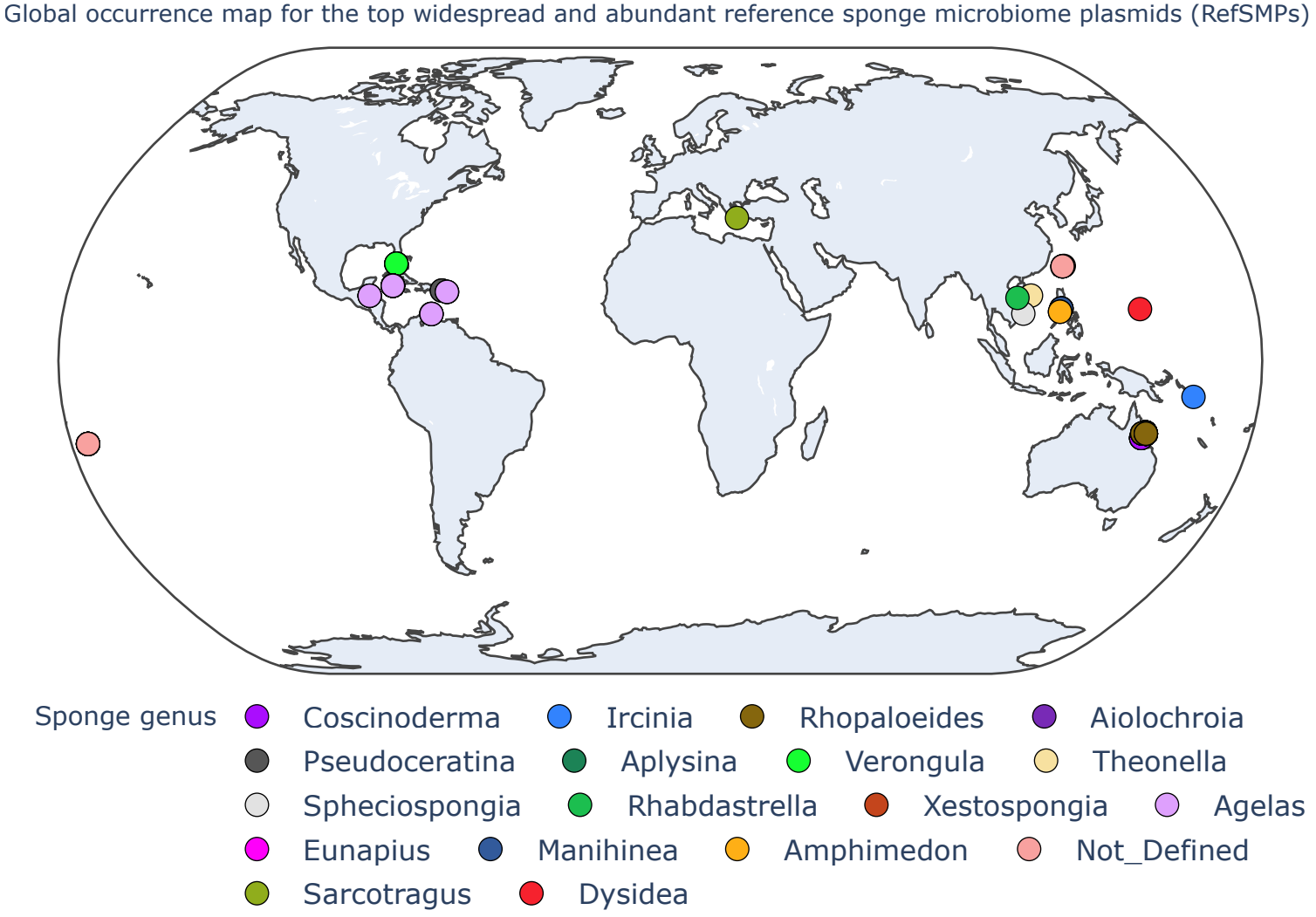

Global occurrence map for the top widespread and abundant reference sponge microbiome plasmids (RefSMPs)
Sponge genus
Coscinoderma
Ircinia
Rhopaloeides
Aiolochroia
Pseudoceratina
Aplysina
Verongula
Theonella
Spheciospongia
Rhabdastrella
Xestospongia
Agelas
Eunapius
Manihinea
Amphimedon
Not_Defined
Sarcotragus
Dysidea

Figure S8

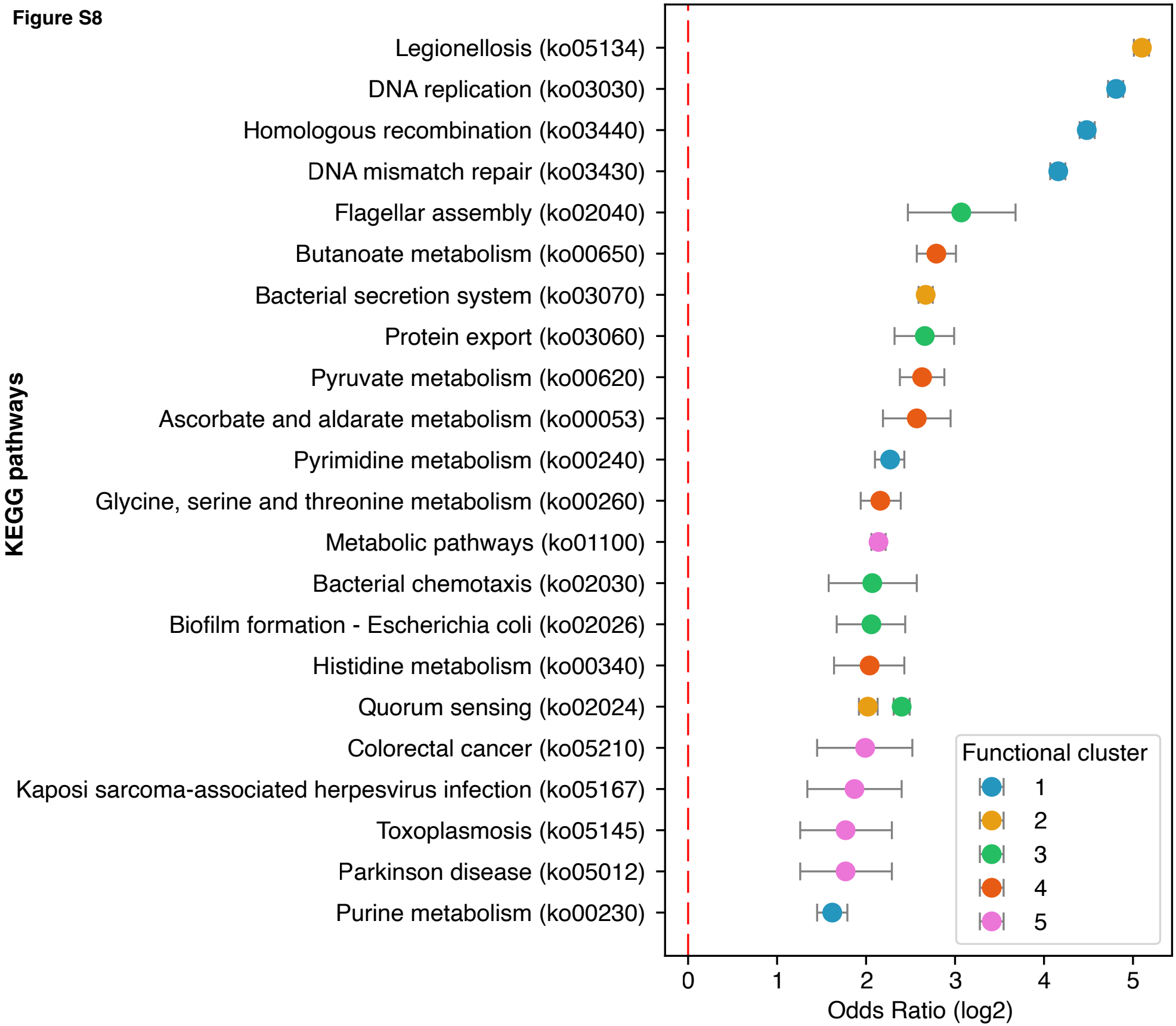

KEGG pathways
Legionellosis (ko05134)
DNA replication (ko03030)
Homologous recombination (ko03440)
DNA mismatch repair (ko03430)
Flagellar assembly (ko02040)
Butanoate metabolism (ko00650)
Bacterial secretion system (ko03070)
Protein export (ko03060)
Pyruvate metabolism (ko00620)
Ascorbate and aldarate metabolism (ko00053)
Pyrimidine metabolism (ko00240)
Glycine, serine and threonine metabolism (ko00260)
Metabolic pathways (ko01100)
Bacterial chemotaxis (ko02030)
Biofilm formation - Escherichia coli (ko02026)
Histidine metabolism (ko00340)
Quorum sensing (ko02024)
Colorectal cancer (ko05210)
Kaposi sarcoma-associated herpesvirus infection (ko05167)
Toxoplasmosis (ko05145)
Parkinson disease (ko05012)
Purine metabolism (ko00230)
Functional cluster
1
2
3
4
5
0
1
2
3
4
5
Odds Ratio (log2)

Figure S9

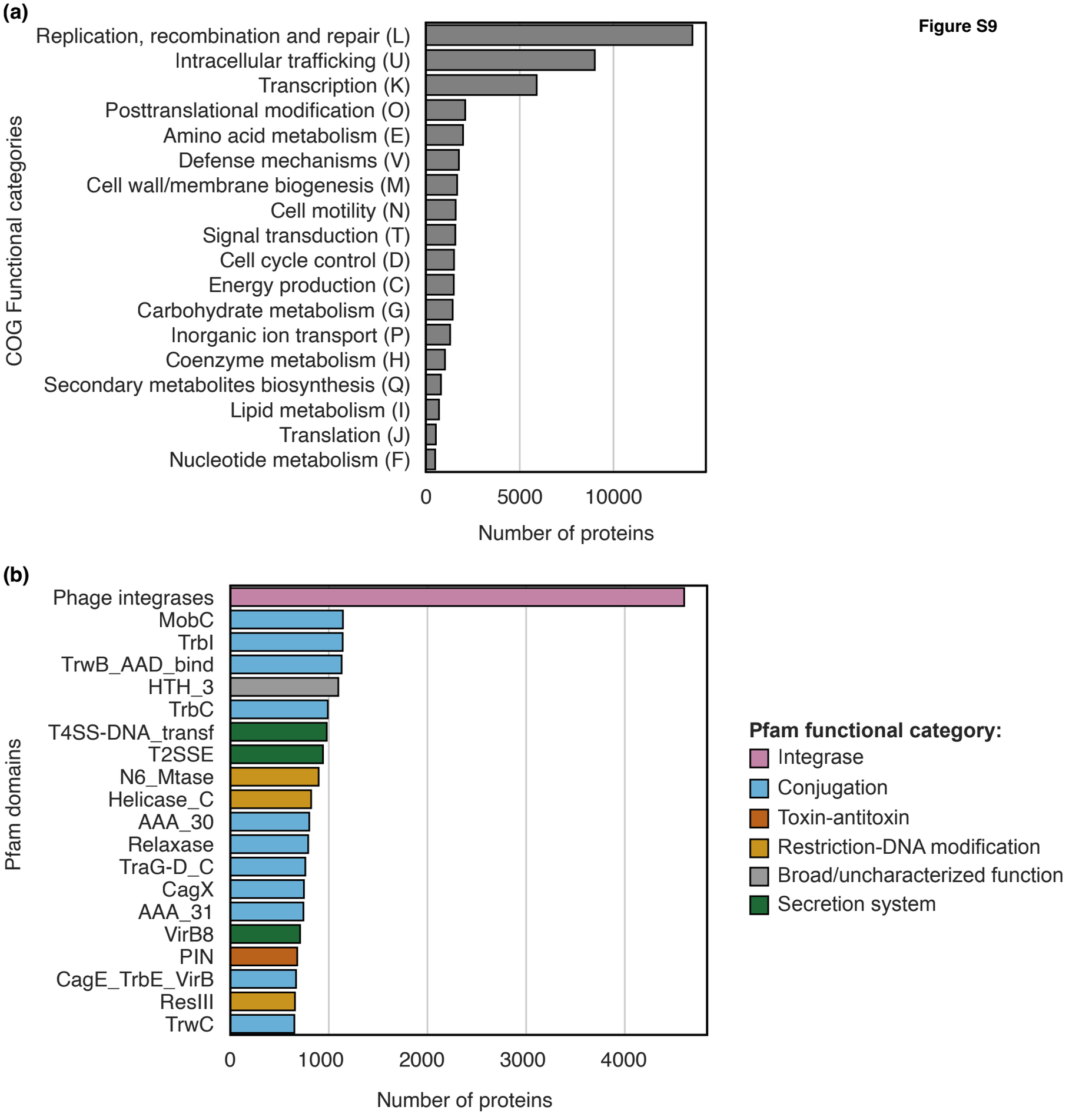

(a)
Replication, recombination and repair (L)
Intracellular trafficking (U)
Transcription (K)
Posttranslational modification (O)
Amino acid metabolism (E)
Defense mechanisms (V)
Cell wall/membrane biogenesis (M)
Cell motility (N)
Signal transduction (T)
Cell cycle control (D)
Energy production (C)
Carbohydrate metabolism (G)
Inorganic ion transport (P)
Coenzyme metabolism (H)
Secondary metabolites biosynthesis (Q)
Lipid metabolism (I)
Translation (J)
Nucleotide metabolism (F)
COG Functional categories
0
5000
10000
Number of proteins
(b)
Phage integrases
MobC
TrbI
TrwB_AAD_bind
HTH_3
TrbC
T4SS-DNA_transf
T2SSE
N6_Mtase
Helicase_C
AAA_30
Relaxase
TraG-D_C
CagX
AAA_31
VirB8
PIN
CagE_TrbE_VirB
ResIII
TrwC
Pfam domains
0
1000
2000
3000
4000
Number of proteins
Pfam functional category:
Integrase
Conjugation
Toxin-antitoxin
Restriction-DNA modification
Broad/uncharacterized function
Secretion system

Figure S10

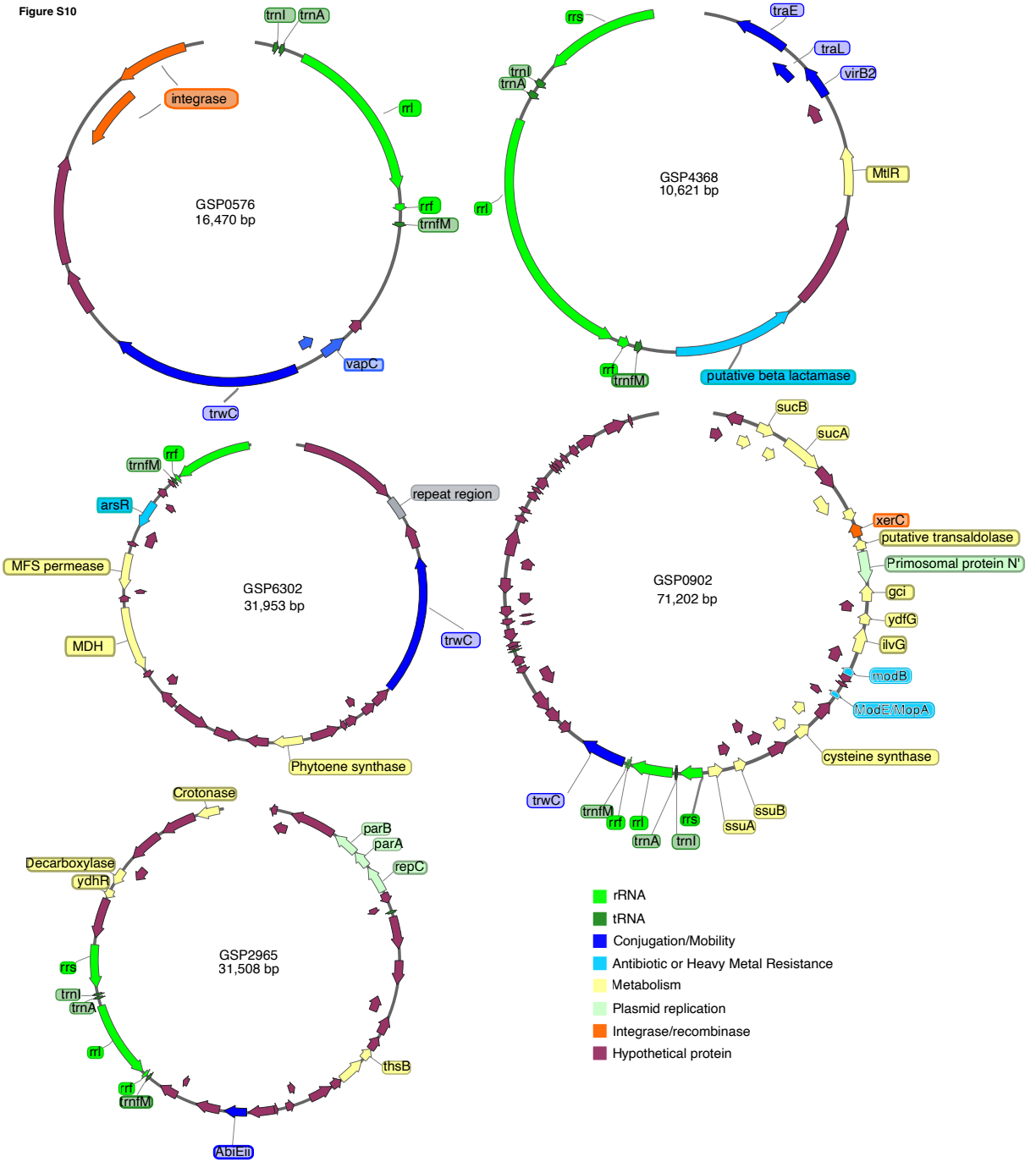
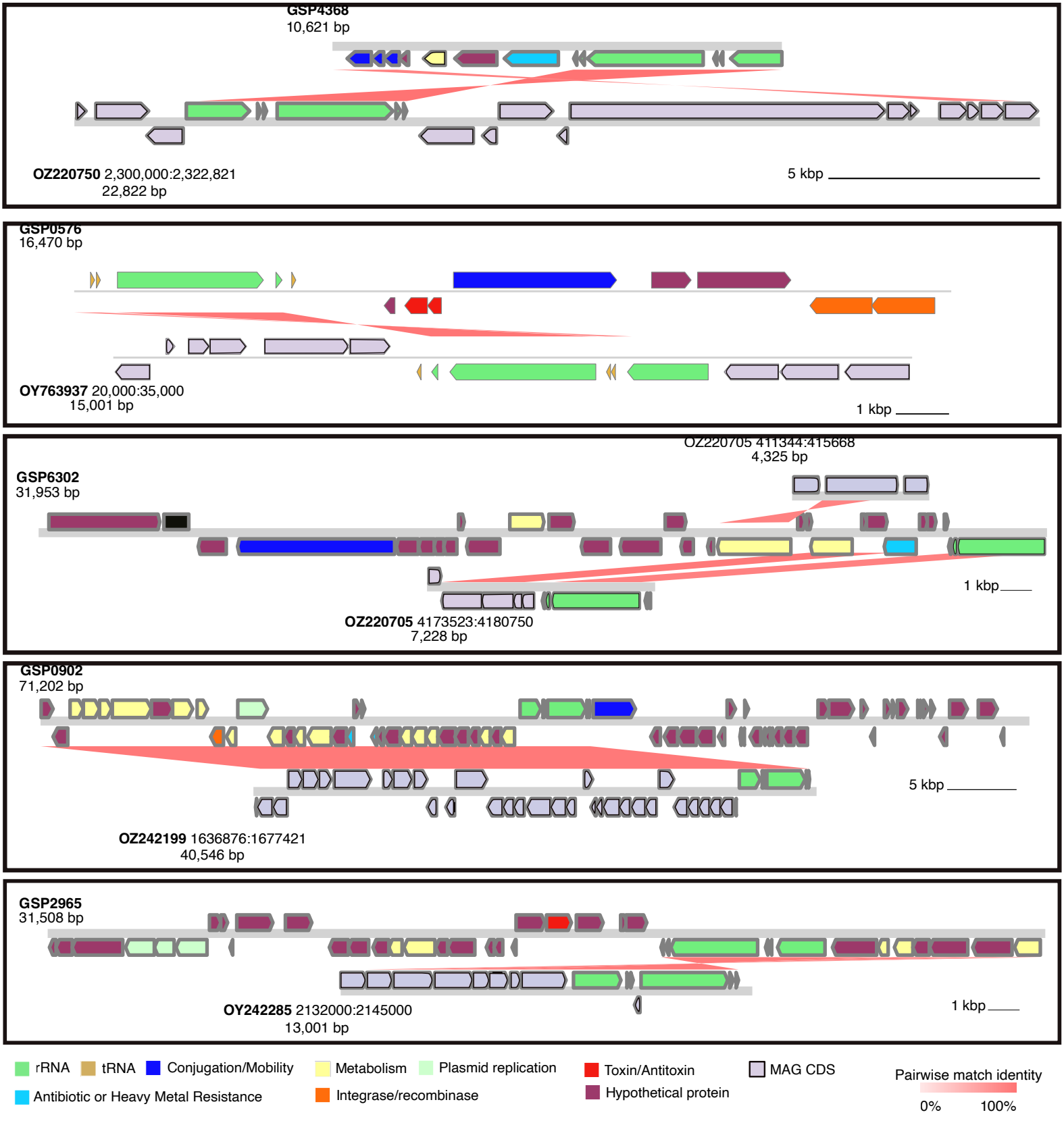

GSP0576
16,470 bp
trnI
trnA
integrase
rrl
rrf
trnfM
vapC
trwC
GSP4368
10,621 bp
rrs
trnI
trnA
rrl
traE
traL
virB2
MtlR
rrf
trnfM
putative beta lactamase
GSP6302
31,953 bp
rrf
trnfM
arsR
repeat region
MFS permease
MDH
trwC
Phytoene synthase
GSP0902
71,202 bp
sucB
sucA
xerC
putative transaldolase
Primosomal protein N'
gci
ydfG
ilvG
modB
ModE/MopA
cysteine synthase
trwC
trnfM
rrf
rrl
trnA
trnI
rrs
ssuA
ssuB
GSP2965
31,508 bp
Crotonase
parB
parA
repC
Decarboxylase
ydhR
rrs
trnI
trnA
rrl
rrf
trnfM
thsB
AbiEii
rRNA
tRNA
Conjugation/Mobility
Antibiotic or Heavy Metal Resistance
Metabolism
Plasmid replication
Integrase/recombinase
Hypothetical protein

Figure S11


GSP4368
10,621 bp
OZ220750 2,300,000:2,322,821
22,822 bp
5 kbp
GSP0576
16,470 bp
OY763937 20,000:35,000
15,001 bp
1 kbp
OZ220705 411344:415668
4,325 bp
GSP6302
31,953 bp
1 kbp
OZ220705 4173523:4180750
7,228 bp
GSP0902
71,202 bp
5 kbp
OZ242199 1636876:1677421
40,546 bp
GSP2965
31,508 bp
1 kbp
OY242285 2132000:2145000
13,001 bp
rRNA
tRNA
Conjugation/Mobility
Metabolism
Plasmid replication
Toxin/Antitoxin
MAG CDS
Antibiotic or Heavy Metal Resistance
Integrase/recombinase
Hypothetical protein
Pairwise match identity
0%
100%