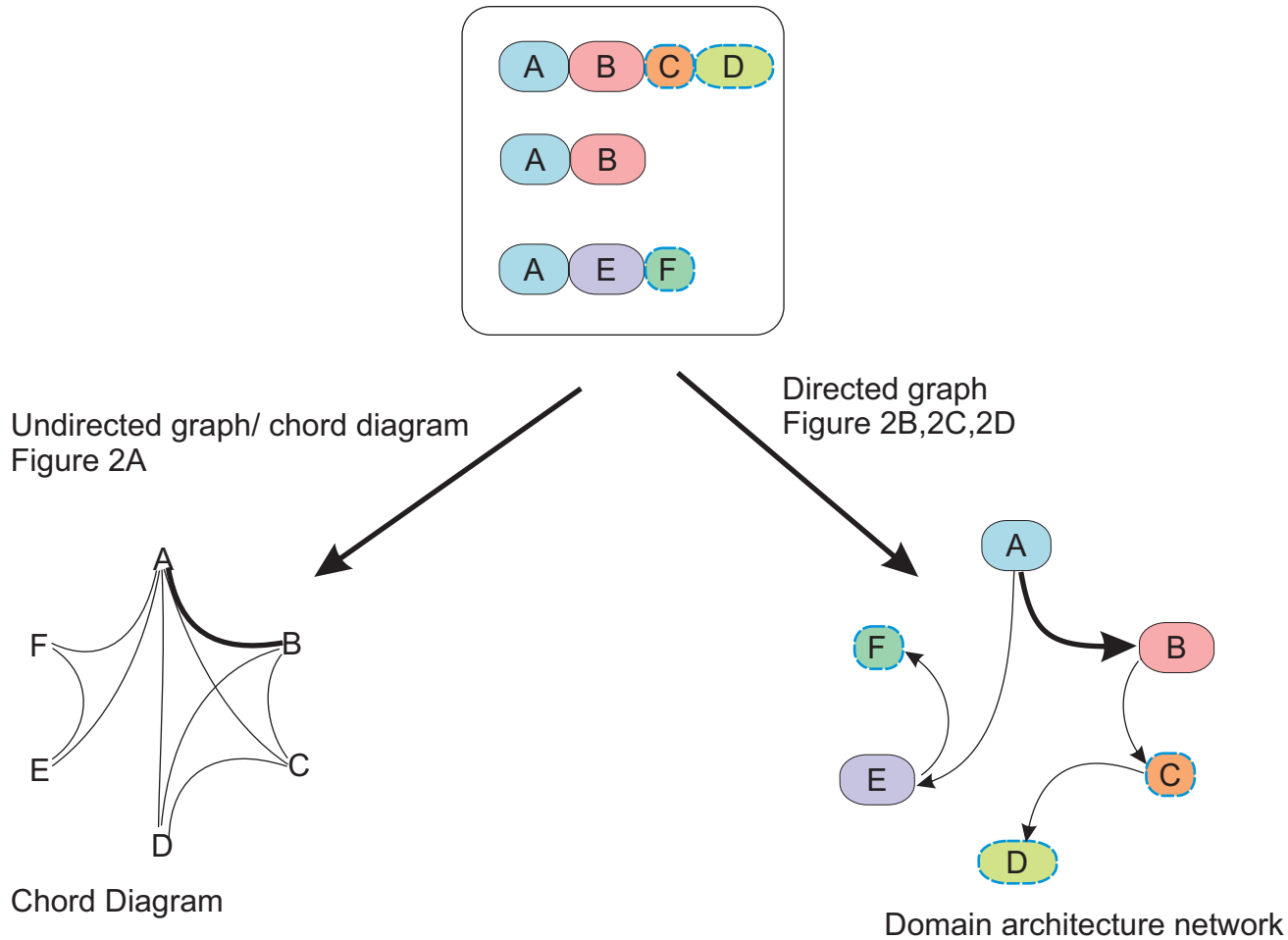
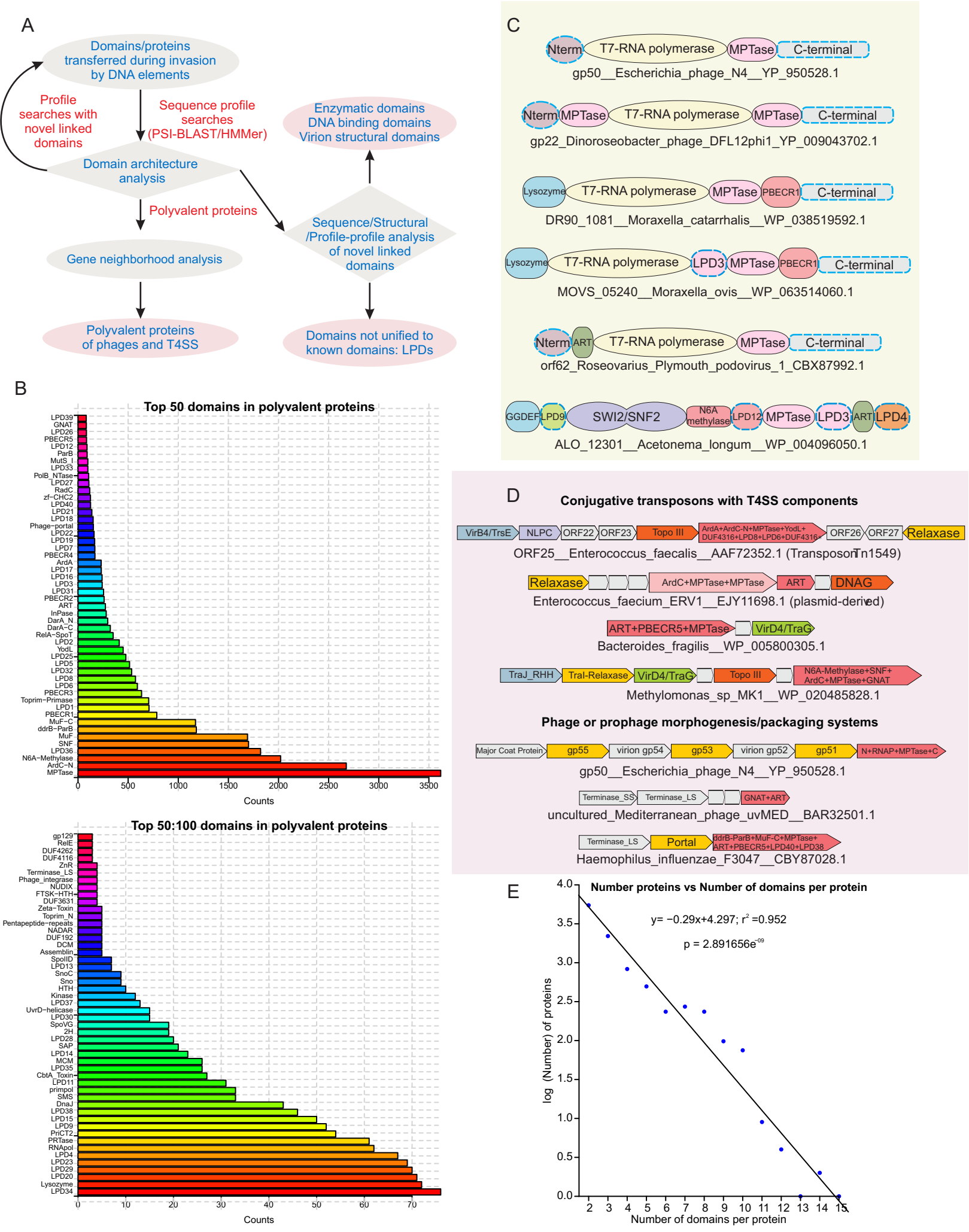
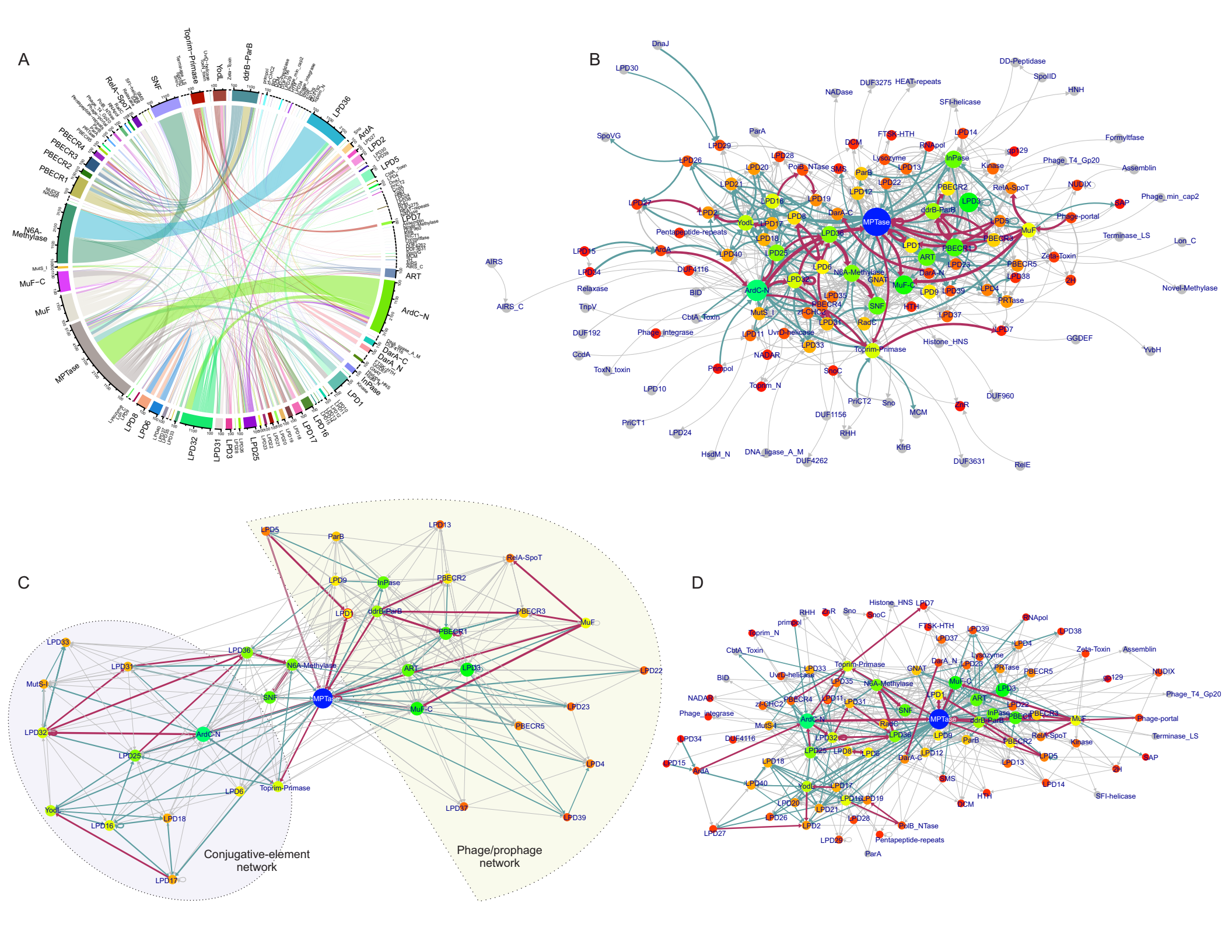
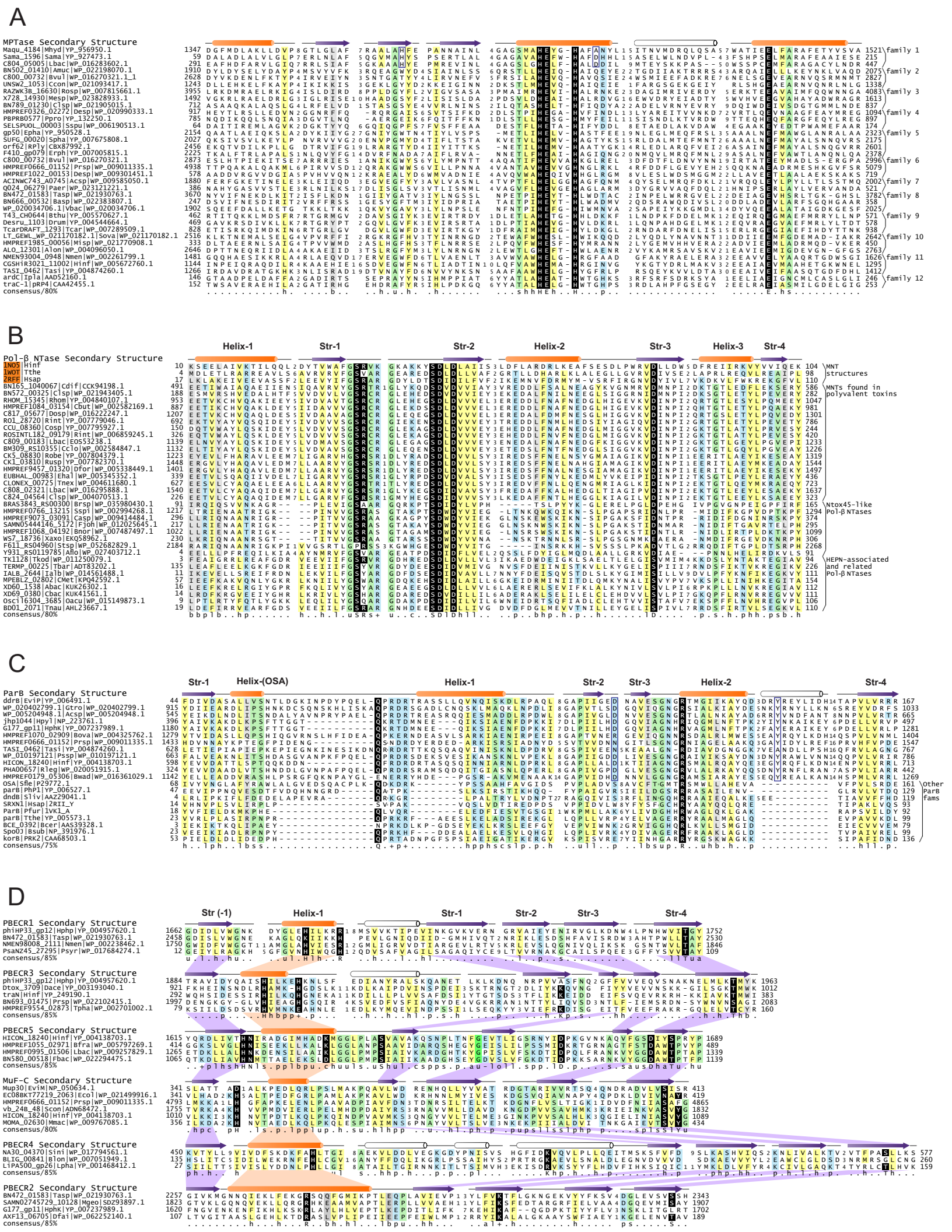
Supplemental S8: high resolution pdf versions of the figures follow starting with a graphical explanation of the algorithm used to construct the graphs in Figure 2.

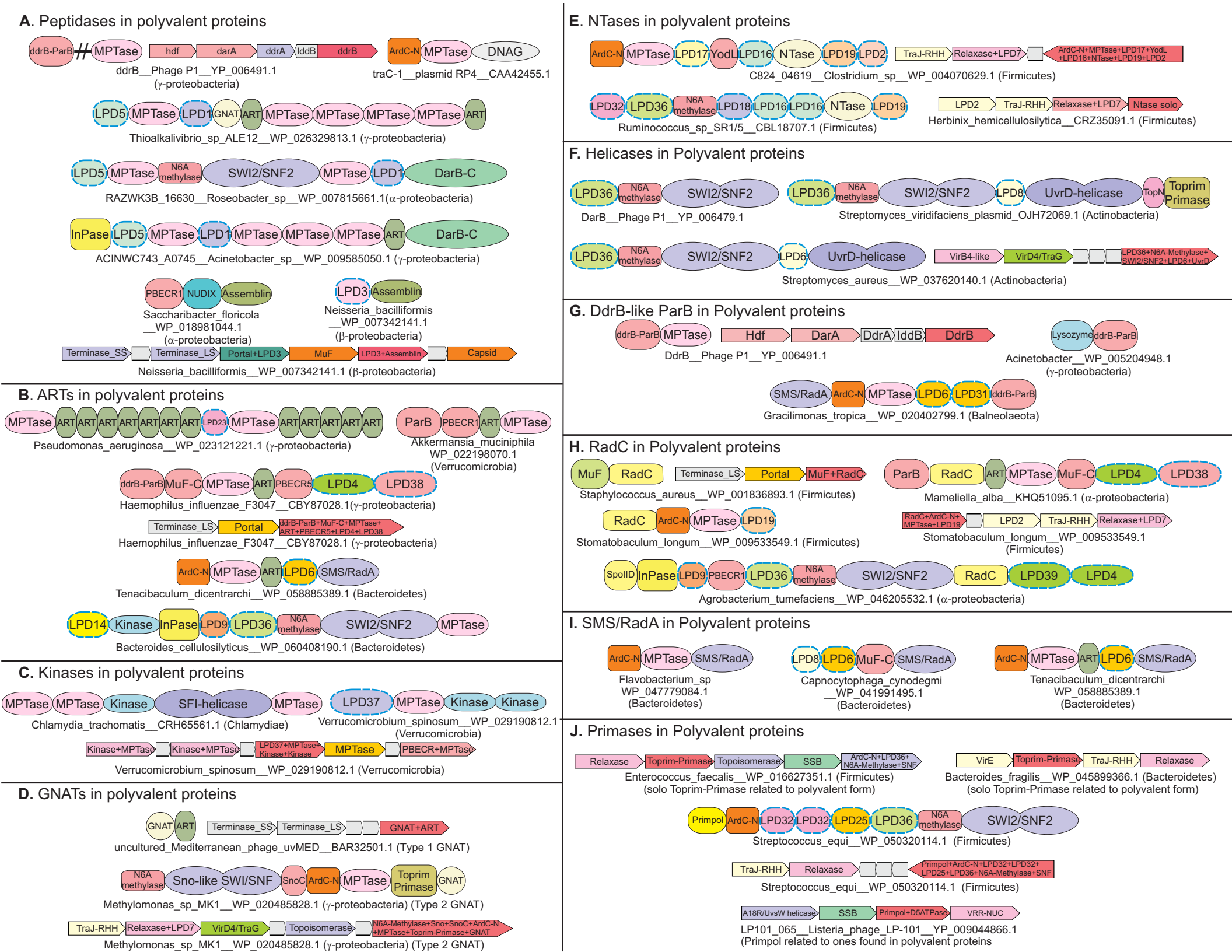


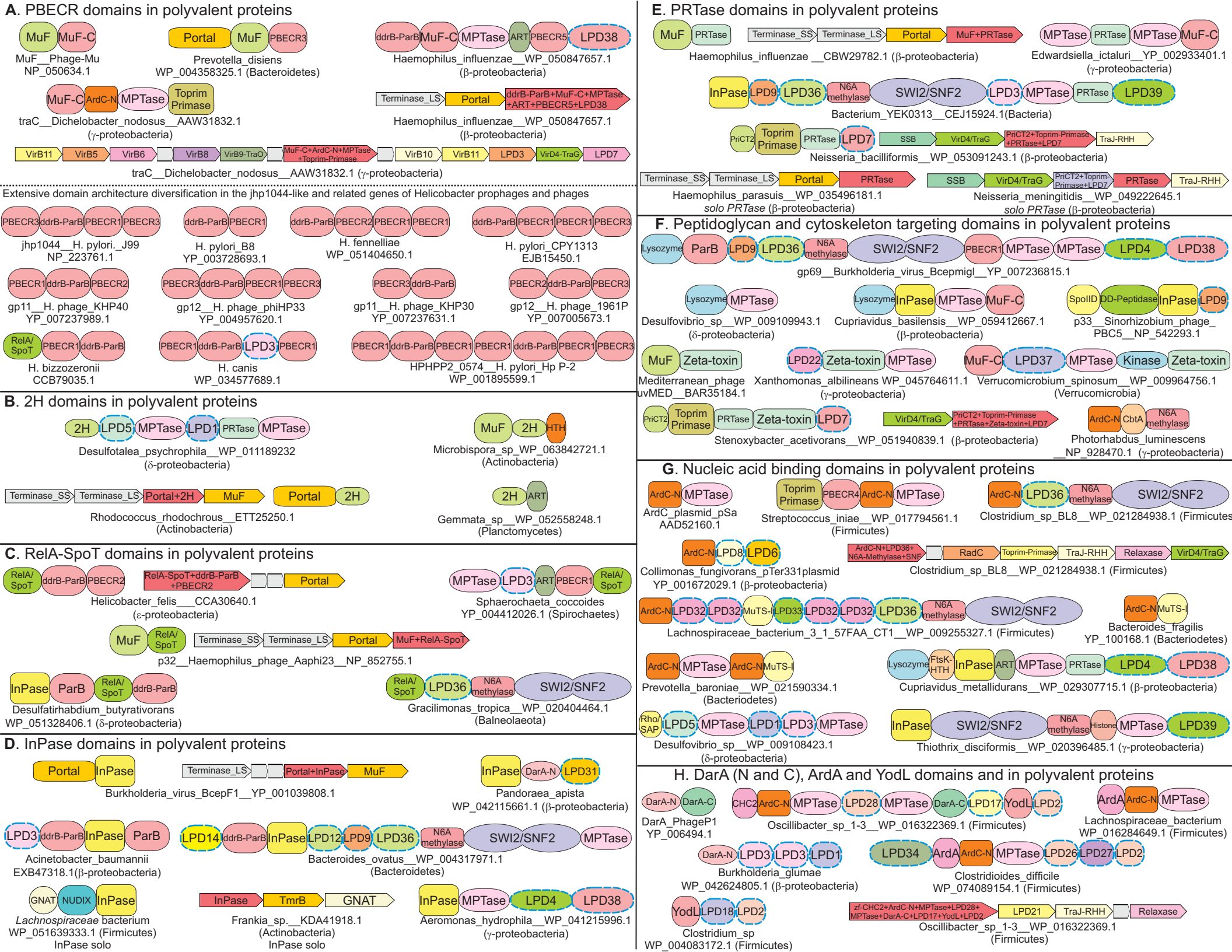
To recover the raw data file used to make the Figure 2A graph, click here.
To recover the raw data file for the graphs in Figure 2B,2C,2D, click here.


A
Domains/proteins transferred during invasion by DNA elements
Profile searches with novel linked domains
Sequence profile searches (PSI-BLAST/HMMer)
Enzymatic domains
DNA binding domains
Virion structural domains
Domain architecture analysis
Polyvalent proteins
Sequence/Structural /Profile-profile analysis of novel linked domains
Gene neighborhood analysis
Polyvalent proteins of phages and T4SS
Domains not unified to known domains: LPDs
B
Top 50 domains in polyvalent proteins
Counts
Top 50:100 domains in polyvalent proteins
Counts
C
Nterm
T7-RNA polymerase
MPTase
C-terminal
gp50__Escherichia_phage_N4__YP_950528.1
Nterm
MPTase
T7-RNA polymerase
MPTase
C-terminal
gp22_Dinoroseobacter_phage_DFL12phi1_YP_009043702.1
Lysozyme
T7-RNA polymerase
MPTase
PBECR1
C-terminal
DR90_1081__Moraxella_catarrhalis__WP_038519592.1
Lysozyme
T7-RNA polymerase
LPD3
MPTase
PBECR1
C-terminal
MOVS_05240__Moraxella_ovis__WP_063514060.1
Nterm
ART
T7-RNA polymerase
MPTase
C-terminal
orf62_Roseovarius_Plymouth_podovirus_1_CBX87992.1
GGDEF
LPD9
SWI2/SNF2
N6A methylase
LPD12
MPTase
LPD3
ART
LPD4
ALO_12301__Acetonema_longum__WP_004096050.1
D
Conjugative transposons with T4SS components
VirB4/TrsE
NLPC
ORF22
ORF23
Topo III
ArdA+ArdC-N+MPTase+YodL+ DUF4316+LPD8+LPD6+DUF4316
ORF26
ORF27
Relaxase
ORF25__Enterococcus_faecalis__AAF72352.1 (Transposon Tn1549)
Relaxase
ArdC+MPTase+MPTase
ART
DNAG
Enterococcus_faecium_ERV1__EJY11698.1 (plasmid-derived)
ART+PBECR5+MPTase
VirD4/TraG
Bacteroides_fragilis__WP_005800305.1
TraJ_RHH
TraI-Relaxase
VirD4/TraG
Topo III
N6A-Methylase+SNF+ ArdC+MPTase+GNAT
Methylomonas_sp_MK1__WP_020485828.1
Phage or prophage morphogenesis/packaging systems
Major Coat Protein
gp55
virion gp54
gp53
virion gp52
gp51
N+RNAP+MPTase+C
gp50__Escherichia_phage_N4__YP_950528.1
Terminase_SS
Terminase_LS
GNAT+ART
uncultured_Mediterranean_phage_uvMED__BAR32501.1
Terminase_LS
Portal
ddrB-ParB+MuF-C+MPTase+ ART+PBECR5+LPD40+LPD38
Haemophilus_influenzae_F3047__CBY87028.1
E
Number proteins vs Number of domains per protein
y= −0.29x+4.297; r² =0.952
p = 2.891656e⁻⁰⁹
log (Number) of proteins
Number of domains per protein


A
B
C
D
Conjugative-element network
Phage/prophage network


A
MPTase Secondary Structure
B
Pol-β NTase Secondary Structure
Helix-1
Str-1
Str-2
Helix-2
Str-3
Helix-3
Str-4
C
ParB Secondary Structure
Str-1
Helix-(OSA)
Helix-1
Str-2
Str-3
Helix-2
Str-4
D
PBECR1 Secondary Structure
Str (-1)
Helix-1
Str-1
Str-2
Str-3
Str-4
PBECR3 Secondary Structure
PBECR5 Secondary Structure
MuF-C Secondary Structure
PBECR4 Secondary Structure
PBECR2 Secondary Structure


A. Peptidases in polyvalent proteins
ddrB-ParB
MPTase
hdf
darA
ddrA
IddB
ddrB
ddrB__Phage P1__YP_006491.1
(γ-proteobacteria)
ArdC-N
MPTase
DNAG
traC-1__plasmid RP4__CAA42455.1
LPD5
MPTase
LPD1
GNAT
ART
MPTase
MPTase
MPTase
MPTase
ART
Thioalkalivibrio_sp_ALE12__WP_026329813.1 (γ-proteobacteria)
LPD5
MPTase
N6A methylase
SWI2/SNF2
MPTase
LPD1
DarB-C
RAZWK3B_16630__Roseobacter_sp__WP_007815661.1(α-proteobacteria)
InPase
LPD5
MPTase
LPD1
MPTase
MPTase
MPTase
ART
DarB-C
ACINWC743_A0745__Acinetobacter_sp__WP_009585050.1 (γ-proteobacteria)
PBECR1
NUDIX
Assemblin
Saccharibacter_floricola
__WP_018981044.1
(α-proteobacteria)
LPD3
Assemblin
Neisseria_bacilliformis
__WP_007342141.1
(β-proteobacteria)
Terminase_SS
Terminase_LS
Portal+LPD3
MuF
LPD3+Assemblin
Capsid
Neisseria_bacilliformis__WP_007342141.1 (β-proteobacteria)
B. ARTs in polyvalent proteins
MPTase
ART
ART
ART
ART
ART
ART
ART
LPD23
MPTase
ART
ART
ART
ART
ART
Pseudomonas_aeruginosa__WP_023121221.1 (γ-proteobacteria)
ParB
PBECR1
ART
MPTase
Akkermansia_muciniphila
WP_022198070.1
(Verrucomicrobia)
ddrB-ParB
MuF-C
MPTase
ART
PBECR5
LPD4
LPD38
Haemophilus_influenzae_F3047__CBY87028.1(γ-proteobacteria)
Terminase_LS
Portal
ddrB-ParB+MuF-C+MPTase+
ART+PBECR5+LPD4+LPD38
Haemophilus_influenzae_F3047__CBY87028.1 (γ-proteobacteria)
ArdC-N
MPTase
ART
LPD6
SMS/RadA
Tenacibaculum_dicentrarchi__WP_058885389.1 (Bacteroidetes)
LPD14
Kinase
InPase
LPD9
LPD36
N6A methylase
SWI2/SNF2
MPTase
Bacteroides_cellulosilyticus__WP_060408190.1 (Bacteroidetes)
C. Kinases in polyvalent proteins
MPTase
MPTase
Kinase
SFI-helicase
MPTase
Chlamydia_trachomatis__CRH65561.1 (Chlamydiae)
LPD37
MPTase
Kinase
Kinase
Verrucomicrobium_spinosum__WP_029190812.1
(Verrucomicrobia)
Kinase+MPTase
Kinase+MPTase
LPD37+MPTase+
Kinase+Kinase
MPTase
PBECR+MPTase
Verrucomicrobium_spinosum__WP_029190812.1 (Verrucomicrobia)
D. GNATs in polyvalent proteins
GNAT
ART
Terminase_SS
Terminase_LS
GNAT+ART
uncultured_Mediterranean_phage_uvMED__BAR32501.1 (Type 1 GNAT)
N6A methylase
Sno-like SWI/SNF
SnoC
ArdC-N
MPTase
Toprim Primase
GNAT
Methylomonas_sp_MK1__WP_020485828.1 (γ-proteobacteria) (Type 2 GNAT)
TraJ-RHH
Relaxase+LPD7
VirD4/TraG
Topoisomerase
N6A-Methylase+Sno+SnoC+ArdC-N
+MPTase+Toprim-Primase+GNAT
Methylomonas_sp_MK1__WP_020485828.1 (γ-proteobacteria) (Type 2 GNAT)
E. NTases in polyvalent proteins
ArdC-N
MPTase
LPD17
YodL
LPD16
NTase
LPD19
LPD2
TraJ-RHH
Relaxase+LPD7
ArdC-N+MPTase+LPD17+YodL
+LPD16+NTase+LPD19+LPD2
C824_04619__Clostridium_sp__WP_004070629.1 (Firmicutes)
LPD32
LPD36
N6A methylase
LPD18
LPD16
LPD16
NTase
LPD19
Ruminococcus_sp_SR1/5__CBL18707.1 (Firmicutes)
LPD2
TraJ-RHH
Relaxase+LPD7
Ntase solo
Herbinix_hemicellulosilytica__CRZ35091.1 (Firmicutes)
F. Helicases in Polyvalent proteins
LPD36
N6A methylase
SWI2/SNF2
DarB__Phage P1__YP_006479.1
LPD36
N6A methylase
SWI2/SNF2
LPD8
UvrD-helicase
TopN
Toprim Primase
Streptomyces_viridifaciens_plasmid_OJH72069.1 (Actinobacteria)
LPD36
N6A methylase
SWI2/SNF2
LPD6
UvrD-helicase
VirB4-like
VirD4/TraG
LPD36+N6A-Methylase+
SWI2/SNF2+LPD6+UvrD
Streptomyces_aureus__WP_037620140.1 (Actinobacteria)
G. DdrB-like ParB in Polyvalent proteins
ddrB-ParB
MPTase
Hdf
DarA
DdrA
IddB
DdrB
DdrB__Phage P1__YP_006491.1
Lysozyme
ddrB-ParB
Acinetobacter__WP_005204948.1
(γ-proteobacteria)
SMS/RadA
ArdC-N
MPTase
LPD6
LPD31
ddrB-ParB
Gracilimonas_tropica__WP_020402799.1 (Balneolaeota)
H. RadC in Polyvalent proteins
MuF
RadC
Terminase_LS
Portal
MuF+RadC
Staphylococcus_aureus__WP_001836893.1 (Firmicutes)
ParB
RadC
ART
MPTase
MuF-C
LPD4
LPD38
Mameliella_alba__KHQ51095.1 (α-proteobacteria)
RadC
ArdC-N
MPTase
LPD19
Stomatobaculum_longum__WP_009533549.1 (Firmicutes)
RadC+ArdC-N+
MPTase+LPD19
LPD2
TraJ-RHH
Relaxase+LPD7
Stomatobaculum_longum__WP_009533549.1
(Firmicutes)
SpoIID
InPase
LPD9
PBECR1
LPD36
N6A methylase
SWI2/SNF2
RadC
LPD39
LPD4
Agrobacterium_tumefaciens__WP_046205532.1 (α-proteobacteria)
I. SMS/RadA in Polyvalent proteins
ArdC-N
MPTase
SMS/RadA
Flavobacterium_sp
WP_047779084.1
(Bacteroidetes)
LPD8
LPD6
MuF-C
SMS/RadA
Capnocytophaga_cynodegmi
__WP_041991495.1
(Bacteroidetes)
ArdC-N
MPTase
ART
LPD6
SMS/RadA
Tenacibaculum_dicentrarchi
WP_058885389.1
(Bacteroidetes)
J. Primases in Polyvalent proteins
Relaxase
Toprim-Primase
Topoisomerase
SSB
ArdC-N+LPD36+
N6A-Methylase+SNF
Enterococcus_faecalis__WP_016627351.1 (Firmicutes)
(solo Toprim-Primase related to polyvalent form)
VirE
Toprim-Primase
TraJ-RHH
Relaxase
Bacteroides_fragilis__WP_045899366.1 (Bacteroidetes)
(solo Toprim-Primase related to polyvalent form)
Primpol
ArdC-N
LPD32
LPD32
LPD25
LPD36
N6A methylase
SWI2/SNF2
Streptococcus_equi__WP_050320114.1 (Firmicutes)
TraJ-RHH
Relaxase
Primpol+ArdC-N+LPD32+LPD32+
LPD25+LPD36+N6A-Methylase+SNF
Streptococcus_equi__WP_050320114.1 (Firmicutes)
A18R/UvsW helicase
SSB
Primpol+D5ATPase
VRR-NUC
LP101_065__Listeria_phage_LP-101__YP_009044866.1
(Primpol related to ones found in polyvalent proteins


A. PBECR domains in polyvalent proteins
MuF MuF-C
MuF__Phage-Mu
NP_050634.1
Portal MuF PBECR3
Prevotella_disiens
WP_004358325.1 (Bacteroidetes)
ddrB-ParB MuF-C MPTase ART PBECR5 LPD38
Haemophilus_influenzae__WP_050847657.1
(β-proteobacteria)
MuF-C ArdC-N MPTase Toprim Primase
traC__Dichelobacter_nodosus__AAW31832.1
(γ-proteobacteria)
Terminase_LS Portal ddrB-ParB+MuF-C+MPTase+ART+PBECR5+LPD38
Haemophilus_influenzae__WP_050847657.1
(β-proteobacteria)
VirB11 VirB5 VirB6 VirB8 VirB9-TraO MuF-C+ArdC-N+MPTase+Toprim-Primase VirB10 VirB11 LPD3 VirD4-TraG LPD7
traC__Dichelobacter_nodosus__AAW31832.1 (γ-proteobacteria)
Extensive domain architecture diversification in the jhp1044-like and related genes of Helicobacter prophages and phages
PBECR3 ddrB-ParB PBECR1 PBECR3
jhp1044__H. pylori._J99
NP_223761.1
ddrB-ParB PBECR1
H. pylori_B8
YP_003728693.1
ddrB-ParB PBECR2 PBECR1 PBECR1
H. fennelliae
WP_051404650.1
ddrB-ParB PBECR1 PBECR1 PBECR3
H. pylori_CPY1313
EJB15450.1
PBECR1 ddrB-ParB PBECR2
gp11__H. phage_KHP40
YP_007237989.1
PBECR3 ddrB-ParB PBECR1 PBECR3
gp12__H. phage_phiHP33
YP_004957620.1
PBECR3 ddrB-ParB
gp11__H. phage_KHP30
YP_007237631.1
PBECR2 ddrB-ParB PBECR3
gp12__H. phage_1961P
YP_007005673.1
RelA/SpoT PBECR1 ddrB-ParB
H. bizzozeronii
CCB79035.1
PBECR1 ddrB-ParB LPD3 PBECR1
H. canis
WP_034577689.1
PBECR1 ddrB-ParB PBECR1 PBECR1 ddrB-ParB PBECR1 PBECR3
HPHPP2_0574__H. pylori_Hp P-2
WP_001895599.1
B. 2H domains in polyvalent proteins
2H LPD5 MPTase LPD1 PRTase MPTase
Desulfotalea_psychrophila__WP_011189232
(δ-proteobacteria)
MuF 2H HTH
Microbispora_sp_WP_063842721.1
(Actinobacteria)
Terminase_SS Terminase_LS Portal+2H MuF Portal 2H
Rhodococcus_rhodochrous__ETT25250.1
(Actinobacteria)
2H ART
Gemmata_sp__WP_052558248.1
(Planctomycetes)
C. RelA-SpoT domains in polyvalent proteins
RelA/SpoT ddrB-ParB PBECR2
RelA-SpoT+ddrB-ParB+PBECR2 Portal
Helicobacter_felis___CCA30640.1
(ε-proteobacteria)
MPTase LPD3 ART PBECR1 RelA/SpoT
Sphaerochaeta_coccoides
YP_004412026.1 (Spirochaetes)
MuF RelA/SpoT Terminase_SS Terminase_LS Portal MuF+RelA-SpoT
p32__Haemophilus_phage_Aaphi23__NP_852755.1
InPase ParB RelA/SpoT ddrB-ParB
Desulfatirhabdium_butyrativorans
WP_051328406.1 (δ-proteobacteria)
RelA/SpoT LPD36 N6A methylase SWI2/SNF2
Gracilimonas_tropica__WP_020404464.1
(Balneolaeota)
D. InPase domains in polyvalent proteins
Portal InPase
Burkholderia_virus_BcepF1__YP_001039808.1
Terminase_LS Portal+InPase MuF
InPase DarA-N LPD31
Pandoraea_apista
WP_042115661.1 (β-proteobacteria)
LPD3 ddrB-ParB InPase ParB
Acinetobacter_baumannii
EXB47318.1(β-proteobacteria)
LPD14 ddrB-ParB InPase LPD12 LPD9 LPD36 N6A methylase SWI2/SNF2 MPTase
Bacteroides_ovatus__WP_004317971.1
(Bacteroidetes)
GNAT NUDIX InPase
Lachnospiraceae bacterium
WP_051639333.1 (Firmicutes)
InPase solo
InPase TmrB GNAT
Frankia_sp.__KDA41918.1
(Actinobacteria)
InPase solo
InPase MPTase LPD4 LPD38
Aeromonas_hydrophila__WP_041215996.1
(γ-proteobacteria)
E. PRTase domains in polyvalent proteins
MuF PRTase
Haemophilus_influenzae __CBW29782.1 (β-proteobacteria)
Terminase_SS Terminase_LS Portal MuF+PRTase
MPTase PRTase MPTase MuF-C
Edwardsiella_ictaluri__YP_002933401.1
(γ-proteobacteria)
InPase LPD9 LPD36 N6A methylase SWI2/SNF2 LPD3 MPTase PRTase LPD39
Bacterium_YEK0313__CEJ15924.1(Bacteria)
PriCT2 Toprim Primase PRTase LPD7 SSB VirD4/TraG PriCT2+Toprim-Primase+PRTase+LPD7 TraJ-RHH
Neisseria_bacilliformis__WP_053091243.1 (β-proteobacteria)
Terminase_SS Terminase_LS Portal PRTase
Haemophilus_parasuis__WP_035496181.1
solo PRTase (β-proteobacteria)
SSB VirD4/TraG PriCT2+Toprim-Primase+LPD7 PRTase TraJ-RHH
Neisseria_meningitidis__WP_049222645.1
solo PRTase (β-proteobacteria)
F. Peptidoglycan and cytoskeleton targeting domains in polyvalent proteins
Lysozyme ParB LPD9 LPD36 N6A methylase SWI2/SNF2 PBECR1 MPTase MPTase LPD4 LPD38
gp69__Burkholderia_virus_Bcepmigl__YP_007236815.1
Lysozyme MPTase
Desulfovibrio_sp__WP_009109943.1
(δ-proteobacteria)
Lysozyme InPase MPTase MuF-C
Cupriavidus_basilensis__WP_059412667.1
(β-proteobacteria)
SpoIID DD-Peptidase InPase LPD9
p33__Sinorhizobium_phage_
PBC5__NP_542293.1
MuF Zeta-toxin
Mediterranean_phage
uvMED__BAR35184.1
LPD22 Zeta-toxin MPTase
Xanthomonas_albilineans WP_045764611.1
(γ-proteobacteria)
MuF-C LPD37 MPTase Kinase Zeta-toxin
Verrucomicrobium_spinosum__WP_009964756.1
(Verrucomicrobia)
PriCT2 Toprim Primase PRTase Zeta-toxin LPD7 VirD4/TraG PriCT2+Toprim-Primase+PRTase+Zeta-toxin+LPD7
Stenoxybacter_acetivorans__WP_051940839.1 (β-proteobacteria)
ArdC-N CbtA N6A methylase
Photorhabdus_luminescens
__NP_928470.1 (γ-proteobacteria)
G. Nucleic acid binding domains in polyvalent proteins
ArdC-N MPTase
ArdC_plasmid_pSa
AAD52160.1
Toprim Primase PBECR4 ArdC-N MPTase
Streptococcus_iniae__WP_017794561.1
(Firmicutes)
ArdC-N LPD36 N6A methylase SWI2/SNF2
Clostridium_sp_BL8__WP_021284938.1 (Firmicutes)
ArdC-N LPD8 LPD6
Collimonas_fungivorans_pTer331plasmid
YP_001672029.1 (β-proteobacteria)
ArdC-N+LPD36+N6A-Methylase+SNF RadC Toprim-Primase TraJ-RHH Relaxase VirD4/TraG
Clostridium_sp_BL8__WP_021284938.1 (Firmicutes)
ArdC-N LPD32 LPD32 MuTS-I LPD33 LPD32 LPD32 LPD36 N6A methylase SWI2/SNF2
Lachnospiraceae_bacterium_3_1_57FAA_CT1__WP_009255327.1 (Firmicutes)
ArdC-N MuTS-I
Bacteroides_fragilis
YP_100168.1 (Bacteriodetes)
ArdC-N MPTase ArdC-N MuTS-I
Prevotella_baroniae__WP_021590334.1
(Bacteriodetes)
Lysozyme FtsK-HTH InPase ART MPTase PRTase LPD4 LPD38
Cupriavidus_metallidurans__WP_029307715.1 (β-proteobacteria)
Rho/SAP LPD5 MPTase LPD1 LPD3 MPTase
Desulfovibrio_sp__WP_009108423.1
(δ-proteobacteria)
InPase SWI2/SNF2 N6A methylase Histone MPTase LPD39
Thiothrix_disciformis__WP_020396485.1 (γ-proteobacteria)
H. DarA (N and C), ArdA and YodL domains and in polyvalent proteins
DarA-N DarA-C
DarA_PhageP1
YP_006494.1
CHC2 ArdC-N MPTase LPD28 MPTase DarA-C LPD17 YodL LPD2
Oscillibacter_sp_1-3__WP_016322369.1 (Firmicutes)
ArdA ArdC-N MPTase
Lachnospiraceae_bacterium
WP_016284649.1 (Firmicutes)
DarA-N LPD3 LPD3 LPD1
Burkholderia_glumae
WP_042624805.1 (β-proteobacteria)
LPD34 ArdA ArdC-N MPTase LPD26 LPD27 LPD2
Clostridioides_difficile
WP_074089154.1 (Firmicutes)
YodL LPD18 LPD2
Clostridium_sp
WP_004083172.1 (Firmicutes)
zf-CHC2+ArdC-N+MPTase+LPD28+MPTase+DarA-C+LPD17+YodL+LPD2 LPD21 TraJ-RHH Relaxase
Oscillibacter_sp_1-3__WP_016322369.1
(Firmicutes)