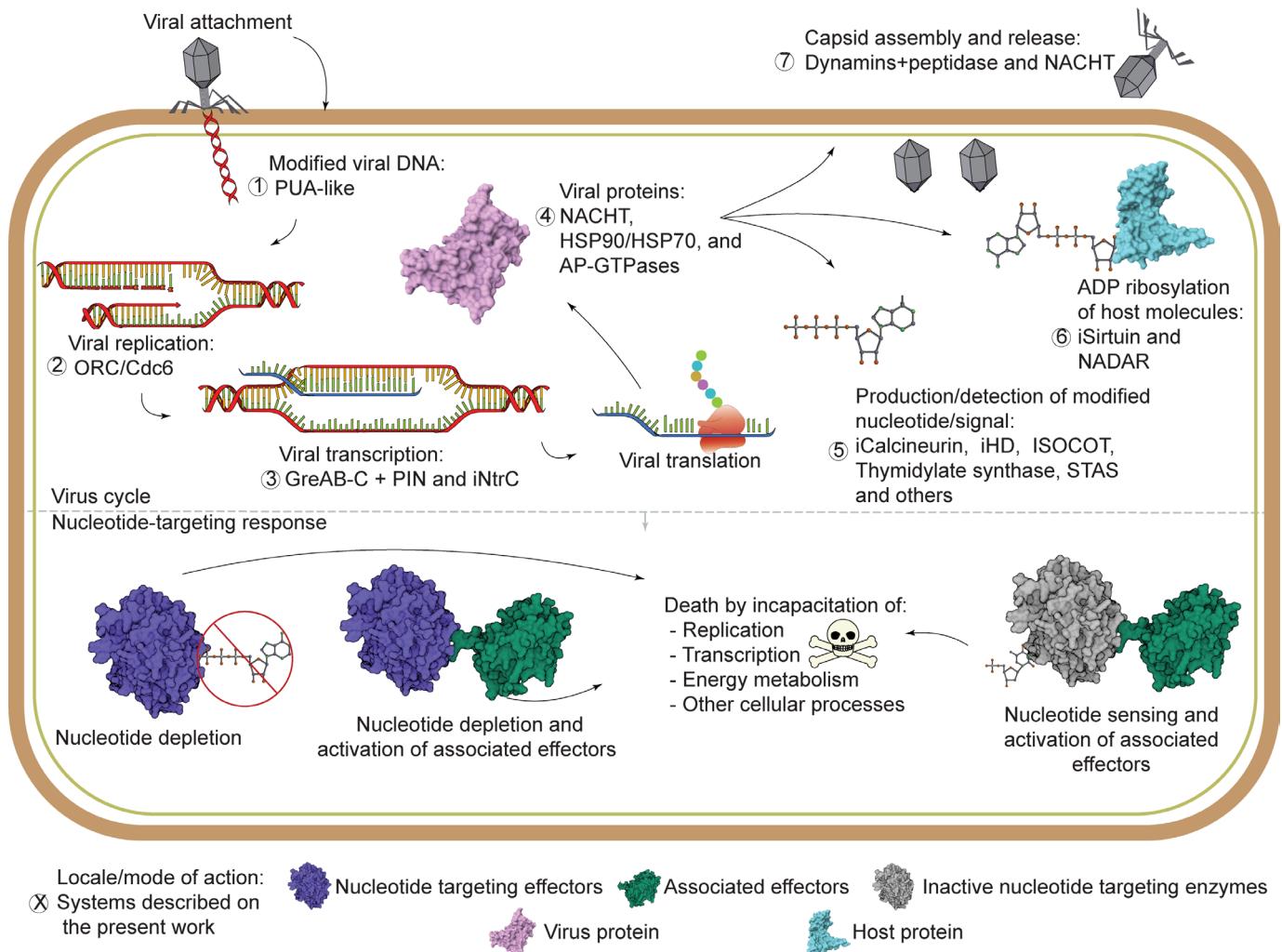


Functionally comparable but evolutionarily distinct nucleotide-targeting effectors help identify conserved paradigms across diverse immune systems

SUPPLEMENTARY MATERIAL

Gianlucca G. Nicastro, A. Maxwell Burroughs, Lakshminarayan M Iyer, and L Aravind

Computational Biology Branch, Intramural Research Program, National Library of Medicine,
National Institutes of Health, Bethesda, MD 20894



You can also access this content through a dedicated website, which can be found at:
<https://aravindlab.github.io/NTdegrade/>

TABLE OF CONTENTS

I. Supplementary Figures.

- A. Supplementary Figure S1 - Permuted GreAB-C structure model.
- B. Supplementary Figure S2 - Double PIN structure model.
- C. Supplementary Figure S3 - Phylogenetic tree of the Orc/Cdc6-like proteins.
- D. Supplementary Figure S4 - Sequence logos of different Calcineurin-like phosphoesterase families.
- E. Supplementary Figure S5 - Activity site and conserved residues from the Calcineurin-CE domain.
- F. Supplementary Figure S6 - Activity site and conserved residues from the HD-CE domain.
- G. Supplementary Figure S7 - Activity site and conserved residues from the PRTase-CE domain.
- H. Supplementary Figure S8 - AlphaFold confidence score for the models produced on the present work.

II. Supplementary Data

A. Genome contextual data

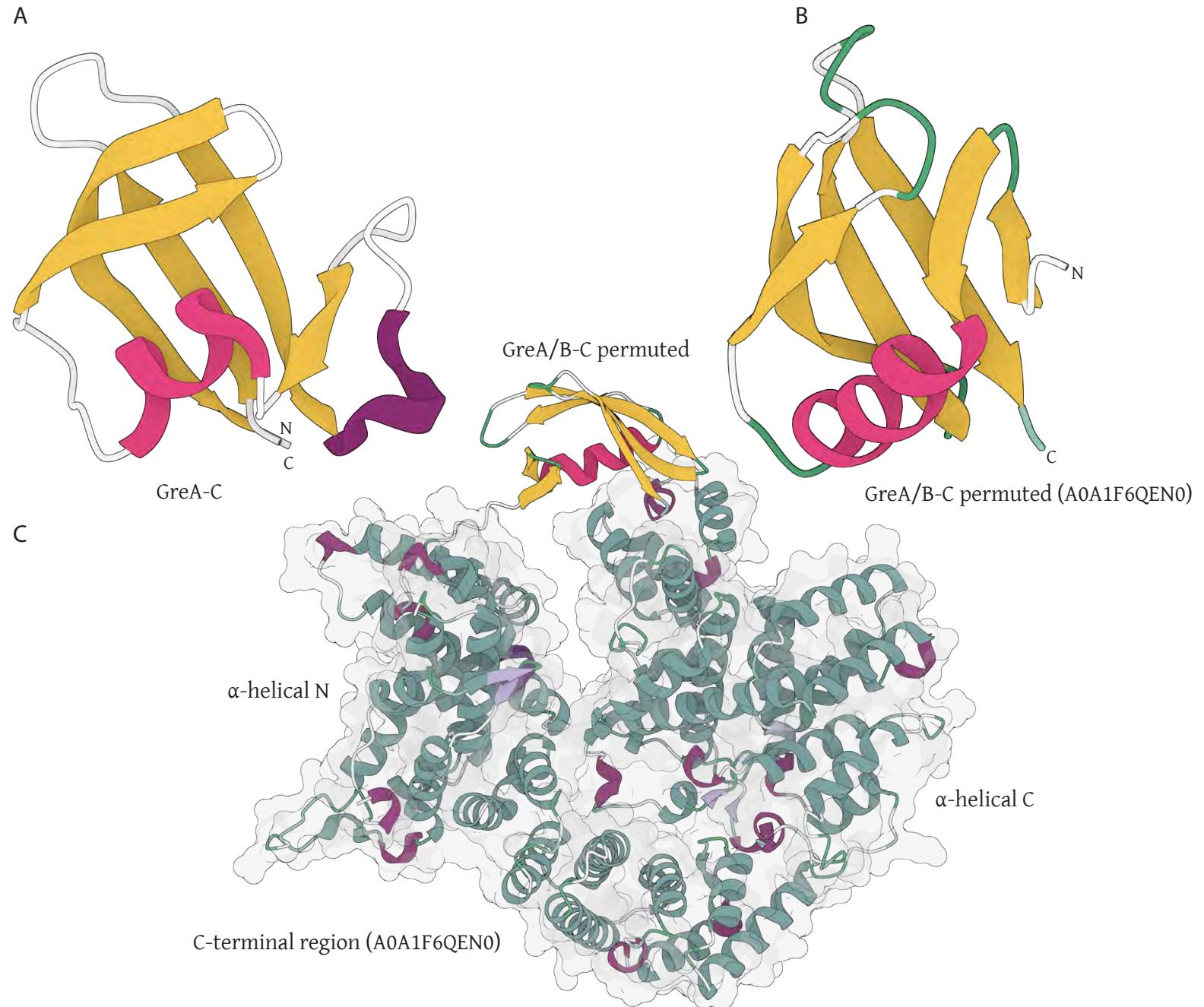
- 1. AEP-1D
- 2. AP-ATPase
- 3. AP-GTPase
- 4. Dynamin
- 5. GreAB-C
- 6. iHD
- 7. HSP70
- 8. HSP90
- 9. ISOCOT
- 10. MNS
- 11. NACHT
- 12. Nadar
- 13. ORC-Cdc6(Calcineurin-CE)
- 14. ORC-Cdc6(HD-CE)
- 15. ORC-Cdc6 (PRTase-CE)
- 16. ORC-Cdc6 (TIR)
- 17. ORC-Cdc6(SIR2)
- 18. ORC-Cdc6(Kinases)
- 19. ORC-Cdc6(Others)
- 20. PUA-like
- 21. ParB-CE
- 22. ParB
- 23. Pepco
- 24. PolBetaCCAA-like
- 25. RVT
- 26. SAMHD1
- 27. nSTAND4
- 28. nSTAND5
- 29. iSTAND3
- 30. iSirtuin
- 31. STAS big protein
- 32. STAS
- 33. Thymidylate synthase
- 34. iThymidylate synthase

B. Multiple sequence alignments

- 1. Calcineurin-CE
- 2. HD-CE
- 3. PRTase-CE
- 4. PUA-like-CE
- 5. iSirtuin-CE
- 6. Calcineurin-CE+iThymidylate-synthase
- 7. AEP-1D
- 8. REC+Calcineurin-CE
- 9. STAS+HisKin
- 10. STAS+PRTase
- 11. STAS+HisKin+iPRTase+PRTase
- 12. Pepco
- 13. iCalcineurin-CE
- 14. iHD

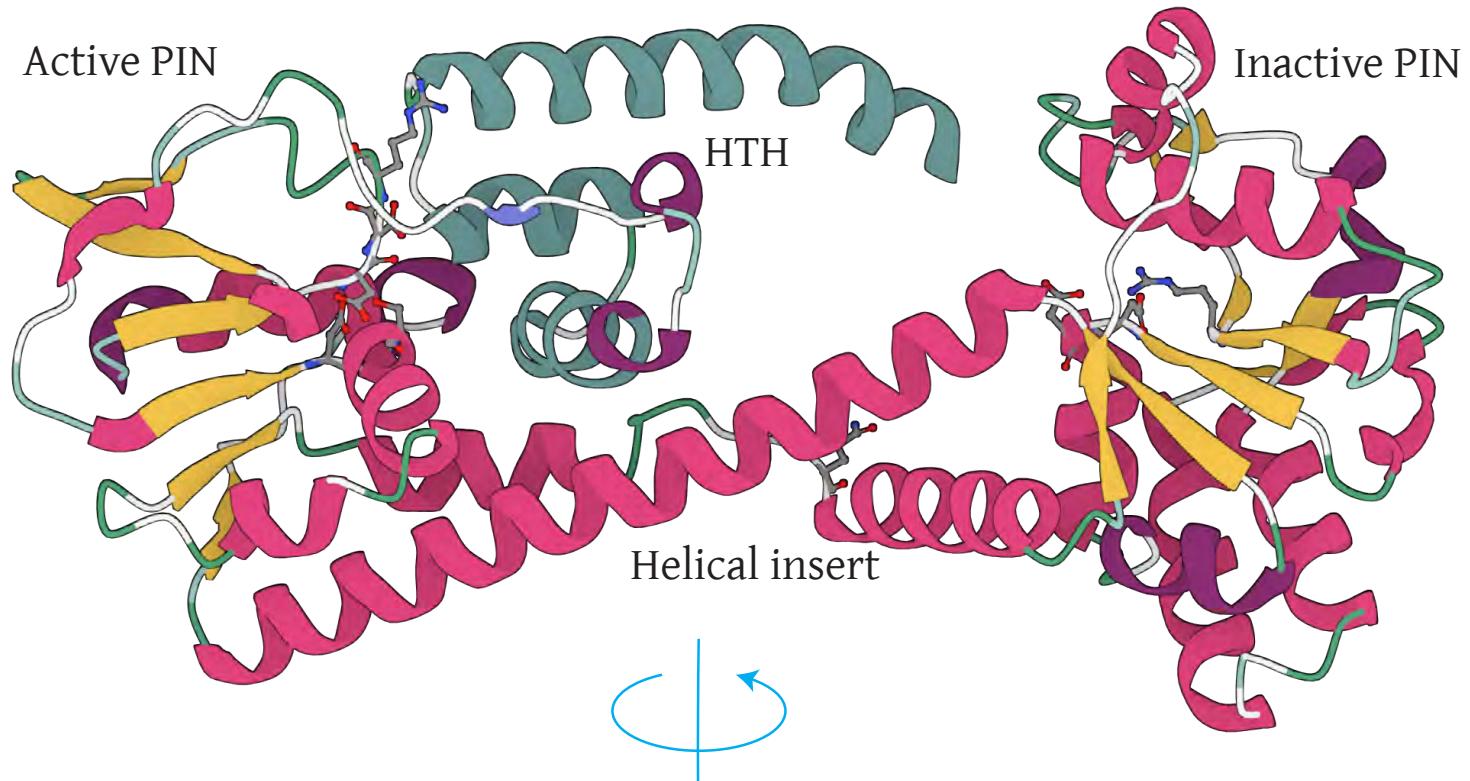
C. Phylogenetic Tree

- 1. Calcineurin-CE
- 2. HD-CE
- 3. PRTase-CE
- 4. ORC-Cdc6

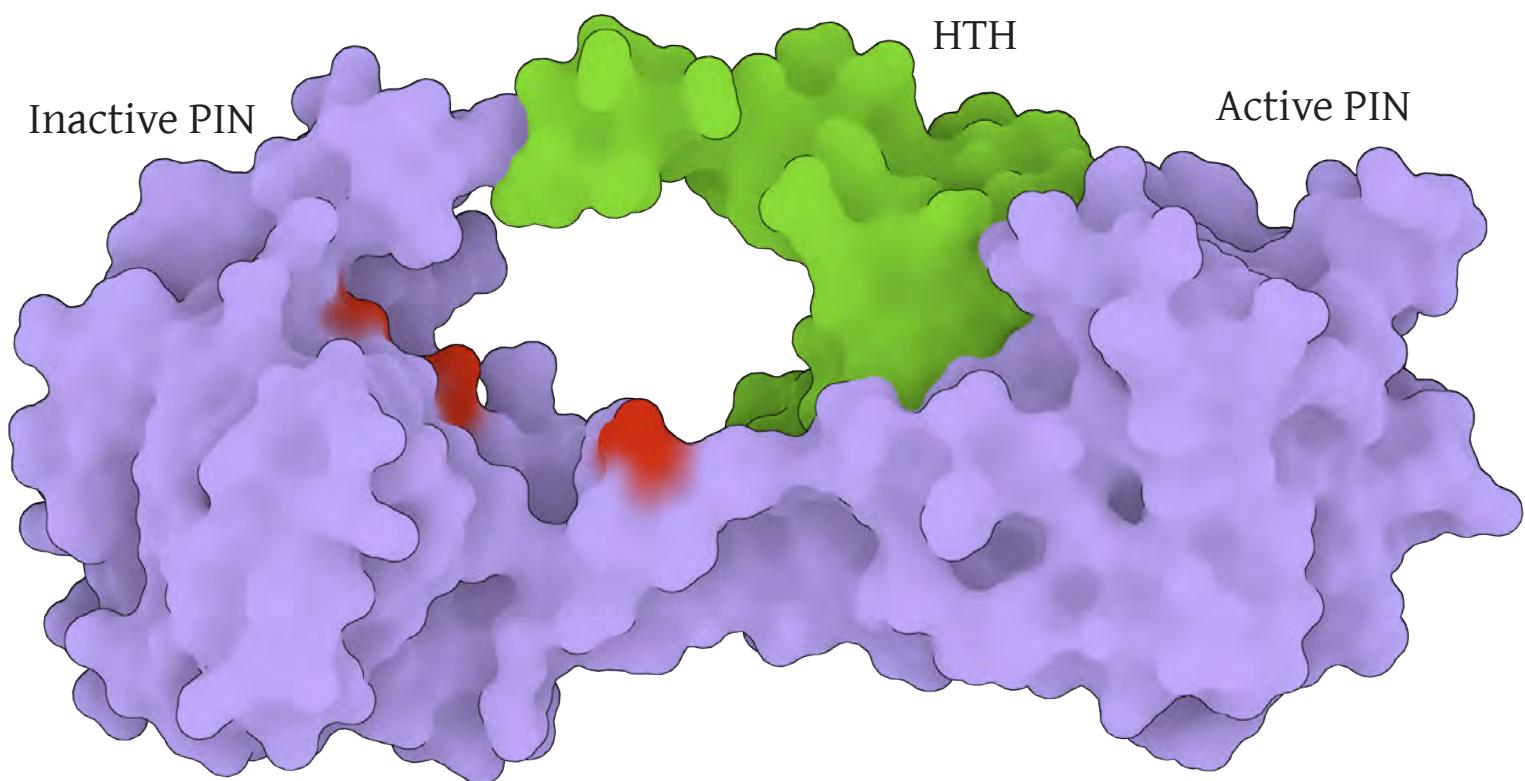


Supplementary Figure S1. Permuted GreAB-C structure model. (A) Structure of the previously identified GreAB-C domain. (B) Alphafold model of a permuted GreAB-C domain. (C) Context of the permuted GreAB-C domain embedded in a sequence of divergent TPRs, forming a toroidal structure.

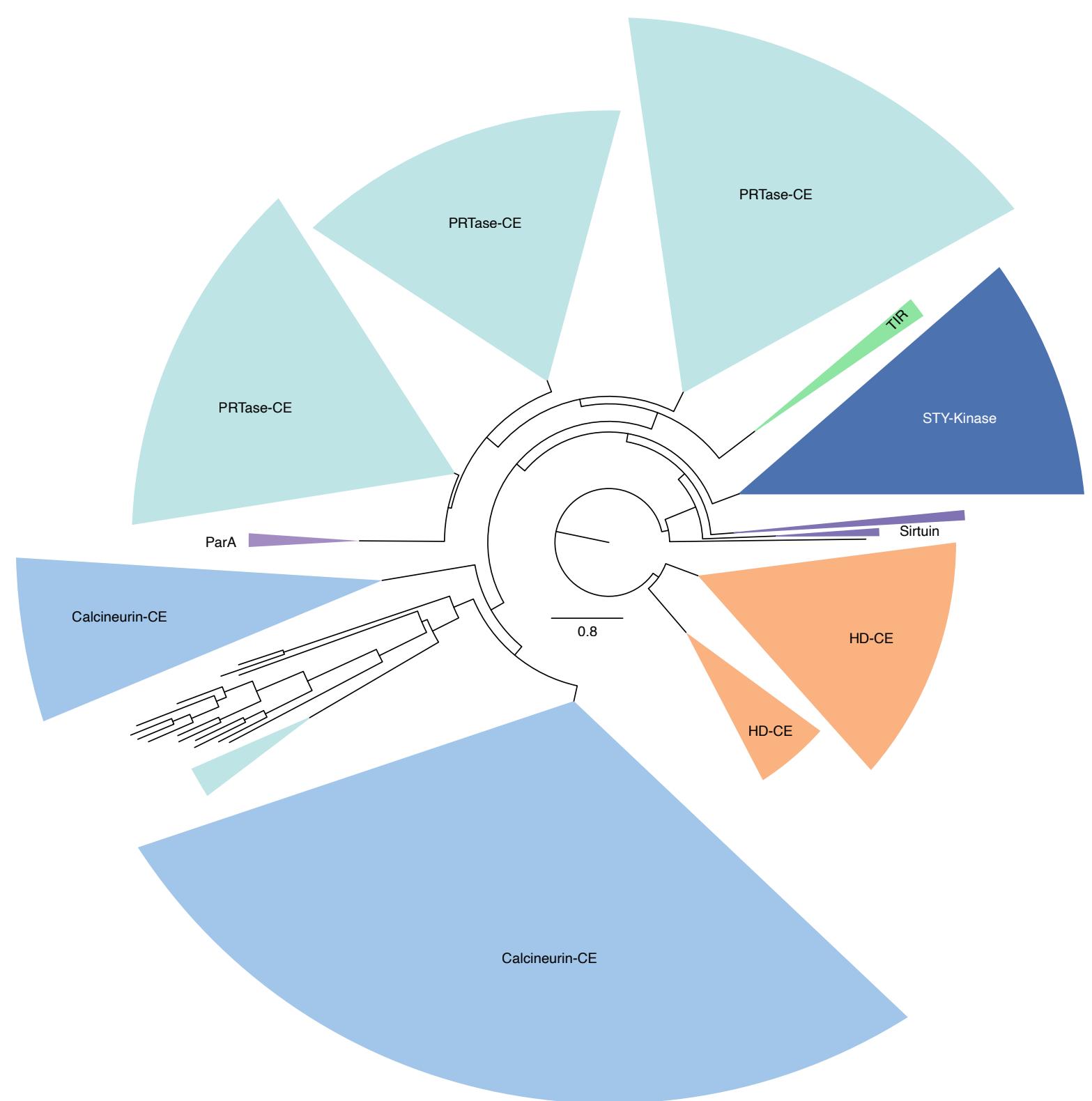
A



B



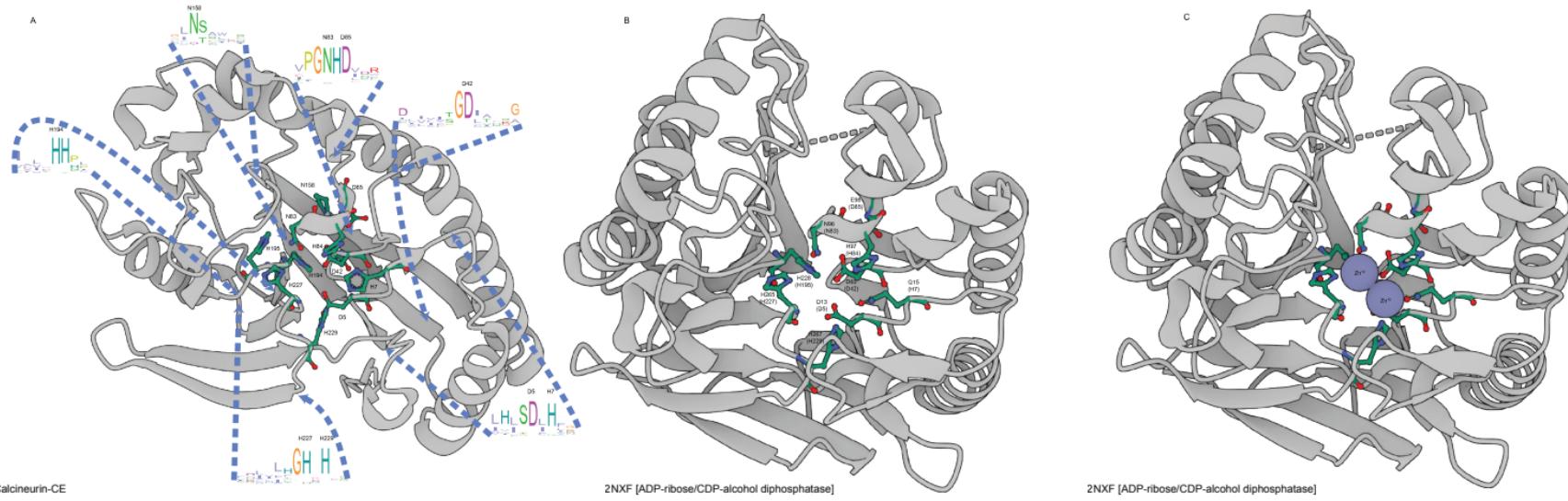
Supplementary Figure S2. Double PIN structure model. (A) cartoon representation of the double PIN domain. β -strands are depicted in yellow, α -helices in dark pink, and the HTH domain in dark cyan. Catalytic residues are shown in ball-and-stick representation. (B) Molecular surface representation demonstrating the occlusion of the active PIN site by the HTH domain. The two PIN domains are represented in purple, the HTH domain in green, and the position of the inactive catalytic residues of the inactive PIN domain is depicted in red.



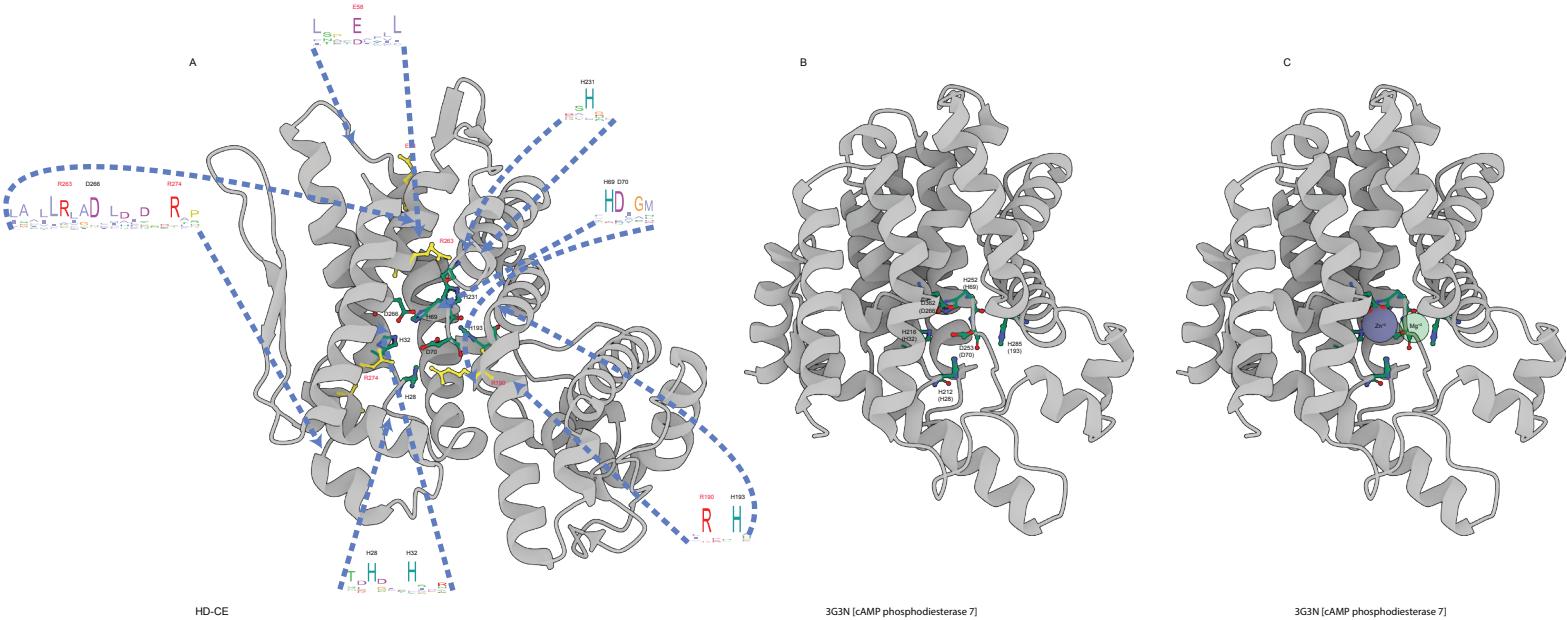
Supplementary Figure S3. Phylogenetic tree of the Orc/Cdc6-like proteins. The tree was constructed with IQtree2 using ultrafast bootstrap and SH-aLRT test as support values. The clades are labeled according to associated effector domains present in the genomic neighborhood



Supplementary Figure S4. Sequence logos of different Calcineurin-like phosphoesterase families. Sequences from various families of calcineurin-like domains were gathered using previously characterized PDB sequences as seeds. These sequences were aligned using the Mafft local pair algorithm, and the logos representing different families were visualized using the Jalview software. Conserved motifs I-V are depicted with boxes.

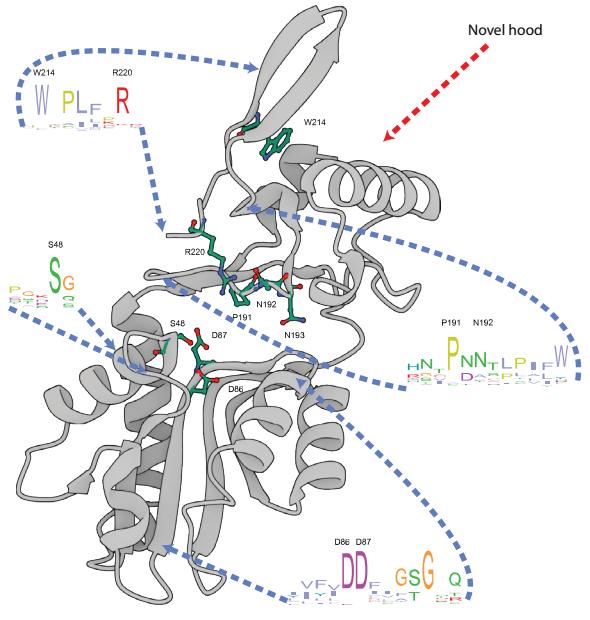


Supplementary Figure S5. Activity site and conserved residues from the Calcineurin-CE domain. (A) Alphafold modeling of the Calcineurin-CE domain (WP_080633877.1). The residues conserved in the activity sites are shown in ball and stick representations. The synapomorphic asparagine is colored in orange. Sequence logos of the conserved regions are shown and the dotted lines indicate the start and end position of the logo in the structural model. (B and C). Activity site of the experimental crystallized ADP-ribose/C-ADP-alcohol diphosphatase domain from the PDB 2NXF without and with the Zn²⁺ ion respectively.



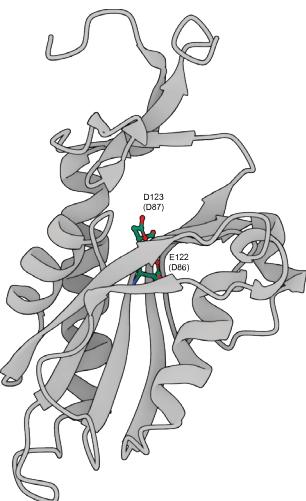
Supplementary Figure S6. Activity site and conserved residues from the HD-CE domain. (A) Alphafold modeling of the HD-CE domain (CZQ95560.1). The residues conserved in the activity sites are shown in ball and stick representations. The synapomorphic residues are colored yellow. Sequence logos of the conserved regions are shown and the dotted lines indicate the start and end position of the logo in the structural model. (B and C). Activity site of the experimental crystallized cAMP phosphodiesterase 7 domain from the PDB 3G3N without and with the Zn⁺² and MG⁺² ions respectively.

A



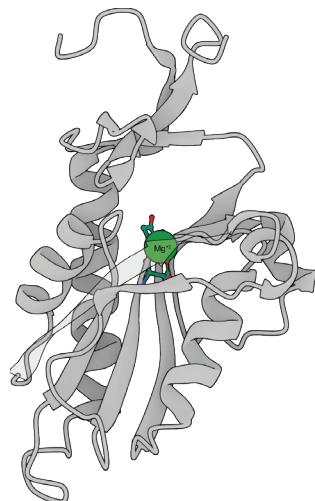
PRTase-CE

B



5KNQ [hypoxanthine guanine PRTase]

C



5KNQ [hypoxanthine guanine PRTase]

Supplementary Figure S7. Activity site and conserved residues from the PRTase-CE domain.
 (A) Alphafold modeling of the PRTase-CE domain (NPV02125.1). The residues conserved in the activity sites are shown in ball and stick representations. Sequence logos of the conserved regions are shown and the dotted lines indicate the start and end position of the logo in the structural model. (B and C). Activity site of the experimental crystallized hypoxanthine guanine PRTase domain from the PDB 5KNQ without and with the Mg⁺² ion respectively.

Figure 3A Calcineurin-CE

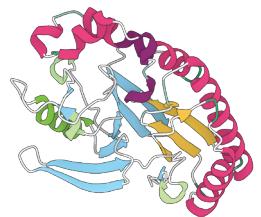
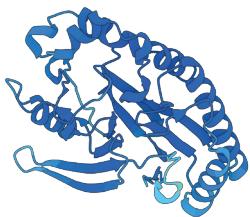


Figure 4G iSirtuin

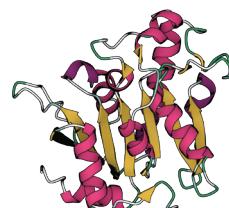
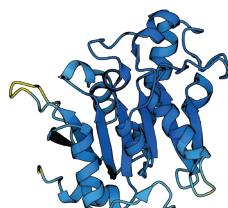


Figure 3B HD-CE

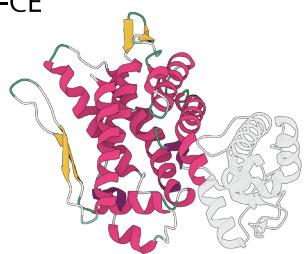
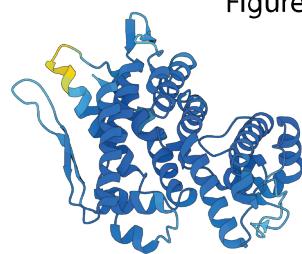


Figure 5G PRTase

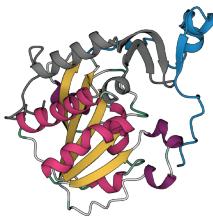
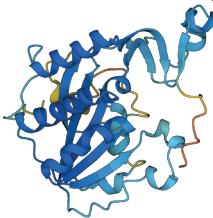


Figure 3C PRTase-CE

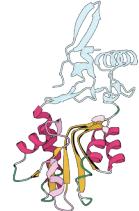


Figure 6B HSP90 system with C-terminal wHTH

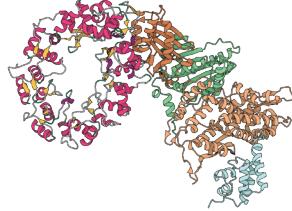
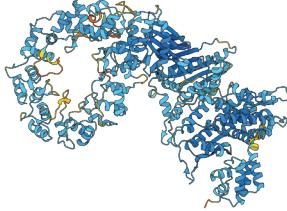


Figure 4B Cdc6•double-CARF•Calcineurin-CE complex

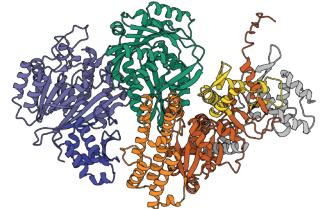
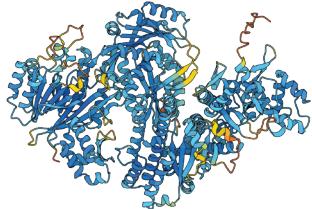


Figure 6E Pepco•Caspase complex



Figure 4D Pua-like

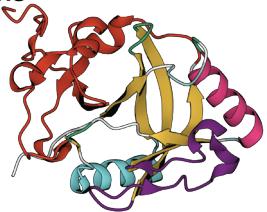
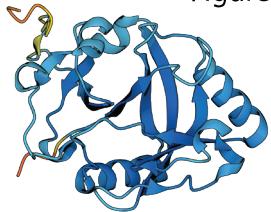
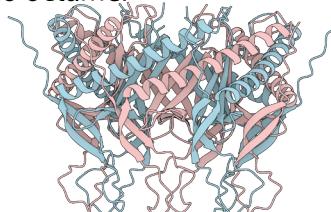
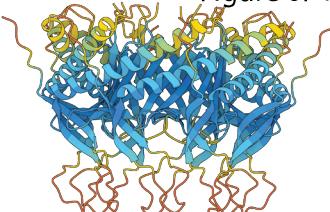


Figure 6F Pepco octamer



Very High (pLDDT > 90)

Confident (90 > pLDDT > 70)

Low (70 > pLDDT > 50)

Very Low(pLDDT < 50)

Supplementary Figure S8. AlphaFold confidence score for the models produced on the present work. On the left the structural models are colored by the pLDDT scores. On the right are the structural models as colored in the figures presented in the main text. Dark blue, light blue, yellow and orange represent a very high, confident, low and very low prediction respectively.

This block contains the full sequence of the 16S rRNA gene from the provided image, spanning approximately 1500 pages of text.

... (content of the file is extremely long and repetitive, containing thousands of entries for various species and their interactions)

WP_07131497.1	TPR_Malt+TPR	NZ_JDVO1000024.1	repeat_region>*TPR-Malt+TPR->
WP_07595594.1	TPR_Malt	NZ_MKOR0100001.1	Pseudonocardiales
KU069334.1	ISOCOT	GCF_001940455.1	Pseudonocardiales
MF854394.1	Rhodanese+ISOCOT	GCA_001516065.1	Actinomycetales
WP_12958116.1	ISOCOT	GCA_0011322615.1	Actinomycetales
MBO33775.1	ISOCOT	GCF_002725625.1	Actinomycetales
MCC6904406.1	ISOCOT	GCA_00245245.1	Actinomycetales
NHJ48206.1	ISOCOT	GCA_007725775.1	Actinomycetales
WP_145614039.1	ISOCOT	GCF_002857665.1	Actinomycetales
MBL7139014.1	ISOCOT	GCA_002857665.1	Actinomycetales
PLX23386.1	ISOCOT	GCF_002869335.1	Actinomycetales
MYX92447.1	ISOCOT	GCA_002869335.1	Actinomycetales
SDR13757.1	TPR_Malt+TPR	GCF_002869335.1	Actinomycetales
NJM11238.1	HD-CE+HSP90	GCA_002869335.1	Actinomycetales
HDD72717.1	ISOCOT	GCF_002869335.1	Actinomycetales
MD6402831.1	ISOCOT	GCA_002869335.1	Actinomycetales
RLJ07437.1	ISOCOT	GCF_002869335.1	Actinomycetales
RLJ09226.1	ISOCOT	GCA_002869335.1	Actinomycetales
RLI19056.1	ISOCOT	GCF_002869335.1	Actinomycetales
OSD41903.1	ISOCOT	GCF_002869335.1	Actinomycetales
MBU0761692.1	ISOCOT	GCA_002869335.1	Actinomycetales
UCE95506.1	ISOCOT	GCF_002869335.1	Actinomycetales
MB57642105.1	ISOCOT	GCA_002869335.1	Actinomycetales
KPV88718.1	ISOCOT	GCF_002869335.1	Actinomycetales
PIE78416.1	SIR2_+ISOCOT	GCF_002869335.1	Actinomycetales
WP_156469446.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
NAQ22496.1	HD-CE+HSP90	GCA_013204585.1	Actinomycetales
WP_090171786.1	TPR_Malt	GCF_002869335.1	Actinomycetales
MBK8392041.1	HD-CE	GCA_016711515.1	Actinomycetales
MBK7808942.1	HD-CE+HSP90	GCA_016714445.1	Actinomycetales
MBK8621213.1	HD-CE+HSP90	GCA_016714895.1	Actinomycetales
WP_006981066.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
WP_144247495.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
WP_112568691.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
WP_179415836.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
WP_149874156.1	Calcineurin-CE+TPR	GCF_002869335.1	Actinomycetales
WP_184629146.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
WP_136878771.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
WP_187591760.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
WP_047801218.1	Calcineurin-CE+TPR	GCF_002869335.1	Actinomycetales
WP_090501330.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
RYE18664.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
WP_130143368.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
WP_148921152.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
WP_003001339.1	TPR_Malt	GCF_002869335.1	Actinomycetales
MPB3709107.1	HD-CE	GCA_017940045.1	Actinomycetales
MBD3599449.1	HD-CE+HSP90	GCA_017940045.1	Actinomycetales
MRR500537.1	HD-CE+HSP90	GCA_017940045.1	Actinomycetales
MPS328502.1	HD-CE+HSP90	GCA_017940045.1	Actinomycetales
WP_135647357.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
NQE05492.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
NQE5392.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
RLG38728.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
WP_080506053.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
NLK32867.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
WP_0135056.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
WP_0480378375.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
HIH74739.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
MCD486236.1	HD-CE+MORC	GCF_002869335.1	Actinomycetales
WP_214421011.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
MPB787036.1	HD-CE	GCA_017940045.1	Actinomycetales
WP_051939857.1	TPR_Malt+TPR+TPR	GCF_002869335.1	Actinomycetales
GFE0654.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
GHE35110.1	HD-CE+MORC	GCF_002869335.1	Actinomycetales
WP_098244913.1	TPR_Malt+TPR	GCF_002869335.1	Actinomycetales
GGY57364.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
WP_046506703.1	TPR_Malt+TPR	GCF_002869335.1	Actinomycetales
WP_055704844.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
NUR41459.1	HSP90	GCF_002869335.1	Actinomycetales
WP_21273682.1	TPR_Malt+TPR+TPR	GCF_002869335.1	Actinomycetales
WP_224283077.1	TPR_Malt	GCF_002869335.1	Actinomycetales
WP_053754779.1	HSP90+DrdHyd	GCF_002869335.1	Actinomycetales
WP_159049678.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
WP_152772032.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
RO08776.1	PARM-HD+CE+HSP90	GCF_002869335.1	Actinomycetales
WP_15155879.1	TPR_Malt+TPR+TPR	GCF_002869335.1	Actinomycetales
WP_144390269.1	TPR_Malt	GCF_002869335.1	Actinomycetales
WP_225271635.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
WP_055502833.1	TPR_Malt	GCF_002869335.1	Actinomycetales
WP_189235199.1	HD-CE+MORC	GCF_002869335.1	Actinomycetales
SDR13757.1	TPR_Malt+TPR	GCF_002869335.1	Actinomycetales
NJM11238.1	HD-CE	GCA_012031845.1	Actinomycetales
HAJ26871.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
WP_01763278.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
RL5181267.1	HD-CE+HSP90+Hete-N1	GCF_002869335.1	Actinomycetales
MBI1462517.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
TAN35343.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
MBD9278363.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
MAS96472.1	HD-CE+HSP90	GCF_002869335.1	Actinomycetales
PAAG0100055.1		GCF_002869335.1	Actinomycetales
#ISOCOT		GCF_002869335.1	Actinomycetales
pid	arch	nucleotideID:position	assembly
HDP70471.1	ISOCOT	DSBK01000096.1:1-5295	DSBK01000096.1
KU069334.1	ISOCOT		DRMN0100001.1
MF854394.1	Rhodanese+ISOCOT	DRMN01000041.1:1-5196	NZ_SLU0100006.1
WP_12958116.1	ISOCOT	NZ_SLU0100006.1:1-41676-5193	PBO10000145.1
MBO33775.1	ISOCOT	PBO10000145.1:1-9235	JADZAD01000092.1
MCC6904406.1	ISOCOT	JADZAD01000092.1:1-7379	VIKN01000099.1
NHJ48206.1	ISOCOT	VIKNN01000099.1:1-4169-24562	NZ_JARAF01000005.1
WP_145614039.1	ISOCOT	NZ_JARAF01000005.1:1-32937-341251	
MBL7139014.1	ISOCOT	JADH01000008.1:1-5608	
PLX23386.1	ISOCOT	PKTE01000086.1:1-222-8926	
MYX92447.1	ISOCOT	VXR01000034.1:1-2711-1395	
SDR13757.1	TPR_Malt+TPR	GCF_002869335.1:1-392884-394618	
NJM11238.1	HD-CE	GCA_003751225.1:1-3772-18737	
HAJ26871.1	HD-CE+HSP90	GCF_002869335.1:1-406-13149	
WP_01763278.1	HD-CE+HSP90	GCF_002869335.1:1-44390-47185	
RL5181267.1	HD-CE+HSP90+Hete-N1	GCF_002869335.1:1-500-16345	
MBI1462517.1	HD-CE+HSP90	GCF_002869335.1:1-519	
TAN35343.1	HD-CE+HSP90	GCF_002869335.1:1-7610-8749	
MBD9278363.1	HD-CE+HSP90	GCF_002869335.1:1-8309-9700	
MAS96472.1	HD-CE+HSP90	GCF_002869335.1:1-908-20623	
PAAG0100055.1		GCF_002869335.1:1-908-20623	
#ISOCOT		GCF_002869335.1:1-908-20623	
pid	arch	nucleotideID:position	assembly
HDP70471.1	ISOCOT	DSBK01000096.1:1-5295	DSBK01000096.1
KU069334.1	ISOCOT		DRMN0100001.1
MF854394.1	Rhodanese+ISOCOT	DRMN01000041.1:1-5196	NZ_SLU0100006.1
WP_12958116.1	ISOCOT	NZ_SLU0100006.1:1-41676-5193	PBO10000145.1
MBO33775.1	ISOCOT	PBO10000145.1:1-9235	JADZAD01000092.1
MCC6904406.1	ISOCOT	JADZAD01000092.1:1-7379	VIKN01000099.1
NHJ48206.1	ISOCOT	VIKNN01000099.1:1-4169-24562	NZ_JARAF01000005.1
WP_145614039.1	ISOCOT	NZ_JARAF01000005.1:1-32937-341251	
MBL7139014.1	ISOCOT	JADH01000008.1:1-5608	
PLX23386.1	ISOCOT	PKTE01000086.1:1-222-8926	

Microbial community analysis									


... (content of the file is too large to be displayed here)

This block contains the full sequence of the 16S rRNA gene from the provided file, starting at position 1 and ending at position 1000.

This block contains the full sequence of the 16S rRNA gene from the provided image, spanning from position 1 to approximately 980.


WP_1688331.1	ORC-CDCG-like	GCF_01995614.5	108	*ORC-CDCG-like->	Vibrio parahemolyticus
WP_1688331.1	ORC-CDCG-like	GCF_01995614.5	108	*ORC-CDCG-like->XARF	Vibrio vulnificus
WP_165387721.1	ORC-CDCG-like	GCF_00424575.3	542	*ORC-CDCG-like->XARF	Vibrio crassostreae
NP_123308974.1	ORC-CDCG-like	GCF_00751457.3	142	*ORC-CDCG-like->XARF	Vibrio cholerae
PTD26861.1	ORC-CDCG-like	GCF_00024524.3	153	*ORC-CDCG-like->XARF	Vibrio cholerae
TIF39892.1	ORC-CDCG-like	GCA_08682465.3	147	*ORC-CDCG-like->XARF	Vibrio cholerae
VTPA01080001.1	ORC-CDCG-like	GCA_08682465.3	147	*ORC-CDCG-like->XARF	Vibrio cholerae
ALAB8577.1	ORC-CDCG-like	GCF_011395.1	11884537-198059	PSE->?>SbC_PSE->ORC-CDCG-like->PSE->	Stenotrophomonas maltophilia
CAS9623453.1	ORC-CDCG-like	CA117081080037-1:5408-4433	559	PSE->?>SbC_PSE->ORC-CDCG-like->PSE->	uncultured Gammaproteobacterium
CA117081080037-1:5408-4433	ORC-CDCG-like	GCA_09831404.1	559	PSE->?>SbC_PSE->ORC-CDCG-like->PSE->	Solomonis sp. 1885
MBV8225564.1	ORC-CDCG-like	JAFATU01080005:1-1221-3464	604	PSE->?>SbC_PSE->ORC-CDCG-like->PSE->	Stenotrophomonas maltophilia
MBV8225564.1	ORC-CDCG-like	GCA_01923653.1	1221-3464	PSE->?>SbC_PSE->ORC-CDCG-like->PSE->	Verrucimonas obliqua bacterium
MB19159874.1	ORC-CDCG-like	JAEH0110800058:1-89938-108252	611	PSE->?>SbC_PSE->ORC-CDCG-like->PSE->	Verrucimonas sp.
MB19159874.1	ORC-CDCG-like	GCA_01679538.1	603	PSE->?>SbC_PSE->ORC-CDCG-like->PSE->	Verrucimonas bacterium
JAD090021080004:1-16603-16761	ORC-CDCG-like	GCA_01679538.1	603	Pse_Ingap_uise->>REL->CARF-double->ERNA->DNA_poi2d0NA_poi2_gamma3->Yba_BNA_b6d>RecR>Toprim_4->	Verrucimonas

candidate division KSL1
candidate division KSL1
candidate division KSL1
candidate division KSL1











1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

