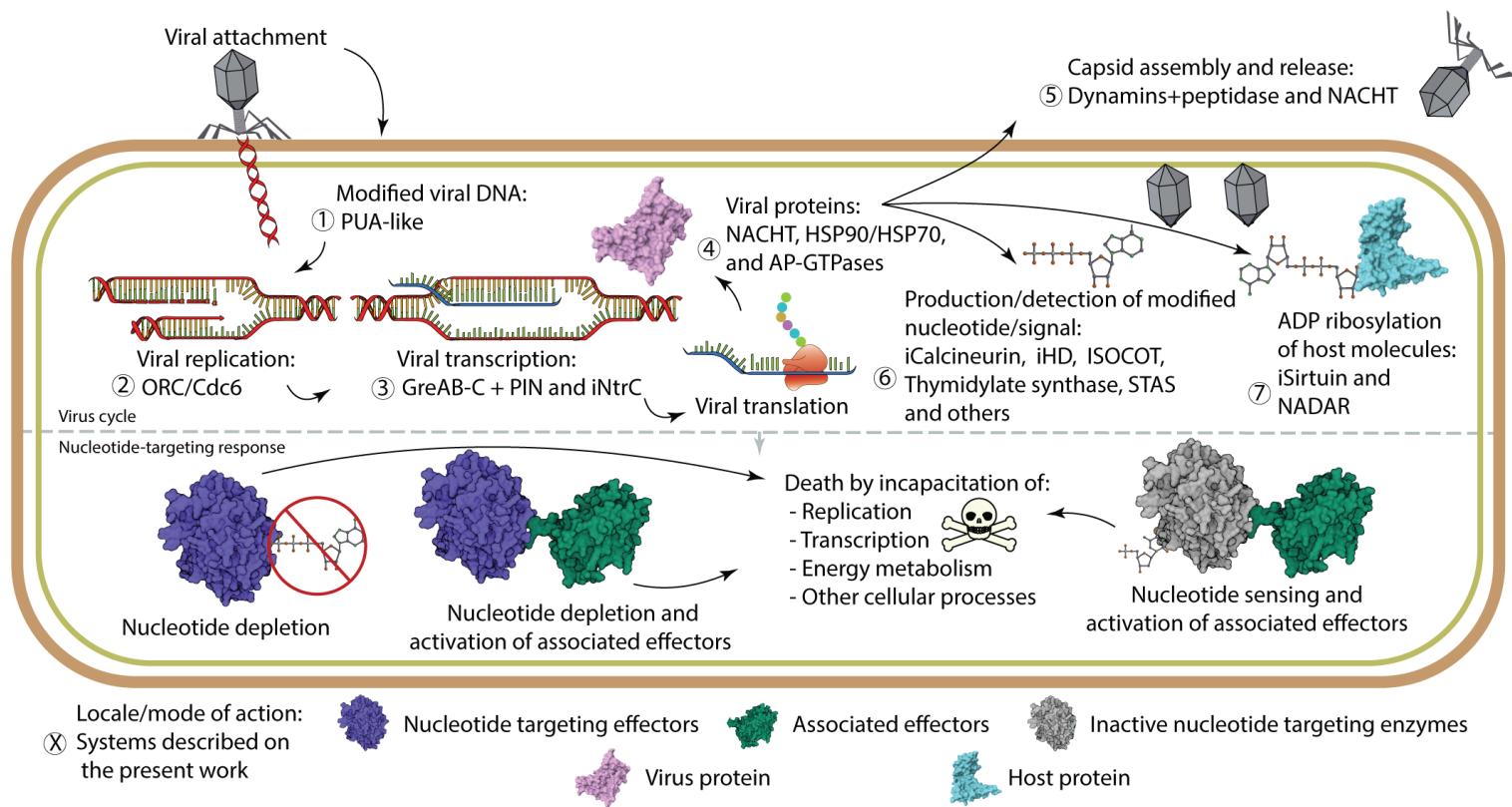


Functionally comparable but evolutionarily distinct nucleotide-targeting effectors help identify conserved paradigms across diverse immune systems

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 1 - Permuted GreAB-C structure model.
- B. Supplementary Figure 2 - Double PIN structure model.
- C. Supplementary Figure 3 - Phylogenetic tree of the Orc/Cdc6-like proteins.
- D. Supplementary Figure 4 - Sequence logos of different Calcineurin-like phosphoesterase families.
- E. Supplementary Figure 5 - Activity site and conserved residues from the Calcineurin-CE domain.
- F. Supplementary Figure 6 - Activity site and conserved residues from the HD-CE domain.
- G. Supplementary Figure 7 - Activity site and conserved residues from the PRTase-CE domain.
- H. Supplementary Figure 8 - AlphaFold confidence score for the models produced on the present work.

II. Supplementary Data

A. Genome contextual data

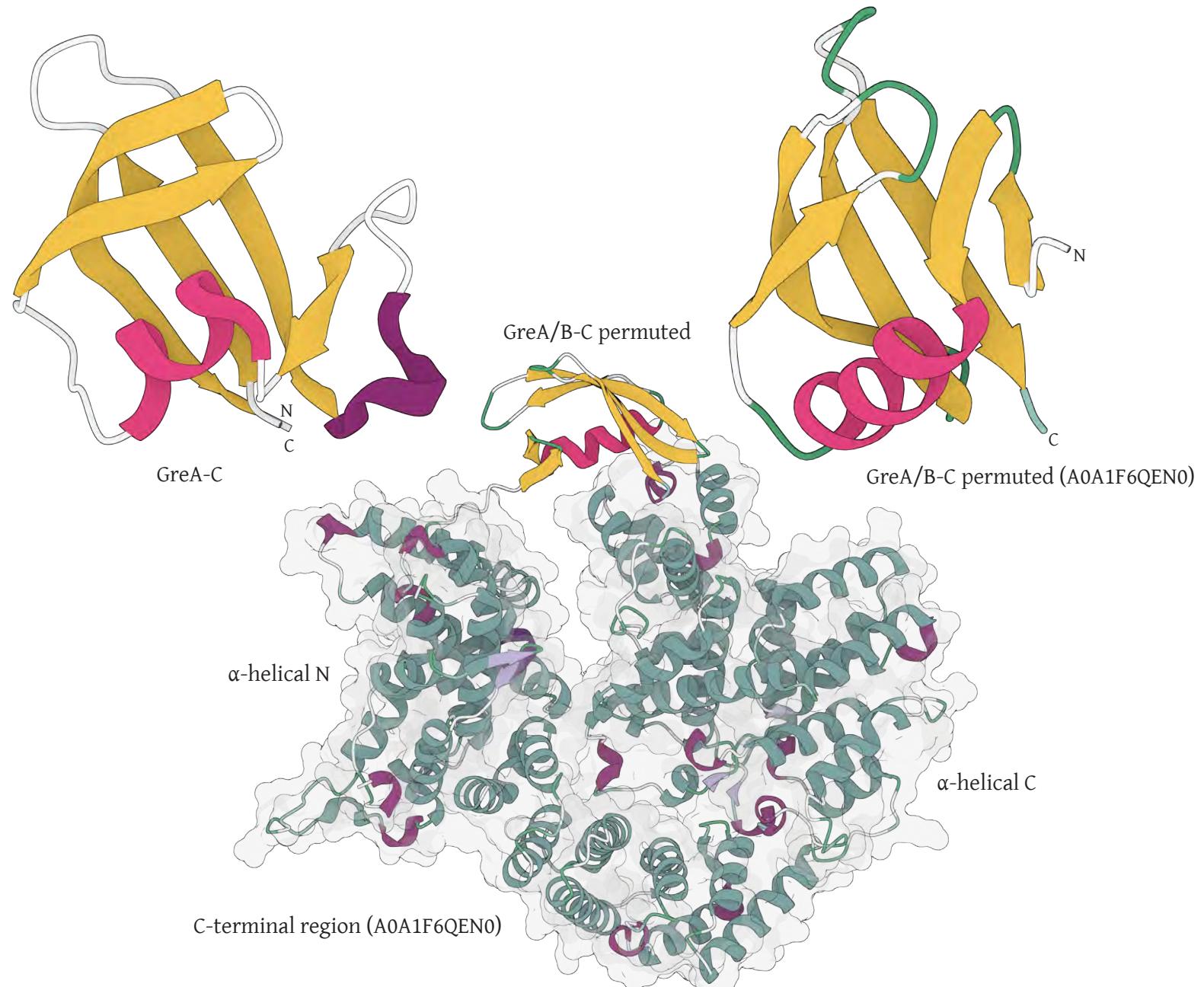
- 1. AEP-1D
- 2. AP-ATPase
- 3. AP-GTPase
- 4. Dynamin
- 5. GreAB-C
- 6. iHD
- 7. HSP70
- 8. HSP90
- 9. ISOCOT
- 10. MNS
- 11. NACHT
- 12. Nadar
- 13. ORC-Cdc6(Calcineurin-CE)
- 14. ORC-Cdc6(HD-CE)
- 15. ORC-Cdc6 (PRTase-CE)
- 16. ORC-Cdc6 (TIR)
- 17. ORC-Cdc6(SIR2)
- 18. ORC-Cdc6(Kinases)
- 19. ORC-Cdc6(Others)
- 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

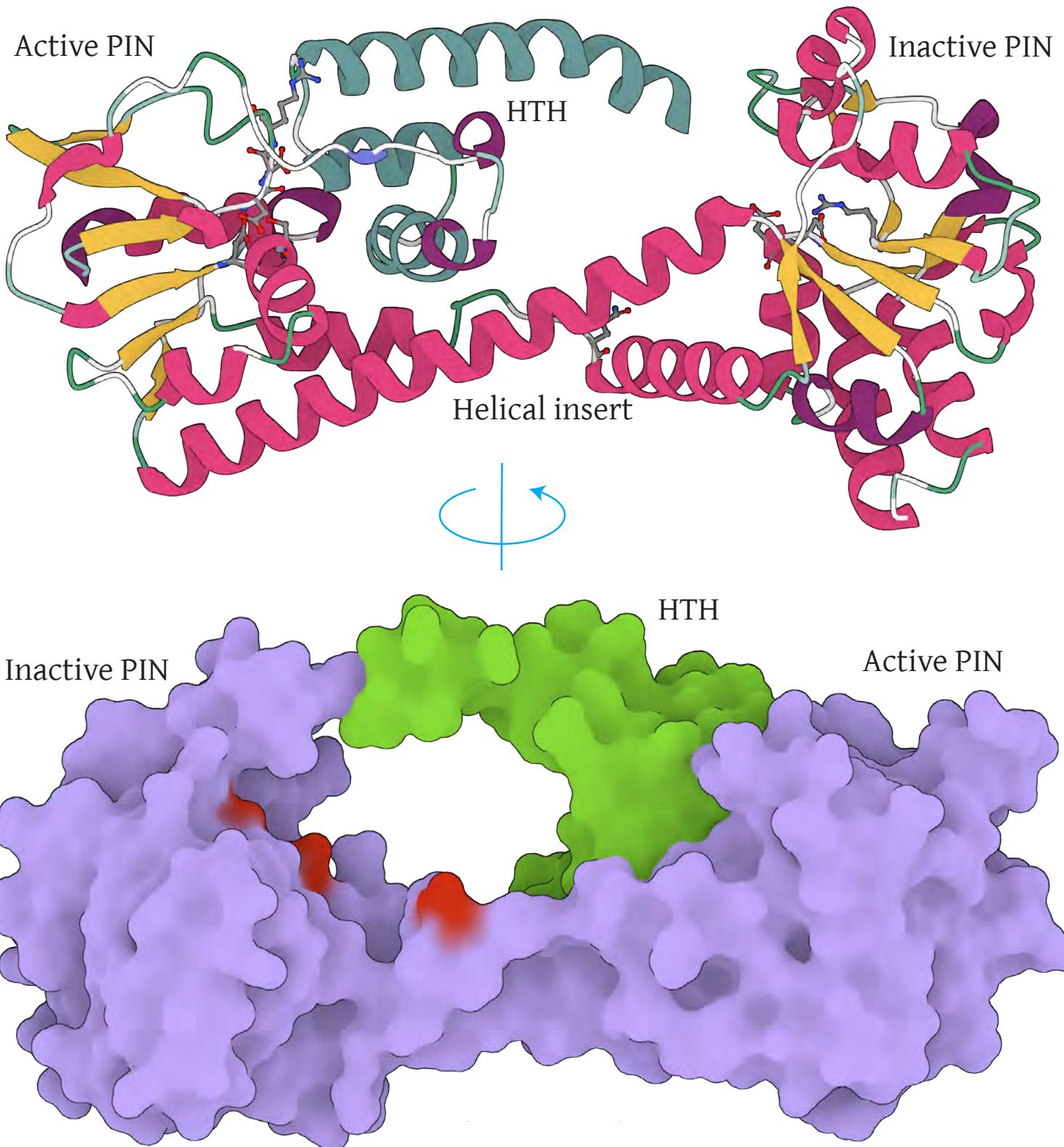
- 1. Calcineurin-CE
- 2. HD-CE
- 3. PRTase-CE
- 4. PUA-like-CE
- 5. iSirtuin-CE
- 6. Calcineurin-CE+iThymidylate-synthase
- 7. AEP-1D
- 8. REC+Calcineurin-CE
- 9. STAS+HisKin
- 10. STAS+PRTase
- 11. STAS+HisKin+iPRTase+PRTase
- 12. Pepco
- 13. iCalcineurin-CE
- 14. iHD

C. Phylogenetic Tree

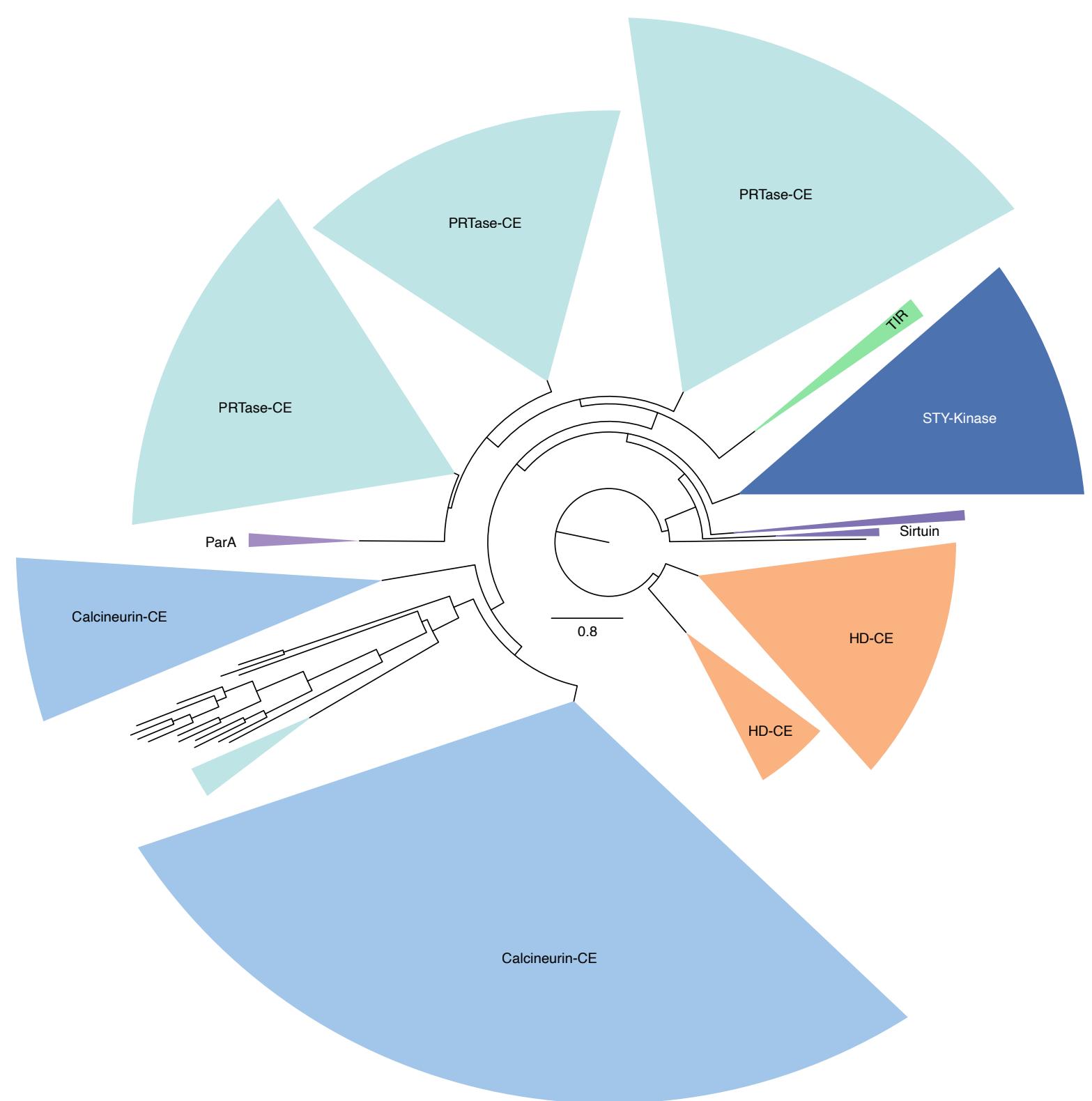
- 1. Calcineurin-CE
- 2. HD-CE
- 3. PRTase-CE
- 4. ORC-Cdc6



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 TPs, forming a toroidal structure.



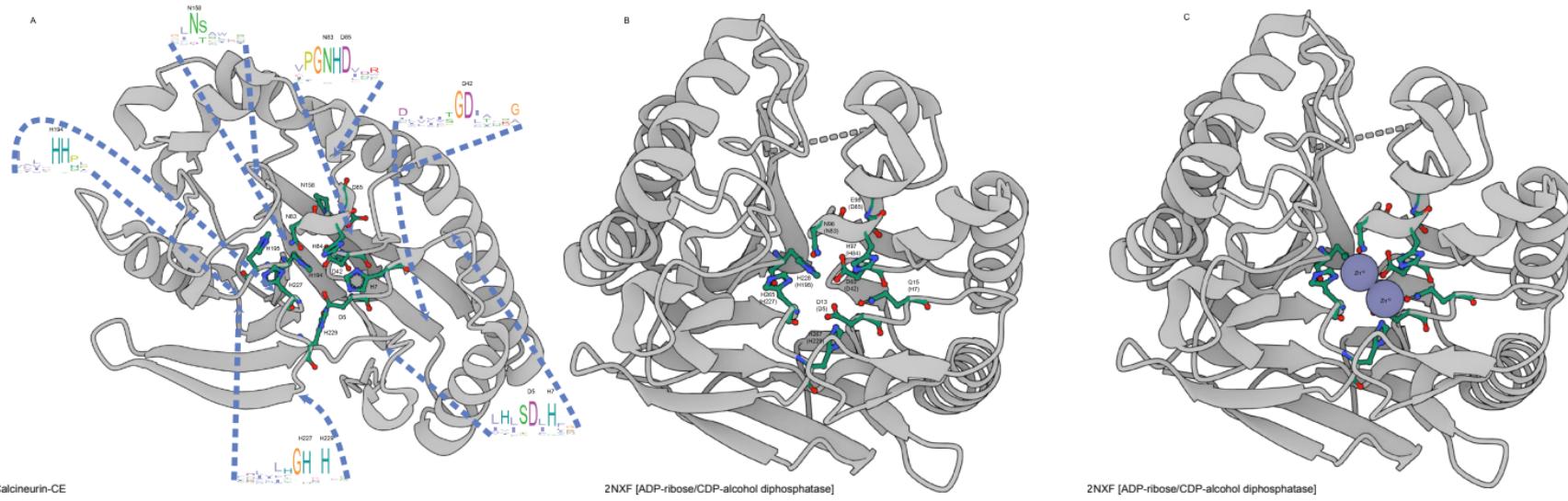
Supplementary Figure S2. Double PIN structure model. (A) cartoon representation of the double PIN domain. β -strands are depicted in yellow, α -helices in dark pink, and the HTH domain in dark cyan. Catalytic residues are shown in ball-and-stick representation. (B) Molecular surface representation demonstrating 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.



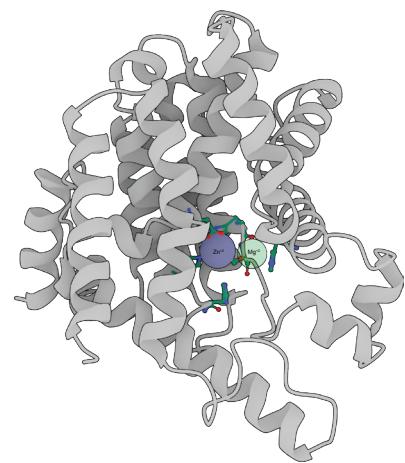
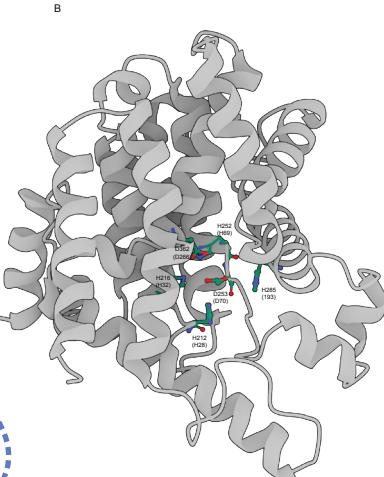
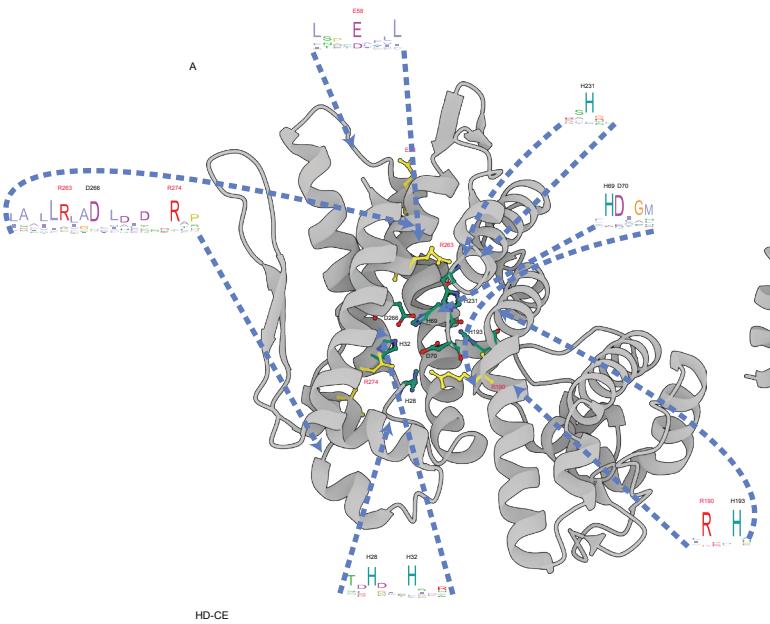
Supplementary Figure S3. Phylogenetic tree of the Orc/Cdc6-like proteins. The tree was constructed with IQtree2 using ultrafast bootstrap and SH-aLRT test as support values. The clades are labeled according to associated effector domains present in the genomic neighborhood



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.

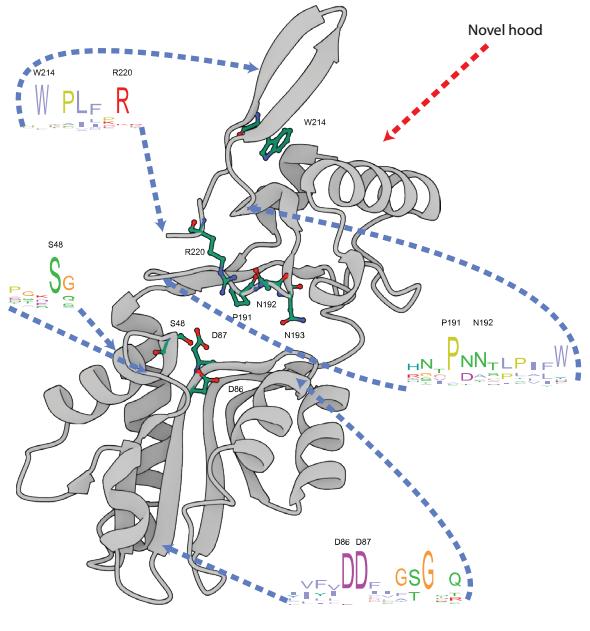


Supplementary Figure S5. Activity site and conserved residues from the Calcineurin-CE domain. (A) Alphafold modeling of the Calcineurin-CE domain (WP_080633877.1). The residues conserved in the activity sites are shown in ball and stick representations. The synapomorphic asparagine is colored in orange. 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-CDP-alcohol diphosphatase domain from the PDB 2NXF without and with the Zn⁺² ion respectively.



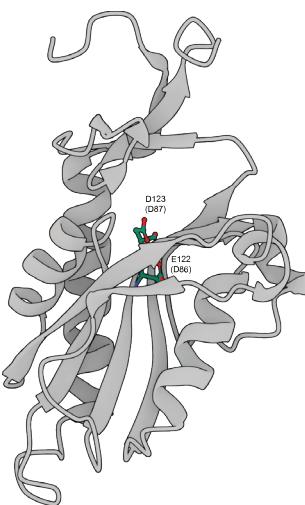
Supplementary Figure S6. Activity site and conserved residues from the HD-CE domain. (A) AlphaFold modeling of the HD-CE domain (CZQ95560.1). The residues conserved in the activity sites are shown in ball and stick representations. The synapomorphic residues are colored yellow. Sequence logos of the 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



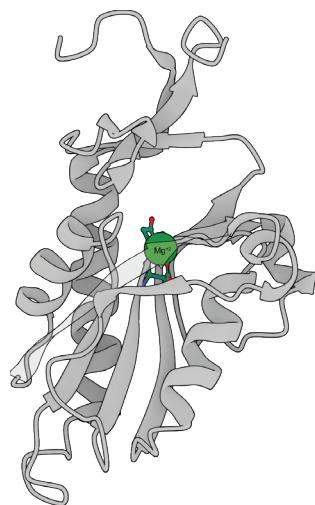
PRTase-CE

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

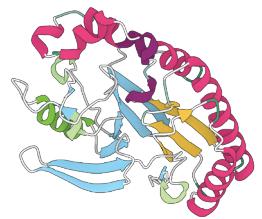
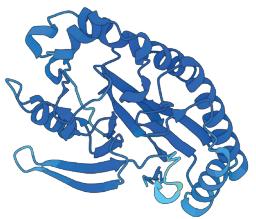


Figure 4G iSirtuin

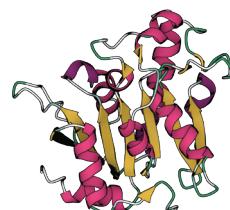


Figure 3B HD-CE

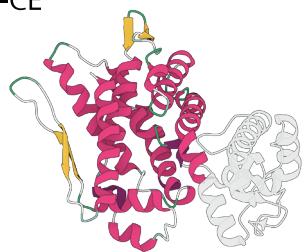
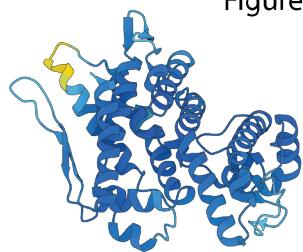


Figure 5G PRTase

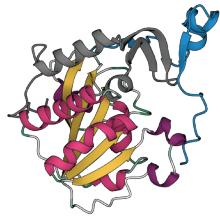


Figure 3C PRTase-CE

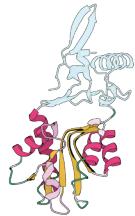


Figure 6B HSP90 system with C-terminal wHTH

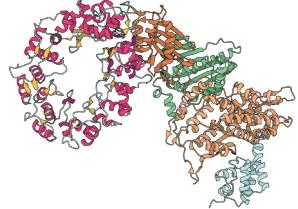
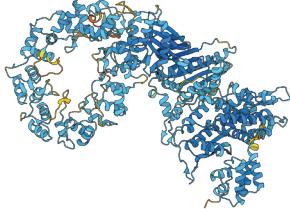


Figure 4B Cdc6-double-CARF•Calcineurin-CE complex

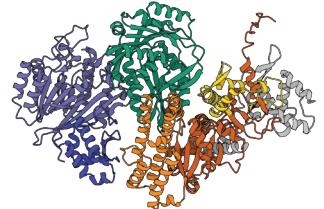
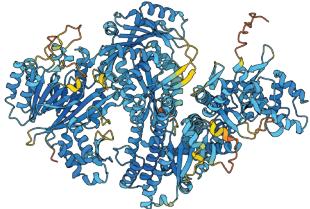


Figure 6E Pepco•Caspase complex



Figure 4D Pua-like

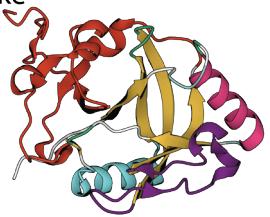
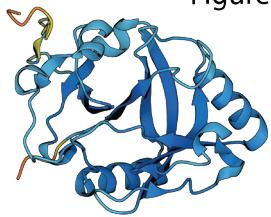
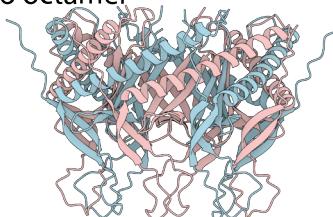
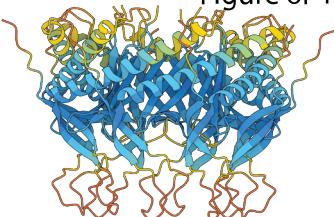


Figure 6F Pepco octamer



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 from position 1 to approximately 980.


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

... (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_169631302.1	ORC-CDC6-like	NZ_JABCLX010000258.1:1-326	GCF_012956145.1	108	*ORC-CDC6-like->	Vibrio parahaemolyticus	Gammaproteobacteria	NZ_JABC010000258
WP_162796445.1	ORC-CDC6-like	NZ_QEDE01000009.1:148204-153123	GCF_003347865.1	324	*ORC-CDC6-like->CARF-double->	Vibrio cholerae	Gammaproteobacteria	NZ_QEDE01000009.1
WP_165387721.1	ORC-CDC6-like	NZ_RBZD01000024.1:34261-41182	GCF_004214575.1	206	*ORC-CDC6-like->CARF-double->	Vibrio vulnificus	Gammaproteobacteria	NZ_RBZD01000024.1
WP_123300074.1	ORC-CDC6-like	NZ_RJVD01000001.1:67514-81212	GCF_003751465.1	655	PSE->*ORC-CDC6-like->	Vibrio crassostreiae	Gammaproteobacteria	NZ_RJVD01000001.1
RZP54190.1	ORC-CDC6-like	RZD01000024.1:34261-41081	GCA_004214575.1	173	*ORC-CDC6-like->CARF-double->	Vibrio vulnificus	Gammaproteobacteria	RZD01000024.1
TQP39892.1	ORC-CDC6-like	VIPA01000296.1:53-498	GCA_006802465.1	147	*ORC-CDC6-like->	Vibrio cholerae	Gammaproteobacteria	VIPA01000296.1
ALA88577.1	ORC-CDC6-like	CP011305.1:1884537-1900596	GCA_001274655.1	604	Na+TrkA_N->THF_DHG_CYH+THF_DHG_CYH_C->IMPDH->?->GATase+tRNA_Me_trans+GMP_synt_C->*ORC-CDC6-like->	Stenotrophomonas maltophilia	Gammaproteobacteria	CP011305.1
CAC9623431.1	ORC-CDC6-like	CAIZN01000037.1:54018-94433	GCA_903814045.1	559	PSE->?->SbcC->PSE->PSE->*ORC-CDC6-like->PSE->	uncultured Gammaproteobacteria bacterium	Gammaproteobacteria	CAIZN01000037.1
WP_078485038.1	ORC-CDC6-like	NZ_MPRL01000095.1:5394-12972	GCF_002020875.1	85	S-methyl_trans+Pterin_bind+B12-binding_2+B12-binding+Met_synt_B12->Glycos_transf_2->wHTH->*ORC-CDC6-like->	Solemya pervernica gill symbiont	Gammaproteobacteria	NZ_MPRL01000095.1
MBV8225564.1	ORC-CDC6-like	JAFAIU010000527.1:221-3464	GCA_019236535.1	320	*ORC-CDC6-like->CARF-double->	Verrucomicrobia bacterium	Verrucomicrobia	JAFAIU010000527.1
MBL9159074.1	ORC-CDC6-like	JAEUHI01000058.1:89930-100252	GCA_016795385.1	611	*ORC-CDC6-like->CARF-double->	Verrucomicrobiales bacterium	Verrucomicrobiae	JAEUHI01000058.1
MBL7075605.1	ORC-CDC6-like	JADHWK01000064.1:2665-12761	GCA_016783505.1	663	Phage_integrase->?->REC->*ORC-CDC6-like->?->CARF-double->tRNA->ncRNA->DNA_pol3_delta2+DNA_pol3_gamma3->YbaB_DNA_bd->RecR+Toprim_4->	candidate division KSB1 bacterium	candidate division KSB1	JADHWK01000064.1











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

