Structural characterization of a soil viral auxiliary metabolic gene product – a functional chitosanase

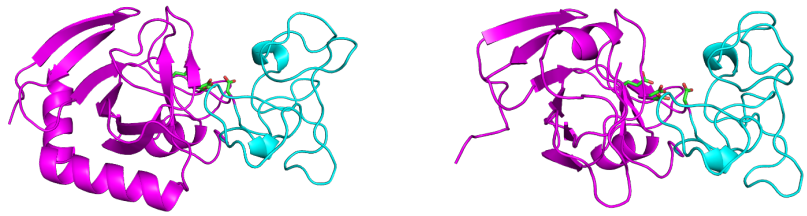
Ruonan Wu[1*], Clyde A. Smith[2*], Garry W. Buchko[1,3], Ian K. Blaby[4], David Paez-Espino[5], Nikos C. Kyrpides[4], Yasuo Yoshikuni[4], Jason E. McDermott[1,6], Kirsten S. Hofmockel[1], John R. Cort[1,7], Janet K. Jansson[1**]

[1]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA 99354, USA; [2]Stanford Synchrotron Radiation Lightsource, Stanford University, Menlo Park, CA 94025, USA; [3]School of Molecular Biosciences, Washington State University, Pullman, WA 99164, USA; [4]US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA; [5]Mammoth Biosciences, Brisbane, CA 94005, USA; [6]Department of Molecular Biology and Immunology, Oregon Health & Science University, Portland, OR, 97239, USA; [7]Institute of Biological Chemistry, Washington State University, Pullman, Washington, USA

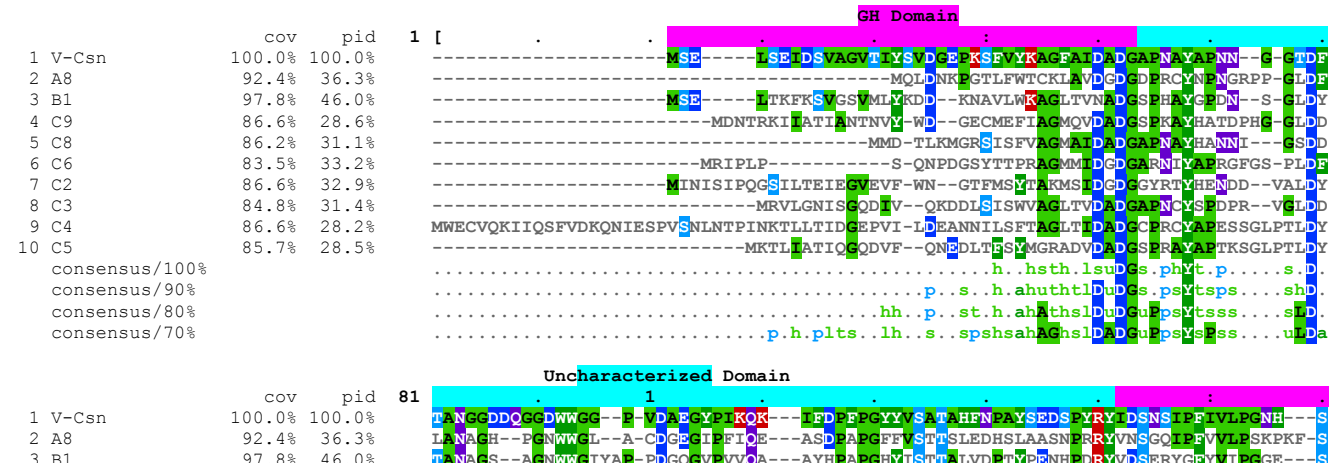
*Co-first authors

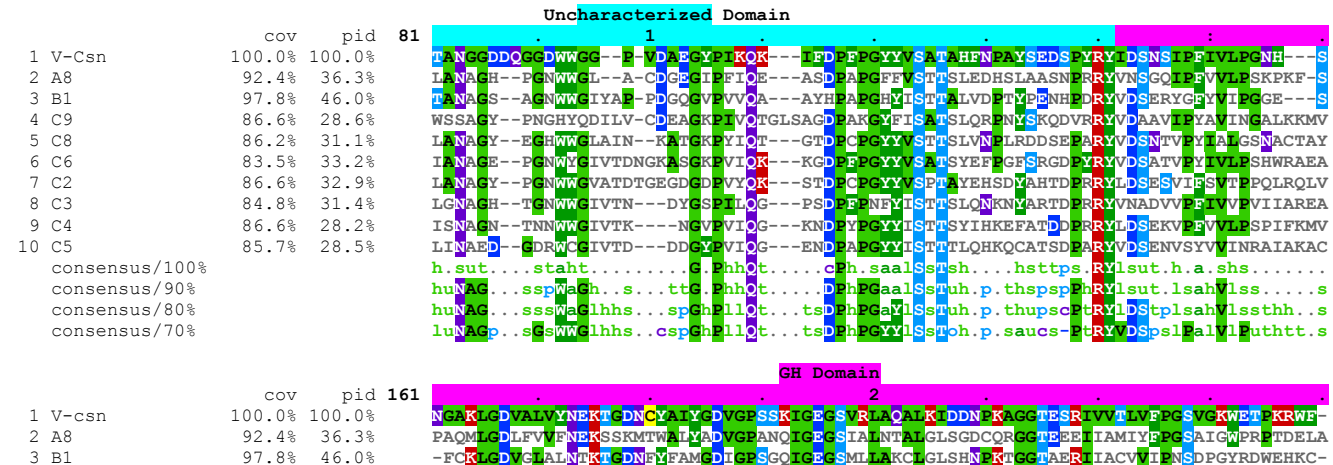
**Senior, corresponding author, janet.jansson@pnnl.gov

a

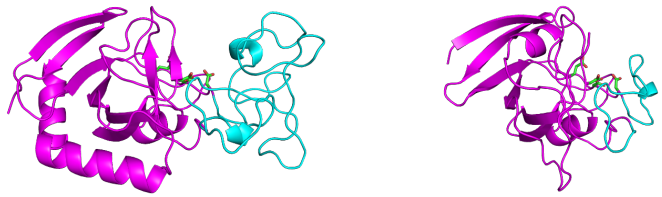


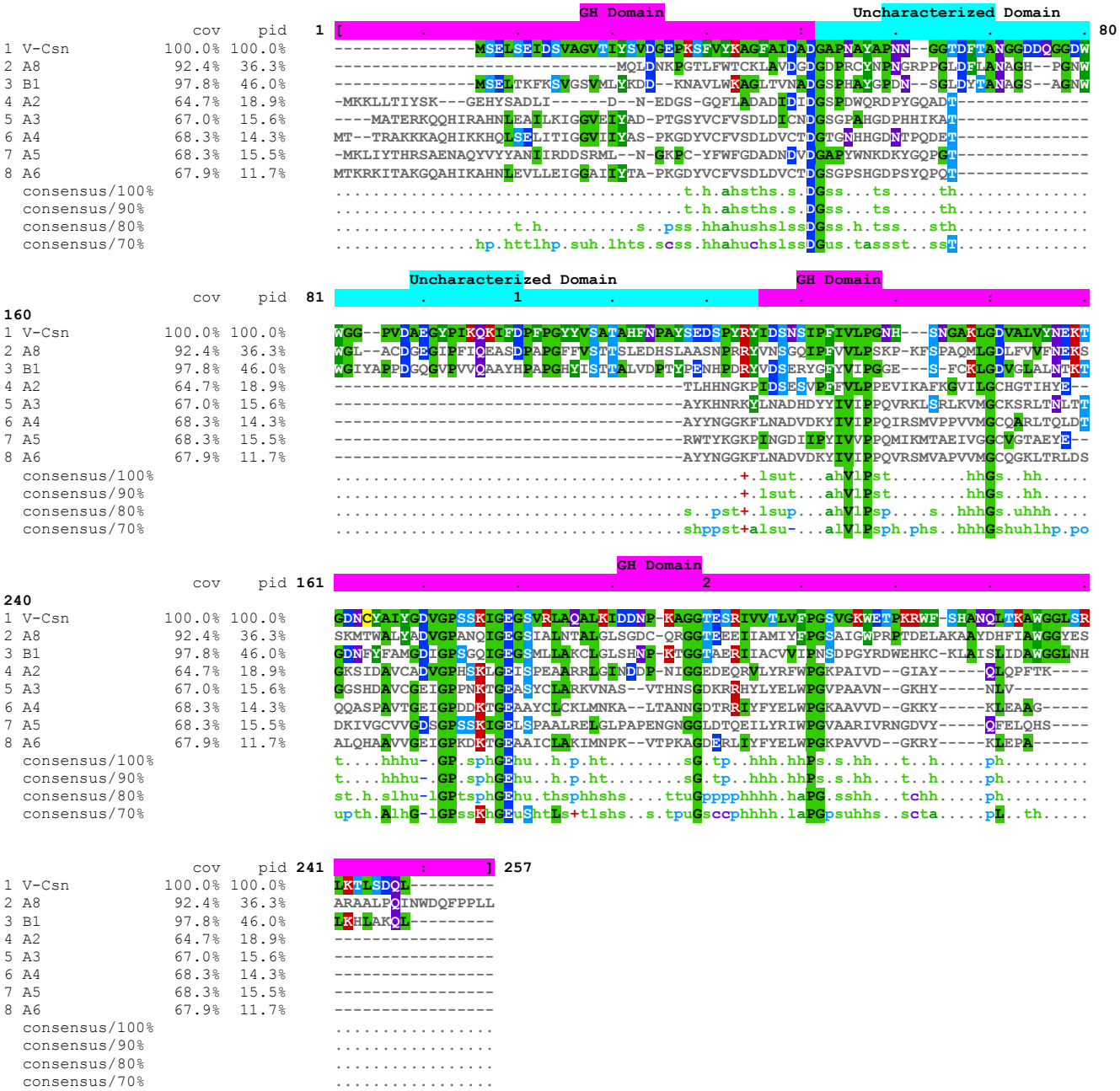
V-Csn X-ray C2 AlphaFold2





b





Supplementary Figure 1 Viral chitosanase protein structure comparisons and partial sequence alignments. **a** Viral AMG chitosanase sequences predicted to have the uncharacterized domain (cyan) and GH domain (magenta) that are similar to those of the V-Csn X-Ray structure. Alignment made with MView 1.63. Conserved active site residues shown as sticks. **b** Viral AMG chitosanase sequences with and without the uncharacterized domain

structure seen in V-Csn (cyan). GH domain in magenta. Sequences lacking the novel domain may also lack the two prominent helices in the GH domain at the C-terminus of the sequence, or a shorter C-terminal helix may be poorly predicted in the structure shown. Conserved active site residues shown as sticks.

Supplementary Table 1 Metadata of viral contigs detected with Glycoside Hydrolase Family 75 chitosanase

The metadata includes 1) information of habitat ('Ecosystem' and 'Habitat') and IMG/VR (v3) dataset ('Sample ID' and 'JGI Proposal Award DOI' if any) where the viral contigs were detected; 2) quality of the viral contigs ('Quality' and 'Completeness'); 3) bioinformatically predicted hosts ('Host assignment'); 4) statistics of chitosanase annotations ('E-value' and 'Bit score'); 5) gene arrangement of each viral contigs ('Upstream' and 'Downstream'); 6) assigned AMG category ('AMG category') and 7) viral chitosanase IDs used in this study ('Short name').

Viral Contig	Length	Sample ID	UVIG	Habitat	Ecosystem	Quality (checkV)	Completeness	Host assignment	Putative viral chitosanase ID	E-value	Bit score	Annotation	Upstream (H:hallmark, O:host;X:NA; V:viral protein)	Downstream (H:hallmark, O:host;X:NA; V:viral protein)	AMG category	Short Name	JGI Proposal Award DOI
3300000364____INPhiseqgaiFeDRAFT_101854920	22176	3300000364	IMGVR_UVIG_3300000364_000020	Terrestrial	Grassland	Low-quality	13.69	Actinobacteria	3300000364____INPhiseqgaiFeDRAFT_101854920_13	2.39E-19	94.7	Chitosanase	VVXVVVXXVXVH	XVXXXXXXXXXXXXV	cat1	A1	
3300012349____Ga0137387_10000020	56247	3300012349	IMGVR_UVIG_3300012349_000002	Terrestrial	Unclassified	Medium-quality	62.57	Proteobacteria	3300012349____Ga0137387_10000020_34	1.40E-07	25.1	Chitosanase	XXXXXXXXXXXXXVXVXVVVVVXXXXXXXXXXXX	XOXXXVVXOXVXHVVXXXXXXVVVVXVXVVXXXXXHVOXXXXX	cat1	A2	10.46936/10.25585/60000754
3300005937____Ga0081455_10000703	43484	3300005937	IMGVR_UVIG_3300005937_000048	Terrestrial	Rhizosphere	Medium-quality	62.52	Unclassified	3300005937____Ga0081455_10000703_29	3.50E-07	23.8	Chitosanase	XXOXXXVXXXXVXXVXXXXXXXXXXXVXV	XXXXXXVXXXXXHXXVXOVXV	cat1	A3	
3300027874____Ga0209465_10001056	12459	3300027874	IMGVR_UVIG_3300027874_000074	Terrestrial	Forest	Low-quality	20.34	Proteobacteria	3300027874____Ga0209465_10001056_2	2.40E-08	27.6	Chitosanase	H	XXXVVVOVHOXHXXXHX	cat0	A4	10.46936/10.25585/60000627
3300012971____Ga0126369_10000332	30760	3300012971	IMGVR_UVIG_3300012971_000062	Terrestrial	Forest	Low-quality	49.72	Proteobacteria	3300012971____Ga0126369_10000332_32	2.00E-09	31.1	Chitosanase	XXXXXXVXOXXXVXVXXXXXVXXXXXXVXXXV	XXXXVVVXXXXXXXXHOVX	cat1	A6	10.46936/10.25585/60000627
3300005937____Ga0081455_10000687	43973	3300005937	IMGVR_UVIG_3300005937_000047	Terrestrial	Rhizosphere	Medium-quality	55.26	Proteobacteria	3300005937____Ga0081455_10000687_48	5.00E-09	29.8	Chitosanase	XXOXXXXXXOXXXXVXXXXXXXXOHXXXVOXXXXXXXVXXXXXXXVXXX	XHXXXXVVVXXXXXXXXH	cat0	A7	
3300010043____Ga0126380_10000006	46305	3300010043	IMGVR_UVIG_3300010043_000005	Terrestrial	Forest	Low-quality	46.25	Unclassified	3300010043____Ga0126380_10000006_37	2.97E-52	180	Chitosanase	VVXVVVXVXXOVXVXXXVXVVVXXVVXXXXVVXXVXVXX	XXXVXXXVVVXVXXXOXXOXXX	cat2	A8	10.46936/10.25585/60000627
3300012582____Ga0137358_10000037	40490	3300012582	IMGVR_UVIG_3300012582_000003	Terrestrial	Unclassified	Medium-quality	55.32	Proteobacteria	3300012582____Ga0137358_10000037_18	2.90E-07	24.1	Chitosanase	XOXXXVVXXXXXVXXHO	XXXXXVXXXXVXXXXXXVXXXXXXXVVVXVXXXXH	cat0	B1	10.46936/10.25585/60000754
3300012926____Ga0137401_10000034	40321	3300012926	IMGVR_UVIG_3300012926_000010	Terrestrial	Unclassified	Medium-quality	54.44	Proteobacteria	3300012926____Ga0137401_10000034_25	2.90E-07	24.1	Chitosanase	VXVXXXXHOXXXXVXXXXXXVXXHO	XXXXXVXXXXVXXXXXXVXXXXXXXVVV	cat1	B1	10.46936/10.25585/60000754
3300032893____Ga0335069_10000845	49235	3300032893	IMGVR_UVIG_3300032893_000075	Terrestrial	Wetland	Medium-quality	55.42	Proteobacteria	3300032893____Ga0335069_10000845_57	6.20E-15	49	Chitosanase	XXVXXXVVXHHVXXVXXXXXXXXXXXX