## Supplementary Figure 7. Taxonomic affiliations of "*Methylacidiphilum infernorum*" V4 proteins.

Taxonomic affiliations of the best BLAST hits for each of the *M. infernorum* proteins to the RefSeq database before (top panel, identical to Figure 3) and after (bottom panel) the inclusion in RefSeq of the proteins encoded in four members of *Verrucomicrobia: Akkermansia muciniphila,* bacterium Ellin514, *Opitutus terrae*, and *Verrucomicrobium spinosum*.



## Supplementary Figure 8 (created by Prof. R. S. Gupta). Multiple alignment of the second protein family specific for the Planctomyces/Verrucomicrobia/ Chlamydia group.

Chlamydia trachomatis D/UW-3/CX<sup>#</sup> NP\_219933 Chlamydia muridarum Nigg NP\_297078 Chlamydophila felis Fe/C-56 YP\_515687 Chlamydophila caviae GPIC NP\_829109 Chlamydophila abortus S26/3 YP\_219659 Chlamydia pneumoniae CWL029<sup>#</sup> NP\_224704 Protochlamydia amoebophila UWE25 YP\_008052 Bacterium Ellin 514 ZP\_02965781 Lentisphaera araneosa HTCC 2155 ZP\_01877577 Akkermansia muciniphila BAA-835 YP\_001876867 Verrucomicrobium spinosum DSM 4136 ZP\_02926670 Opitutus terrae PB90-1 YP\_001819597 Planctomyces maris DSM 8797 ZP\_01856939 C. Kuenenia stuttgartiensis CAJ71823 M. infernorum V4 Minf\_0061 YP\_001938720.2<sup>##</sup>

MSRHRSYGKSV-KGETKRNVLKRFERIEVLRKLGRWDDATAKKATGLLKTPAIK MSRHRSYGKSI-KGETKRNVLKRFERIEVLRKLGRWDDATAKKATGLLKTPVIK MSRHRSYGKSI-KGETKRNVLKRFERIEVLRKLGRWNDATAKKATGLPKTPVMK MSRHRSYGKSI-KGETKRNVLKRFERIDVLRKLGRWNDVTAKKATGLPKTPVMK MSRHRSYGKSI-KGETKRNVLKRFERIDLLRKLSRWNDTTAKKATGLPKTPVIK MSRHRSYGKSV-KGVTKRNVLKRFERVEVLRKLGRWNDSTAKKVTGLPKTPILK MSRHPSFGKAG-KTATKRNVLKRFERIDVLKKIGRWKDSENKRITGLPKTPVL-MSQHRSLRAVA-TMGGKRNVLKRFERVGLLKKRGQWKE--GDRITGLRKTKPEA MSIHRSLKVKG-NTAGKKNVLKRFERVDQLIQEGRLKP--GDQVLGLPKTKVNI MSKHSSLKATG-TVGGKRSVLKRFERVKLLKERGEWKK--GOSPLGLPKTKHEA MSQHRSLKSSGGSVGTKRSVLKRGERIKLMKARGQWNE--GRSLYNLPKTKPEA MSQHKSLQGSS-GIVIKRNVLKRFERVDILKKRGQWKA--GDRVQGLRKTKPDV VSLDKSLKSKS-TLVRARNVLKRAERIEKLKFEDRWVE--GQGALGLPKVRVEK MSIDKSLKPKG-KLSRPRNVYRKVERIAILKAENRWDQ--DTSVFGLPKVKVEK MSQHRSYRTGS-LLVAKRNVLKRYERINILKKQGKWKE--GDKVLGLPKTKPI-

Multiple sequence alignment of the protein CT421.2 (NP\_219933), for which all BLAST hits are observed from the PVC group (*Planctomycetes, Verrucomicrobia, Lentisphaerae* and *Chlamydiae*) species. The full-length sequences are shown for most species (length 51-53 aa), except for the two *Planctomycetes*, where sequences are 81 aa long and contain extra amino acids at the N – or C-terminal ends. The positions that are identical in all sequences are marked with \*. Other conserved positions are marked with the symbols : and .

<sup>#</sup> In *C. trachomatis* and *C. pneumoniae*, the protein is present in all of the sequenced strains and the sequences are identical to those shown here.

<sup>##</sup>The updated sequence of Minf\_0061 (YP\_001938720.2) was added by the authors.

## Supplementary Figure 9 (created by Prof. R. S. Gupta). Partial sequence alignment of the RpoB protein.

Proteobacteria    E. coli    CAA23629    IIPYRGSWLDFEFDPKDNLFVRID    RRRKLPATIILf      V. cholerae    EAY40360	ALNYTTEQ GK-SAE (SIGL-P (G-NN IGMDG-E (GMDG-E (G-SA-A (G-NS	ILDLF I- NA- I- AH- I- ST- TT- L-NY- L-EY-
V. cholerae    EAY40360   Y    SS      Pse. aeruginosa    NP_252960   CV    SV      Pas. multocida    NP_246676   YA	GK-SAE G-SE (SIGL-P (G-NN IGMDG-E (GMDG-E (G-SA-A (G-NS	I- NA- I- AH- I- ST- TT- L-NY- L-EY-
Proteobacteria    Pse. aeruginosa    NP_252960   CV   SV      Pas. multocida    NP_246676   YA   SV      Ral. solanacearum    NP_521155   I-YF   N-V-L-F      Nei. meningitidis    YP_974254   VI-YF   V-SL-F      A. tumefaciens    NP_354931    VI-YF   V-SL-F      Ca. crescentus    NP_419321    VI-VYA   V-SL-F      Geo. sulfurreducens    NP_953905    VH-I-I   V-L-F      Bde. bacteriovorus    NP_969761    VAOLIH   F-V-L-F      Hel. pylori    NP_207989   Y-Y-SV-YAN    KV-V-LF	(SIGL-P (SIGL-P (G-NN IGMDG-E (GMDG-E (G-SA-A (G-NS	NA- I- AH- I- ST- TT- L-NY- L-EY-
Pas. multocida    NP_246676   YA      Ral. solanacearum    NP_521155   YF   N-V-L-F      Nei. meningitidis    YP_974254	(SIGL - P ( G - NN I GMDG - E ( GMDG - E ( G - SA - A ( G - NS	I- AH- I- ST- TT- L-NY- L-NY- L-EY-
Proteobacteria    Ral. solanacearum    NP_521155   I-YF-V-   N-V-V-L-F      Nei. meningitidis    YP_974254   L-YF   W-V-L-F      A. tumefaciens    NP_354931    VV-SL-F      Ca. crescentus    NP_419321    V	(SIGL-P (G-NN IGMDG-E (GMDG-E (G-SA-A (G-NS	AH- I- ST- TT- L-NY- L-EY-
Proteobacteria    Nei. meningitidis    YP_974254   L-YF   W-V-L-F      A. tumefaciens    NP_354931    VIAIVYA   V-SL-M      Ca. crescentus    NP_419321    VAIVYA   V-SL-M      Geo. sulfurreducens    NP_953905    VAIVY   MVL-F      Bde. bacteriovorus    NP_969761    VAQLIH   F-V-L-F      Hel. pylori    NP_207989   YY-S-V-YAN    KV-V-LF	(G-NN IGMDG-E (GMDG-E (G-SA-A (G-NS	I- ST- TT- L-NY- L-EY-
A. tumefaciens  NP_354931  VIAIVYA V-SL-I    Ca. crescentus  NP_419321  VAIVY TF-I    Geo. sulfurreducens  NP_953905  VHIMVL-I    Bde. bacteriovorus  NP_969761  VAQLIH F-VL-I    Hel. pylori  NP_207989 DYY-SV-YAN  KV-V-L-F	/GMDG-E /GMDG-E (G-SA-A (G-NS	ST- TT- L-NY- L-EY-
Ca. crescentus    NP_419321    VAIVY   TF-'      Geo. sulfurreducens    NP_953905    VHI   MVL-'      Bde. bacteriovorus    NP_969761    VAQLIH   F-VL-'      Hel. pylori    NP_207989   DYY-SV-YAN    KV-V-L-F	(GMDG-E (G-SA-A (G-NS	TT- L-NY- L-EY-
Geo.    sulfurreducens    NP_953905    VHI   NVL-F      Bde.    bacteriovorus    NP_969761    VAQLIH   F-VL-F      Hel.    pylori    NP_207989   DYY-SV-YAN    KV-V-L-F	(G-SA-A (G-NS	L-NY- L-EY-
Bde.    bacteriovorus    NP    969761    VAQLIH   F-VL-H      Hel.    pylori    NP    207989   DYY-SV-YAN    KV-V-LF	(G-NS	L-EY-
Hel. pylori NP_207989DYY-SV-YAN KV-VLF-		
	·-MD-QKQD	-IKM-
Camp. jejuni YP 178546 DYY-AV-YN KV-V-MLF	G-KKQD	-IK
Amificales [Hydrogenobaculum Y04AAS1 ZP 02061359 -H-DKRIELSTNSDIISARI DKKRIS-IS	-FGLE-AY	DSQ-
Agu. aeolicus NP 214331 DKRVELSGATDLISARI D-KSAV-	-FGLE-AY	DKY-
(Chl. muridarum NP 296965EASIN-LIYIH BKKIL-MTFI	G-S-DA	D-IEE-
Chl. trachomatis NP 219820EAVIN-LIYIH RKKIL-MTFI	G-S-DA	D-IEE-
Chlam, pneumoniae NP 445236EAIIN-LIYIH RKKIL-ITFI	G-SSDA	D-IEE-
Chlam, abortus YP 220059EAIIN-LIYIH BKKIL-ITFI	G - SSDA	D-IEE-
Chiamyatae Chiam, felis YP 515237EAIIN-LIYIH BKKIL-MTFI	G - SSDA	D-IEE-
Chlam, caviae NP 829555EAIIN-LIYIH BKKIL-MTFI	G-SSDA	D-IEE-
Chlam, psittaci AAV92911EAIIN-LIYIH RKKIL-ITFI	G-SSDA	D-IEE-
Proto amoebophila YP 007603EGATN-LIHTY BKKILTET	G-SSNS	D-TEE-
Sim. negevensis TIGRESSIN-LIY-YV- RKKVLSFI	T-G-SSDS	D-IEE-
C Bacterium Ellin514 ZP 02970829DYEAQTS-L-Y-YL- RKKFLT-TFF	G-GSDA	рк
Akkerman. muciniphila YP 001877648 EVQTN-L-Y-YL- RRRFLTFM-	Y-GFK-DR	D-VSQ-
Vermining Line / Opitutaceae bacterium ZP 02012834DTEVQIN-L-Y-YL- RRRFLI-TLF-	G-SDNT	КЕ
Victivallis vadensis ZP 01924354 DM-VQIN-FIYIYL- RRR FYI-TF	- IG - P - NR	DAE
Lentisphaera araneosa NZ ABCK01000017DMEVQY-NH-LIYIFM- RRRFLISTF	- VYSESNY	DL-
Opitutus terrae PB90-1 YP_001817123DTEVQNN-L-Y-YL- RRRFLI-TL	-IG-SSDI	DL-N
└Verrucomicro. spinosum ZP_02930291DTEVQTN-L-Y-YL- RRRFLTL	VIG-PGDE	DK
r Rho. baltica NP_866682 VEIEVNVTKA-T QSG-FAML	- MDPKYST	DA
Planctomycetes { K. stuttgartiensis CAJ73710EIEL-VGKV-T QSGCF	SAEYANI	DE-II
LBIa. marina EAQ82611EIEINITKSFT QSG-FATL	- MSPELSD	DAS-I-
Bacteroidetes- Por. gingivalis NP_904700FKIE-AT-INNVMYAYKKV-TL	- IGFEADK	DI-
Chlorobi Group   Bact. fragilis YP_213594FKIE-AT-INNVMYAYKKV-TL	-IGFENDK	DEI-
Cb. tepidum AAM71403 -V-TIE-QT-INNQIY QKKNFLV-AL	- IGFAKDE	DG
Spirochetes ↓ Tre. pallidum NP_218681EI-Q-KDLIYAKL KK-RILG-VF	H-E-RE(	Q-IEA-
└ Bor. burgdorferi NP_212523EI-S-KDYLYVKI -KKRILI-LF-	GFD-RE	K-IET-
Cyanobacteria ∫ Gloe. violaceus NP_925229 LNAKT-AN-LVW KTS-VVL-F	(GLSDNE	A -
し Nostoc sp. PCC 7120 NP_485634 LNAKT-RN-LVW KTS-QVL-ト	(GLSDNE	-F-A-
Deinococcus-Thermus { The. thermophilus YP_005430 PL-KP-I-L-VE-NGVVSMKVN KF-LVLL	V-G-DQ-TI	LRE
D. radiodurans NP_294636 PM-KP-IELAGDVLEMKVNKF-VSLL	V-GMDDAS	- RA
T. maritima NP_228268 FL-SAEVILY-GVLYAGL DGK-VNLFLF-	(TIG-EKDE	DS
∫ Myc. tuberculosis ABK96916 VSAE-DV-KR-TVGK-RQ-V-VL-⊦	(GW-S	- VER -
Actinobacteria 🚽 Cor. glutamicum NP_599733 VSAE-DV-KR-SVGK-RQ-V-VL-P	(GW	- TER -
Str. coelicolor NP_628815SAEM-I-KR-MVGKQSV-VL-	(GW	EE -
└Bif.longum AAN25011SAEI-KQPQVKQS-IVF-M	I-IGM-KSE	- AQA -
Bac. subtilis NP_387988 VNAEY-T-AVVYTV-VL	GFGSD-F	E
Firmicutes J Sta. aureus CAA45512NAEY-T-AVVYTL-VL	GFSSD-F	E-VL
Strep. pneumoniae NP_346388 VNAEL-S-SITYTTI-F-TLV	GFSGDDF	E-F-I-
└Clo.perfringens NP_563329 VNAEY-T-SNSVIHTI-LV	MGFGSDAE	_\/NV

Partial sequence alignment of the RpoB protein showing a 3 aa insert (boxed) that is uniquely found in all available *Chlamydiae* and *Verrucomicrobia-Lentisphaerae* groups of species, but not found in any other bacteria, including the *Planctomycetes* species. This insert was originally described in the Ref. 18 and its presence in various species from the *Chlamydiae* and *Verrucomicrobia-Lentisphaerae* phyla strongly indicates that species from these two phyla are most closely related to each other. The dashes in this alignment denote identity with the amino acid on the top line.