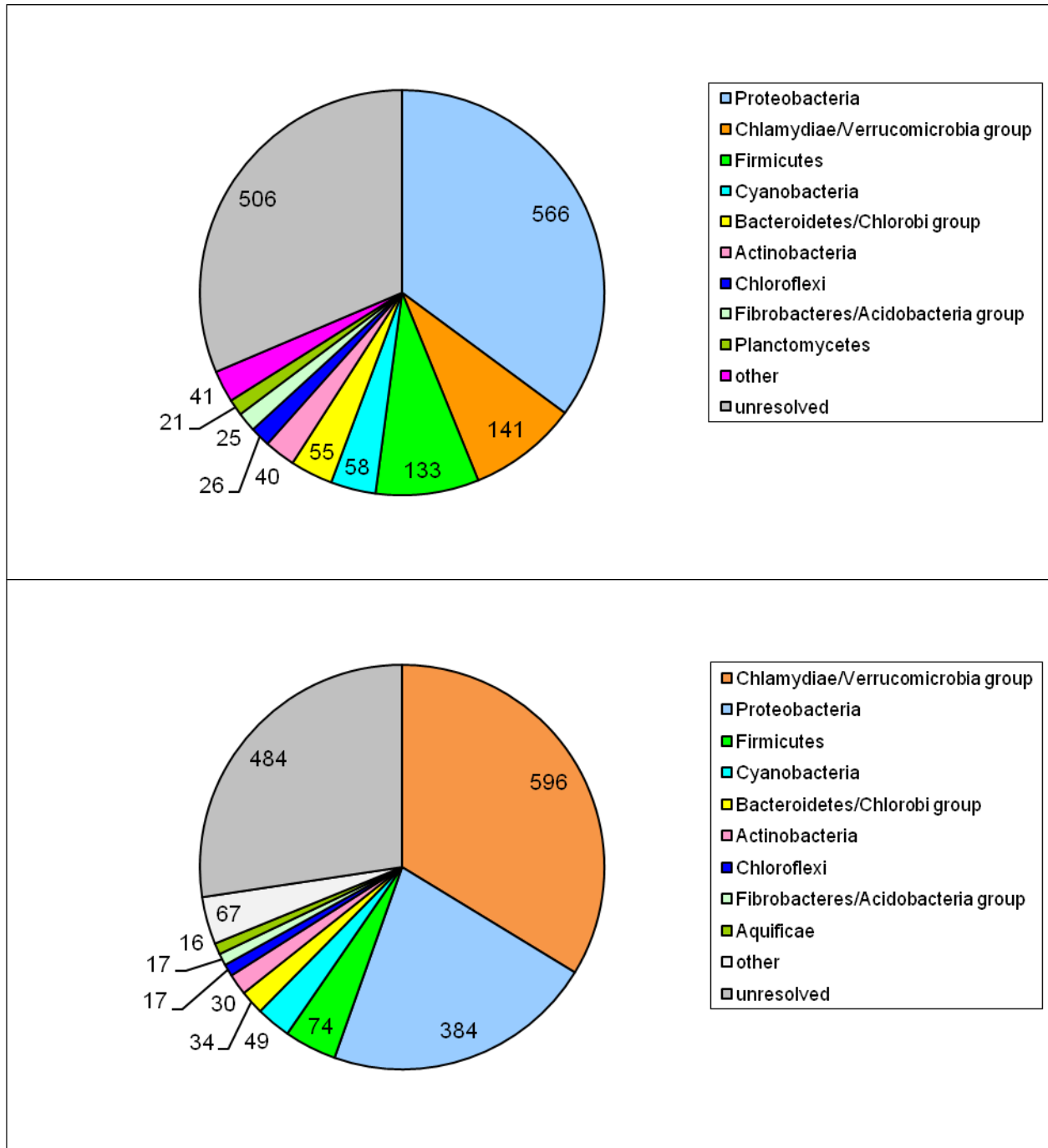


**Supplementary Figure 7. Taxonomic affiliations of “*Methyacidiphilum inferorum*” V4 proteins.**

Taxonomic affiliations of the best BLAST hits for each of the *M. inferorum* 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, *Opiritatus 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	MSRHRSYGKSV-KGETKRNVLKRFERIEVLRKLGRWDDATAKKATGLLKTPAIK
Chlamydia muridarum Nigg NP_297078	MSRHRSYGKSI-KGETKRNVLKRFERIEVLRKLGRWDDATAKKATGLLKTPVIK
Chlamydophila felis Fe/C-56 YP_515687	MSRHRSYGKSI-KGETKRNVLKRFERIEVLRKLGRWDDATAKKATGLPKTPVMK
Chlamydophila caviae GPIC NP_829109	MSRHRSYGKSI-KGETKRNVLKRFERIDVLRKLGRWDDVTAKKATGLPKTPVMK
Chlamydophila abortus S26/3 YP_219659	MSRHRSYGKSI-KGETKRNVLKRFERIDLLRKLGRWDDTTAKKATGLPKTPVIK
Chlamydia pneumoniae CWL029 <sup>#</sup> NP_224704	MSRHRSYGKSV-KGVTKRNVLKRFERVEVLRKLGRWDDSTAKKVTGLPKTPILK
Protochlamydia amoebophila UWE25 YP_008052	MSRHPSFGKAG-KTATKRNVLKRFERIDVLRKLGRWDDSENKRITGLPKTPVL-
Bacterium Ellin 514 ZP_02965781	MSQHRSRAVA-TMGGKRNVLKRFERVGLLKKRGQWKE--GDRITGLRKTPEA
Lentisphaera araneosa HTCC 2155 ZP_01877577	MSIHRSCLKVKG-NTAGKKNVLKRFERVDQLIQEGRKLP--GDQVLGLPKTKVNI
Akkermansia muciniphila BAA-835 YP_001876867	MSKHSSCLKATG-TVGGKRSVLKRFERVLLKRGQWKE--GQSPLGLPKTKHEA
Verrucomicrobium spinosum DSM 4136 ZP_02926670	MSQHRSCLKSSGGSVGTKRSVLKRGERIKLMKARGQWNE--GRSLYNLPKTKPEA
Opitutus terrae PB90-1 YP_001819597	MSQHKSLLQSS-GIVIKRNVLKRFERVDILKRGQWKA--GDRVQGLRKTTPDV
Planctomyces maris DSM 8797 ZP_01856939	VSLDKSLKSKS-TLVRARNVLKRAERIEKLFEDRWVE--GQALGLPKVRVEK
C. Kuenenia stuttgartiensis CAJ71823	MSIDKSLKPKG-KLSRPRNVYRKVERIAILKAENRWDQ--DTSVFGLPKVKVEK
M. infernorum V4 Minf_0061 YP_001938720.2 <sup>##</sup>	MSQHRSYRTGS-LLVAKRNVLKRYERINILKQGWKE--GDKVLGLPKTKPI-
	:*...*..... :* : : ** : ..... :* * .

Multiple sequence alignment of the protein CT421.2 (NP\_219933), for which all BLAST hits are observed from the PVC group (*Planctomyces*, *Verrucomicrobia*, *Lentisphaerae* and *Chlamydiae*) species. The full-length sequences are shown for most species (length 51-53 aa), except for the two *Planctomyces*, 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.**

			176	224
			IIPYRGSWLDFFEDPKDNLVFRID	RRRKLPTAII LRALNYTTEQ ILDLF
			-----Y----	-----S-----GK-SAE ---I-
			-----CV----	-----SV-----G-S--E --NA-
			-----YA----	-----YA-----I-----
			-----I-YF-V-	-----M-V--L-KSIGL-P-- --AH-
			-----L-YF----	-----M-V--L-K--G-NN-- --I-
			V-----I---A--IVYA---	-----V-SL-M--GMDG-E --ST-
			V-----A--IVY----	-----TF-Y--GMDG-E --TT-
			V-----H--I-----	-----M--VL-K--G-SA-A L-NY-
			V-----A---Q--LIH----	-----F-V--L-K--G-NS-- L-EY-
			---D---Y--Y-S--V-YA--N	K---V-V--LF--MD-QKQD -IKM-
			---D---Y--Y-A--V-Y--N	K---V-V-MLF--G-KKQD -IK--
			-H-DK--RIE--LSTNSDIISARI	DKKRIS-IS---FGLE-AYD--SQ-
			---DK--RVE--LSGATDLISARI	D-K--S--AV--FGLE-AYD--KY-
			-----EAS--IN-LIYIH--	RKK ---IL-MTFI--G-S-DAD-IEE-
			-----EAV--IN-LIYIH--	RKK ---IL-MTFI--G-S-DAD-IEE-
			-----EAI--IN-LIYIH--	RKK ---IL-ITFI--G-SSDAD-IEE-
			-----EAI--IN-LIYIH--	RKK ---IL-ITFI--G-SSDAD-IEE-
			-----EAI--IN-LIYIH--	RKK ---IL-MTFI--G-SSDAD-IEE-
			-----EAI--IN-LIYIH--	RKK ---IL-ITFI--G-SSDAD-IEE-
			-----EAI--IN-LIYIH--	RKK ---IL-ITFI--G-SSDAD-IEE-
			-----EGA--TN-LIHIY--	RKK ---IL--TFI--G-SSNSD-IEE-
			-----ESS--IN-LIY-YV-	RKK ---VL--SFI-T-G-SSSD-IEE-
			---D---YEAQ--TS-L-Y-YL	RKK ---FLT-TFF--G-GSDAD--K--
			---D---EVQ--TN-L-Y-YL	RRR ---FL--TFM-Y-GFK-DRD-VSQ-
			---D--T--EVQ--IN-L-Y-YL	RRR ---FLI-TLF--G-SDNTK--E--
			---D---M-VQ--IN-FIYIYL	RRR ---FYI-TF--IG-P-NRD--AE-
			---D---MEVQY-NH-LIYIFM-	RRR ---FLISTF--VYSESNYDL-
			---D--T--EVQ--NN-L-Y-YL	RRR ---FLI-TL--IG-SSDIDL-N--
			---D--T--EVQ--TN-L-Y-YL	RRR ---FL--TL--VIG-PGED--K--
			V--E---IEVNVTK--A-T----	QSG-FA--ML---MDPKYSTDA----
			---E---IEL-VGK--V-T----	QSG----CF---SAEYANDE-II
			---E---IEINITK--SFT----	QSG-FA--TL---MSPELSDDAS-I-
			---FK---IE-AT-INNVMYAY--	-KK---V-TL---IGFEADKD--K-
			---FK---IE-AT-INNVMYAY--	-KK---V-TL---IGFENDKD--EI-
			-V-T---IE-QT-INNQI--Y-Y-	QKKNFLV-AL---IGFAKDED--G--
			-----E--I-Q-KDLIYAKL	KK-RILG-VF---H-E-REQ-IEA-
			-----E--I-S-KDYLIVKI	-KKRILI-LF---GFD-REK-IET-
			L--N--A--K--T-AN-LW----	KT---S-VVL-K--GLSDNE ---A-
			L--N--A--K--T-RN-LW----	KT---S-QL-K--GLSDNE -F-A-
			PL-K--P-I-L-VE-NGVSMKVN	K--F-LVLL--V-G-DQ-TLRE---
			PM-K--P-I-EL--AGDVLEMKVNK	--F-VSLL--V-GMDDAS-RA---
			FL-S--A--EVIL--Y-GVLYAGL	DGK-VNLFLF-KTIG-EKDED--S--
			V--S--A--E-DV-KR-TVG----	-K-RQ-V-VL-K--GW-S-- -VER-
			V--S--A--E-DV-KR-SVG----	-K-RQ-V-VL-K--GW---- -TER-
			---S--A--EM-I-KR-MVG----	-K--QSV-VL-K--GW---- -EE-
			---S--A--E--I-K-QPQ--V-	-K--QS-IVF-M-IGM-KSE -AQA-
			V--N--A--EY-T-A--VVY----	-T---V-VL---GFGSD-E-----
			---N--A--EY-T-A--VVY----	-T---L-VL---GFSSD-E-V--L
			V--N--A--EL-S-S--ITYT---	-T--I-F-TLV--GFGSDD-E-F-I-
			V--N--A--EY-T-SNSVIH----	-T---I-LV--MGFGSDAE -VNY-

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.