Distinct approaches for the detection and removal of chimeric 16S rRNA sequences can significantly affect the outcome of between-site comparisons

Nika Zajec, Blaž Stres, Gorazd Avguštin*

University of Ljubljana, Biotechnical Faculty, Animal Science Department, Chair for Microbiology and Microbial Biotechnology, Groblje 3, 1230 Domžale, Slovenia

*Corresponding author. Email: gorazd.avgustin@bf.uni-lj.si

Supplement.

Publicly available clone libraries containing 16S rRNA sequences from marine sediment environments were used as primary data for chimera search analysis. Chimera_check (Maidak et al. 2001), CCODE (Gonzalez et al. 2005), Pintail (Ashelford et al. 2005), Chimera Slayer (Schloss et al. 2009) and Bellerophon (Huber et al. 2004) computer programs were used, the latter with 4 different settings. Eight chimera identification procedures were thus tested on selected clone libraries and as a baseline, and to complement the clone libraries from published literature, 8 clone libraries, 4 archaeal and 4 bacterial, were prepared from a submarine spring located in the northern Adriatic Sea (Izola32) and analyzed, too. The identified chimeras were removed from primary datasets, creating subsets named Chimera Identification Approach (CIA) filtered datasets, and the effect of the removal of chimeras through various CIA on between-site relationships was analyzed

Table S1. Chimeric sequences, detected with different methods in 16S rRNA archaeal and bacterial clone libraries generated from Slovene marine sediment samples and selected public clone libraries. Numbers in the brackets designate the total number of identified chimeric sequences. HH: Huber Hugenholtz, J: Jukes Cantor, K: Kimura. A: archaeal clone library; B: bacterial clone library

		Derivatives of clone libraries according to chimera identification approach							
	RDP II Chimera_Check	Bellerophon Greengenes				CCODE	Chimera Slayer	Pintail	All chimeras identified
		HH	JC	K	none				
A-spring-out	22, 40, 43, 53, 55, 64, 65, 67, 71, 78, 83, 86	3, 21, 35, 41	/	/	40, 43, 45, 52, 60, 81	3, 4, 6, 8, 9, 13, 19, 25, 30, 32, 34, 35, 38, 41, 43, 44, 46, 60, 62, 63, 67, 69, 71, 73, 74, 76, 79, 86, 88, 91	36, 52, 69,	13, 23, 25, 32, 67, 71, 73, 78, 83, 86, 90	3, 4, 6, 8, 9, 13, 19, 21, 22, 23, 25, 30, 32, 34, 35, 36, 38, 40, 41, 43, 44, 45, 46, 52, 53, 55, 60, 62, 63, 64, 65, 67, 69, 71, 73, 74, 76, 78, 79, 81, 83, 86, 88, 90, 91 (=45)
A-spring-wall	35, 72, 73	36, 38, 45, 53, 60, 65, 72, 76	/	/	/	13, 14, 21, 28, 29, 31, 34, 36, 45, 67, 70, 71, 73	/	9, 14, 29, 30, 41, 48, 50, 52, 57, 70, 71	9, 13, 14, 21, 28, 29, 30, 31, 34, 35, 36, 38, 41, 45, 48, 50, 52, 53, 57, 60, 65, 67, 70, 71, 72, 73, 76 (=27)
A-spring-up	2, 31, 33, 49, 53	1, 2, 8, 12, 13, 14, 17, 27, 37, 41, 49	/	/	3, 8, 13, 14, 32, 41, 49, 54	9, 11, 13, 14, 25, 28, 38, 43, 51, 53,	/	3, 7, 11, 40, 41	2, 3, 7, 8, 9, 11, 13, 14, 25, 28, 31, 32, 33, 38, 40, 41, 43, 49, 51, 53, 54 (=21)
A-spring-down	2, 50, 55	20, 27, 29	/	/	19, 20, 32, 53, 66, 68, 75	3, 12, 17, 18, 20, 22, 37, 40, 42, 43, 44, 46, 48, 51, 55, 57, 63, 66, 67, 69, 71, 74	/	3, 57, 69	2, 3, 12, 17, 18, 19, 20, 22, 27, 29, 32, 37, 40, 42, 43, 44, 46, 48, 50, 51, 53, 55, 57, 63, 66, 67, 68, 69, 71, 74, 75 (=31)
B-spring-out	9, 15, 28, 34, 45	4, 14, 33, 35	1	/	4, 33, 46	4, 8, 11, 14, 18, 28, 28, 30, 40, 41, 43, 50	9, 14, 41	20	1, 4, 8, 9, 11, 14, 15, 18, 20, 28, 30, 33, 34, 35, 40, 41, 43, 45, 46, 50 (=21)
B-spring-wall	46, 70	7, 43, 52, 70, 71	/	/	2, 52, 68, 70	1, 4, 8, 9, 10, 11, 12, 14, 16, 17, 19, 20, 22, 23, 24, 27, 33, 37, 40, 45, 46, 49, 50, 51, 52, 54, 56, 57, 65, 66, 67, 69, 71	/	4, 20, 45, 54, 67	1, 2, 4, 8, 7, 9, 10, 11, 12, 14, 16, 17, 19, 20, 22, 23, 24, 27, 33, 37, 40, 43, 45, 46, 49, 50, 51, 52, 54, 56, 57, 65, 66, 67, 68, 69, 70, 71 (=38)
B-spring-up	22, 45, 46, 56, 68, 69, 87, 91, 97, 101, 137, 141	41, 31, 57, 60, 66, 75, 77, 81, 93, 112, 131, 132	/	/	12, 17, 29, 44, 65, 67, 75, 77, 85, 93, 104, 106, 107, 108, 109, 110, 118, 131, 132	3, 5, 6, 7, 11, 12, 16, 17, 19, 23, 25, 29, 31, 35, 42, 44, 47, 48, 53, 58, 59, 61, 64, 65, 67, 72, 74, 79, 82, 83, 88, 90, 93, 95, 98, 104, 107, 108, 109, 110, 113, 116, 119, 122, 124, 131, 132, 135, 138, 140	46	16, 74, 91, 113, 132, 140	3, 5, 6, 7, 11, 12, 16, 17, 19, 22, 23, 25, 29, 31, 35, 41, 42, 44, 45, 46, 47, 48, 53, 56, 57, 58, 59, 60, 61, 64, 65, 66, 67, 68, 69, 72, 74, 75, 77, 79, 81, 82, 83, 85, 87, 88, 90, 91, 93, 95, 97, 98, 101, 104, 106, 107, 108, 109, 110, 112, 113, 116, 118, 119, 122, 124, 131, 132, 135, 137, 138, 140, 141 (=73)
B-spring-down	10	2, 10, 14, 18, 29, 31, 32, 35, 39, 42, 49, 51, 52, 58, 85, 92, 102, 105, 114	/	/	32, 51, 89, 114	1, 3, 6, 9, 10, 13, 14, 17, 18, 22, 25, 27, 29, 30, 31, 33, 34, 39, 46, 51, 52, 54, 55, 58, 60, 64, 66, 67, 68, 70, 71, 72, 73, 75, 76, 77, 78, 79, 80, 82, 83, 84, 86, 87, 88, 89, 91, 94, 95, 96, 97, 99, 101, 102, 103, 104, 107, 108, 110, 111, 112, 113, 115, 116, 120, 121, 122, 125, 126, 127, 130, 132,	/	1, 30, 55, 67, 71, 80	1, 2, 3, 6, 9, 10, 13, 14, 17, 18, 22, 25, 27, 29, 30, 31, 32, 33, 34, 35, 39, 42, 46, 49, 51, 52, 54, 55, 56, 58, 60, 64, 66, 67, 68, 70, 71, 72, 73, 75, 76, 77, 78, 79, 80, 82, 83, 84, 85, 86, 87, 88, 89, 91, 92, 94, 95, 96, 97, 99, 101, 102, 103, 104, 105, 107, 108, 110, 111, 112, 113, 114, 115, 116, 120, 121, 122, 125, 126, 127, 130, 132 (=82)
Garcia et al., 2009	17, 38	7, 8, 9, 19, 26	6, 7, 9, 21, 26	6, 7, 9, 21, 26	6, 7, 9, 26	1, 2, 6, 7, 9, 10, 12, 15, 16, 17, 20, 21, 23, 24, 25, 26, 31, 33, 37, 44, 45, 49	/	19	1, 2, 6, 7, 8, 9, 10, 12, 15, 16, 17, 19, 20, 21, 23, 24, 25, 26, 31, 33, 37, 38, 44, 45, 49 (=25)
Orphan et al., 2001	/	18	14, 16	14, 16	18, 2	2, 13, 14, 15, 18, 23, 42	2, 10, 13	/	2, 10, 13, 14, 15, 16, 18, 23, 42 (=9)
Sievert et al., 2000	10, 11, 12, 13, 25, 26, 27, 28, 29, 34	6, 7, 19	5, 14, 19, 22, 23, 24, 30, 31, 33, 34, 35	5, 14, 19, 22, 23, 24, 30, 31, 33, 34, 35	2, 5, 8, 14, 19, 22, 23, 24, 30, 31, 33, 34, 35	1, 2, 3, 4, 5, 6, 7, 8, 15, 16, 20, 21, 22, 24	/	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	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 (=35)
Chaudhary et al., 2009	12, 24, 28, 33, 57, 59, 65	24, 27, 34, 40, 56, 60, 61, 75	27	27	27, 40, 43, 75	2, 3, 5, 8, 12, 20, 25, 27, 48, 56, 68, 75	/	/	2, 3, 5, 8, 12, 20, 24, 25, 27, 28, 33, 34, 40, 43, 48, 56, 57, 59, 60, 61, 65, 68, 75 (=23)
Knittel et al., 2005	/	2, 10, 26, 29, 34, 50, 60, 63, 68	2, 44, 45, 62, 63	2, 44, 45, 62, 63	2, 19, 26, 45, 63	5, 16, 23, 35, 42, 43, 44, 45, 46, 50, 59, 60	31, 34, 45, 63	35, 43, 44, 46	2, 5, 10, 16, 19, 23, 26, 29, 31, 34, 35, 42, 43, 44, 45, 46, 50, 59, 60, 62, 63, 68 (=22)
Oline et al., 2006	2, 3, 4, 12, 13, 15, 16, 18, 19, 21, 22, 23, 25, 26, 28, 29, 32, 33, 34, 35	25, 28	/	/	35	23	/	/	2, 3, 4, 12, 13, 15, 16, 18, 19, 21, 22, 23, 25, 26, 28, 29, 32, 33, 34, 35 (=20)

Fig. S1. A non-metric multidimensional scaling (NMDS) ordination of archaeal unweighted (A), weighted (B) and bacterial unweighted (C) and weighted (D) composite datasets of the pooled datasets analyzed by RDP Classifier files according to the chimera identification approach. Each dot represents the phylogenetic signal of a pool of seven clone libraries or their CIA filtered datasets analyzed by a ingle chimera identification approach
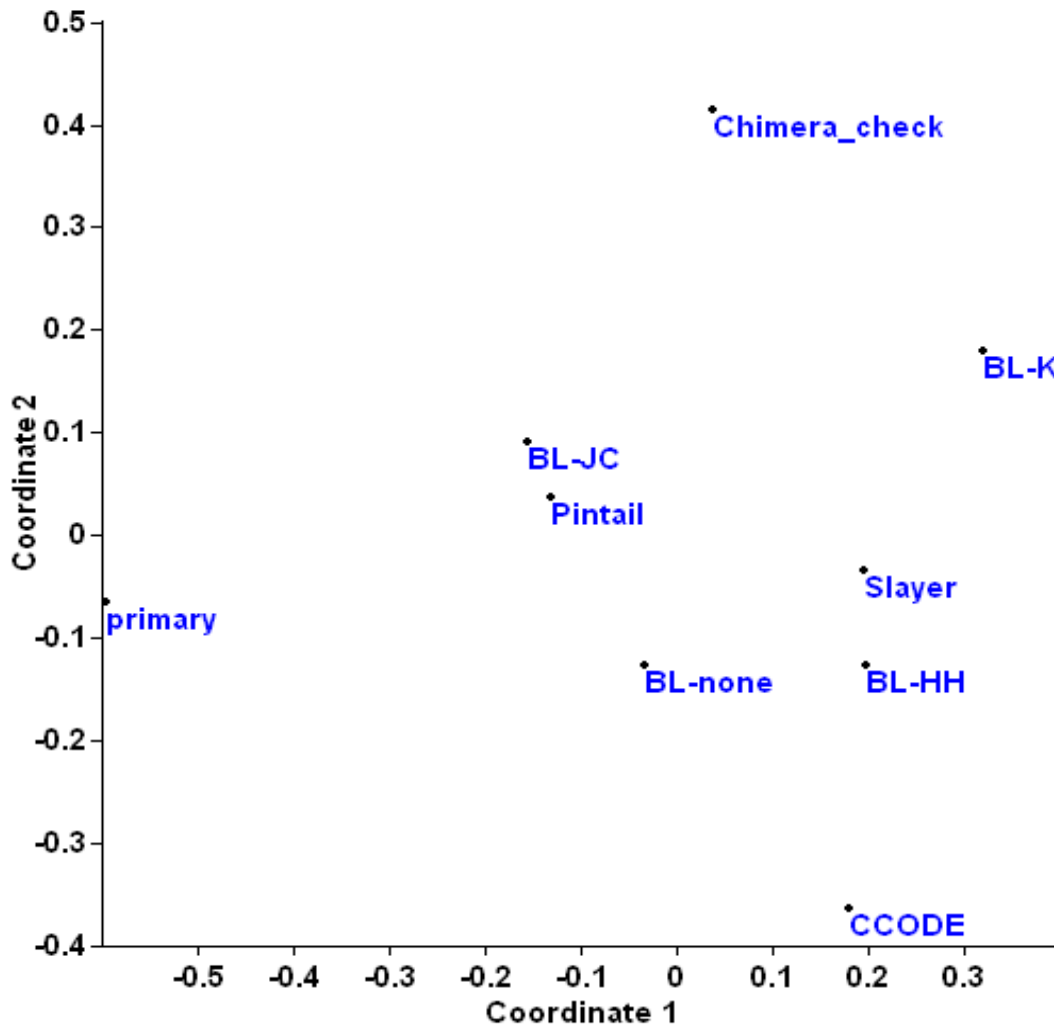

Fig. S1A. Archaea unweighted UniFrac - chimera identification approach congruency using Classifier hierarchy NMDS
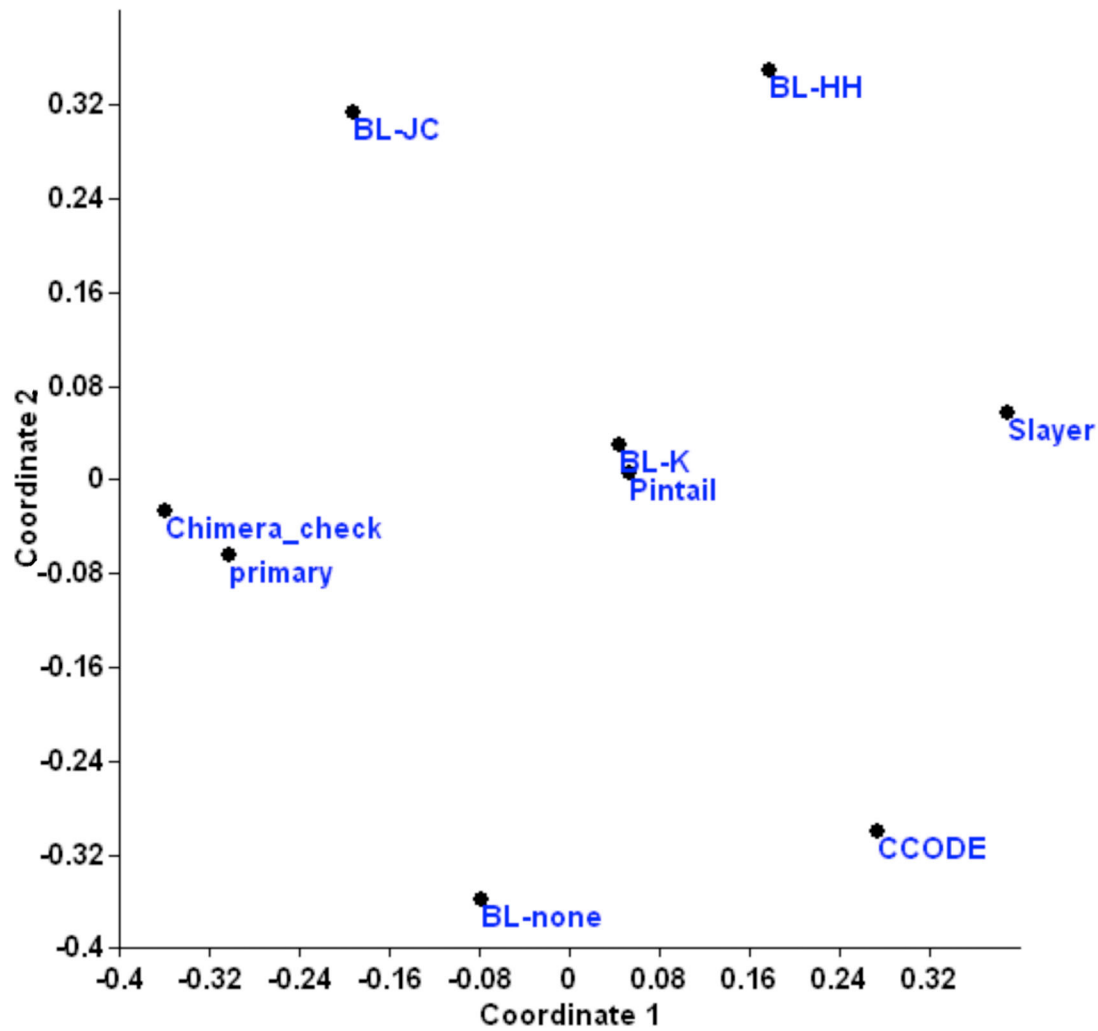
Fig. S1B. Archaea weighted UniFrac - chimera identification approach congruency using Classifier hierarchy NMDS
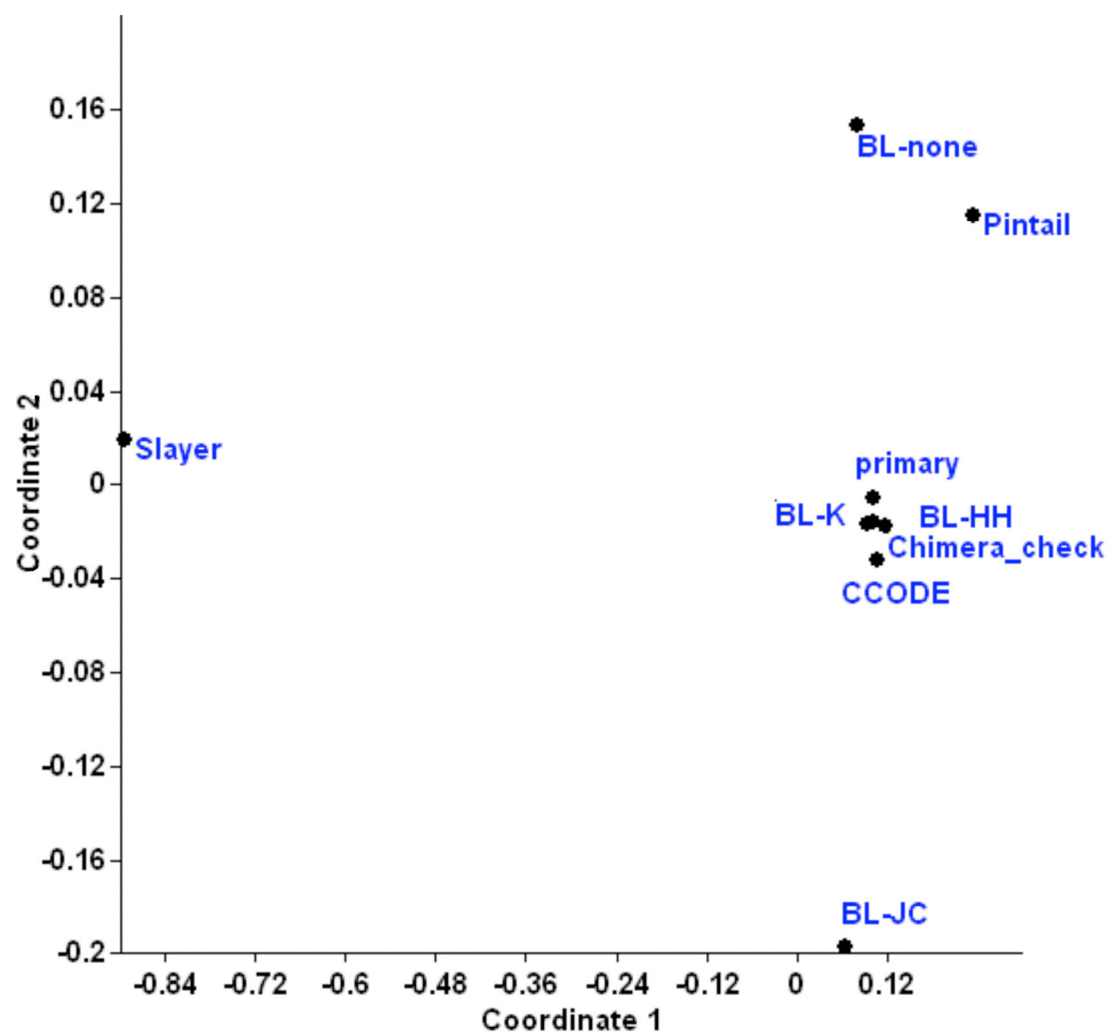
Fig. S1C. Bacteria unweighted UniFrac - chimera identification approach congruency using Classifier hierarchy NMDS
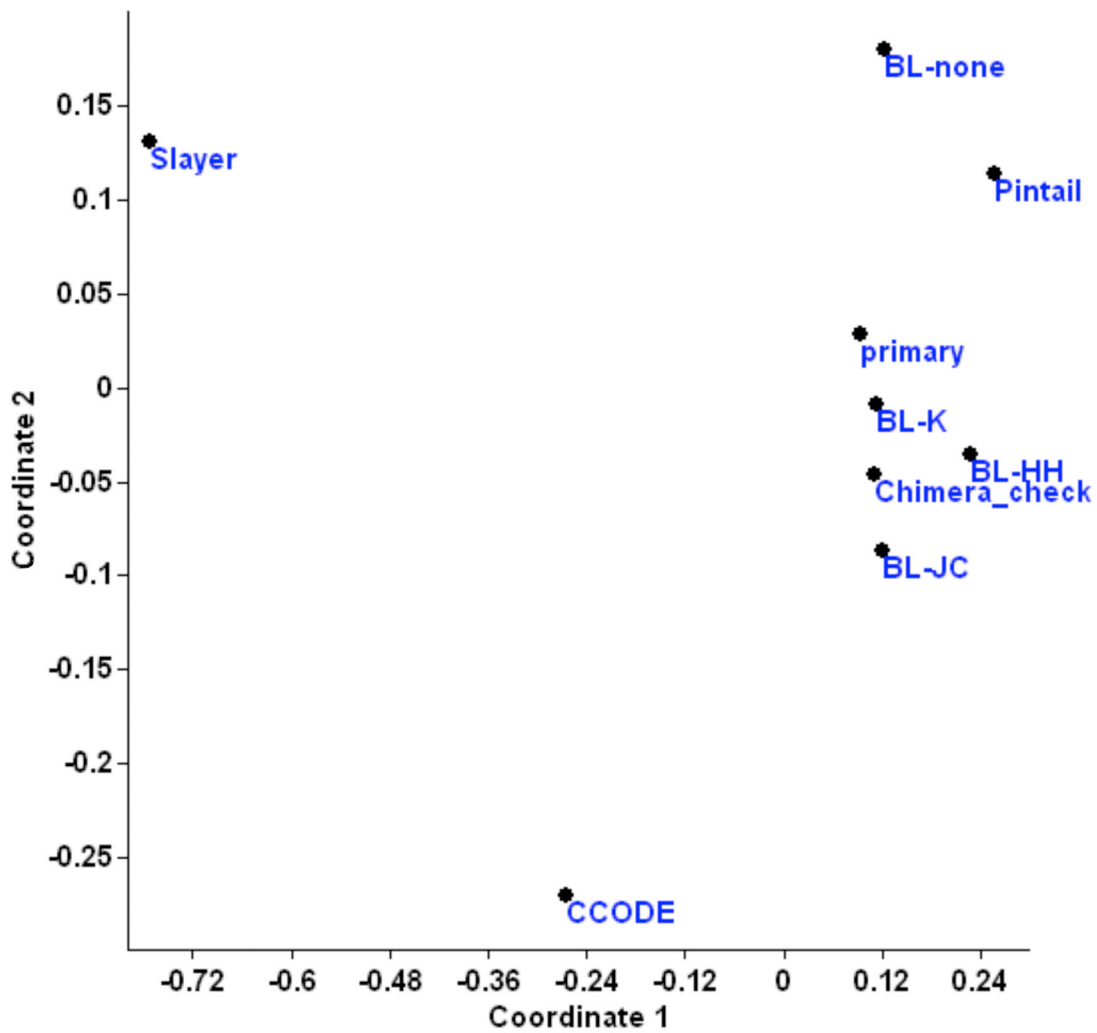
Fig. S1D. Bacteria weighted UniFrac - chimera identification approach congruency using Classifier hierarchy NMDS

Fig. S2. A NMDS ordination of archaeal unweighted (A), weighted (B) and bacterial unweighted (C) and weighted (D) datasets of the Classifier hierarchy files organized according to the clone library. Each dot represents the phylogenetic signal of a single clone library or its CIA filtered dataset generated by one of the eight chimera identification approaches. Colour keys are presented in Figs. S2B and S2D for archaeal and bacterial datasets, respectively
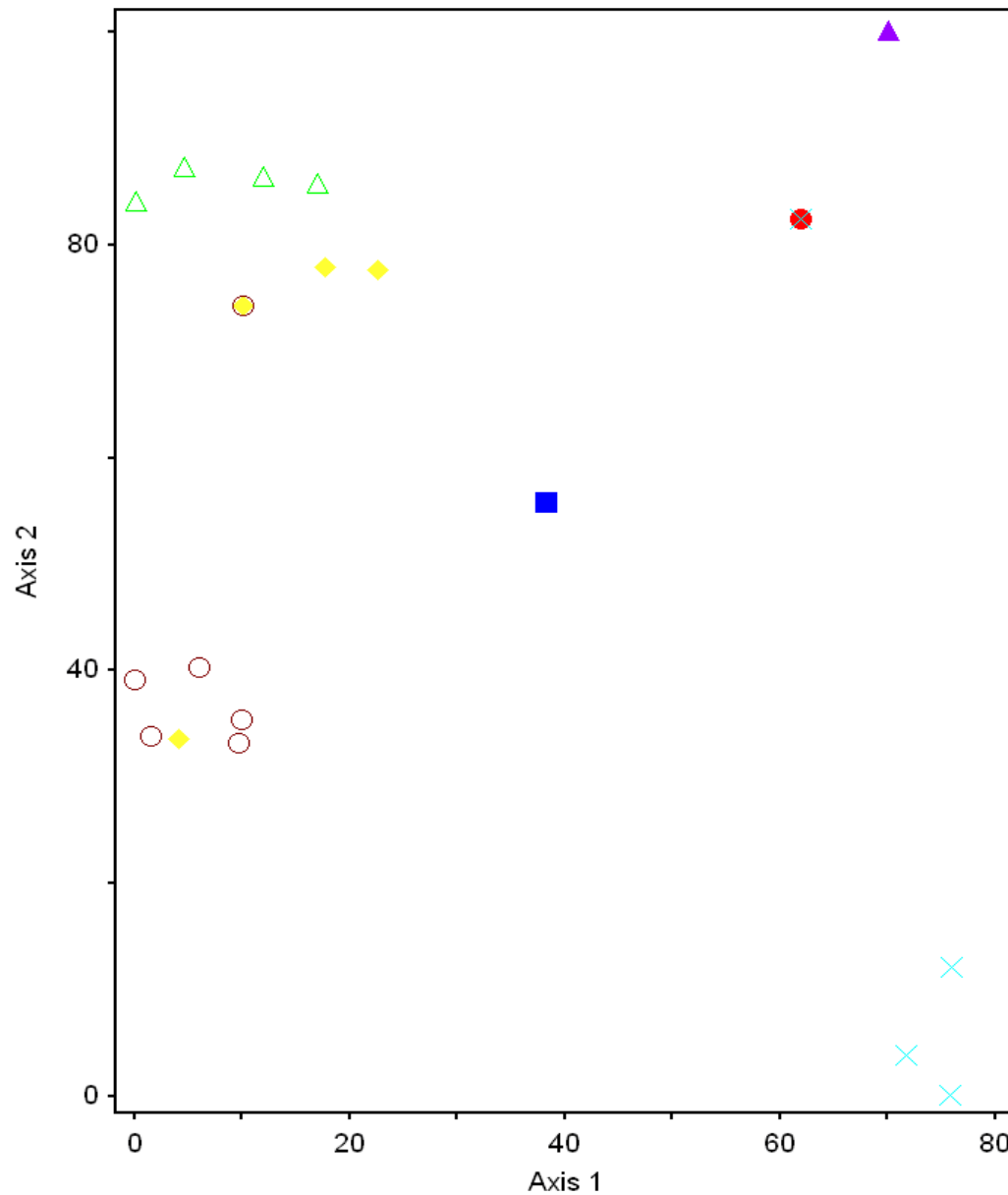
Fig. S2A. Archaeal unweighted datasets according to clone libraries, RDP Classifier hierarchy files
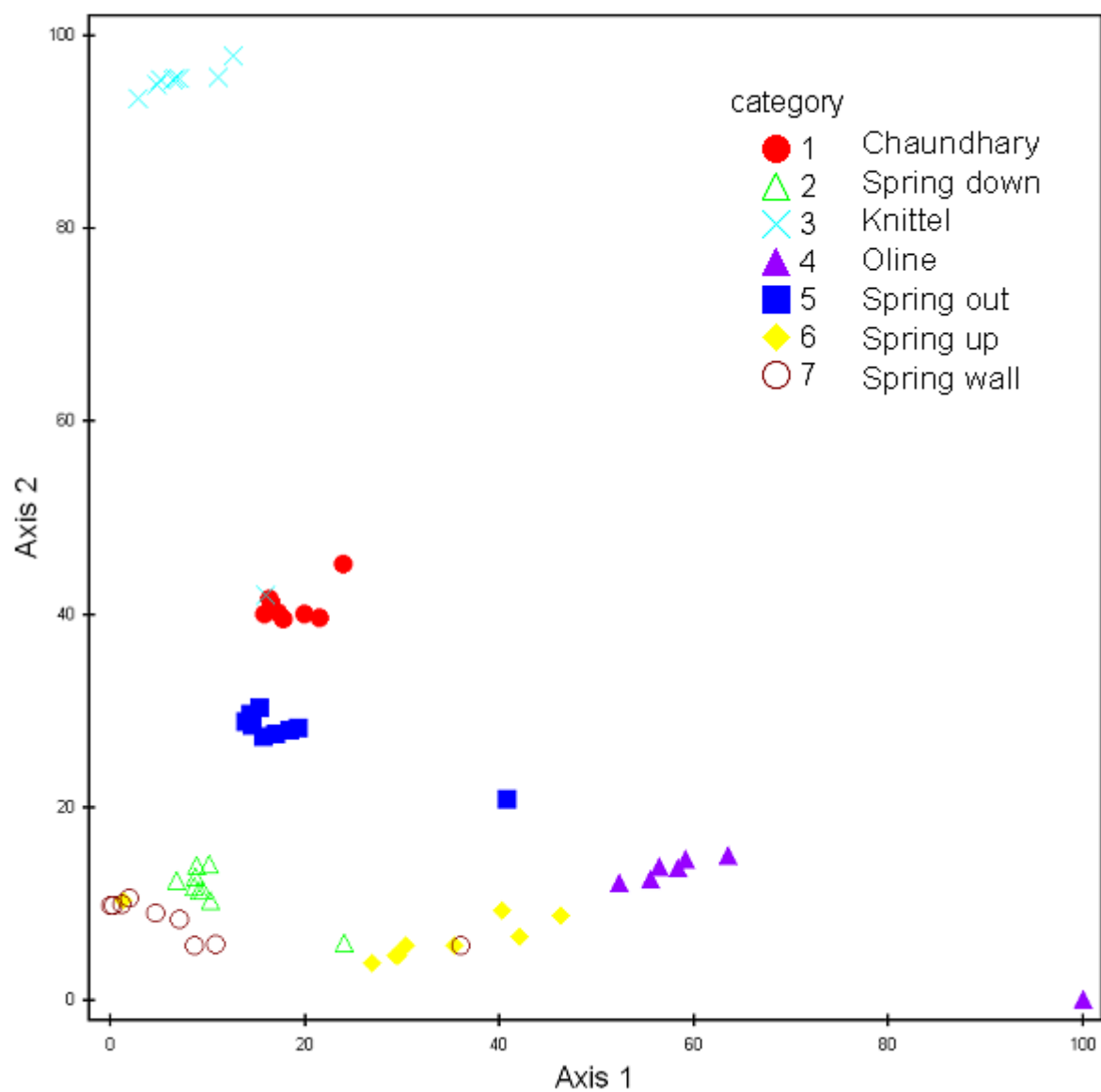
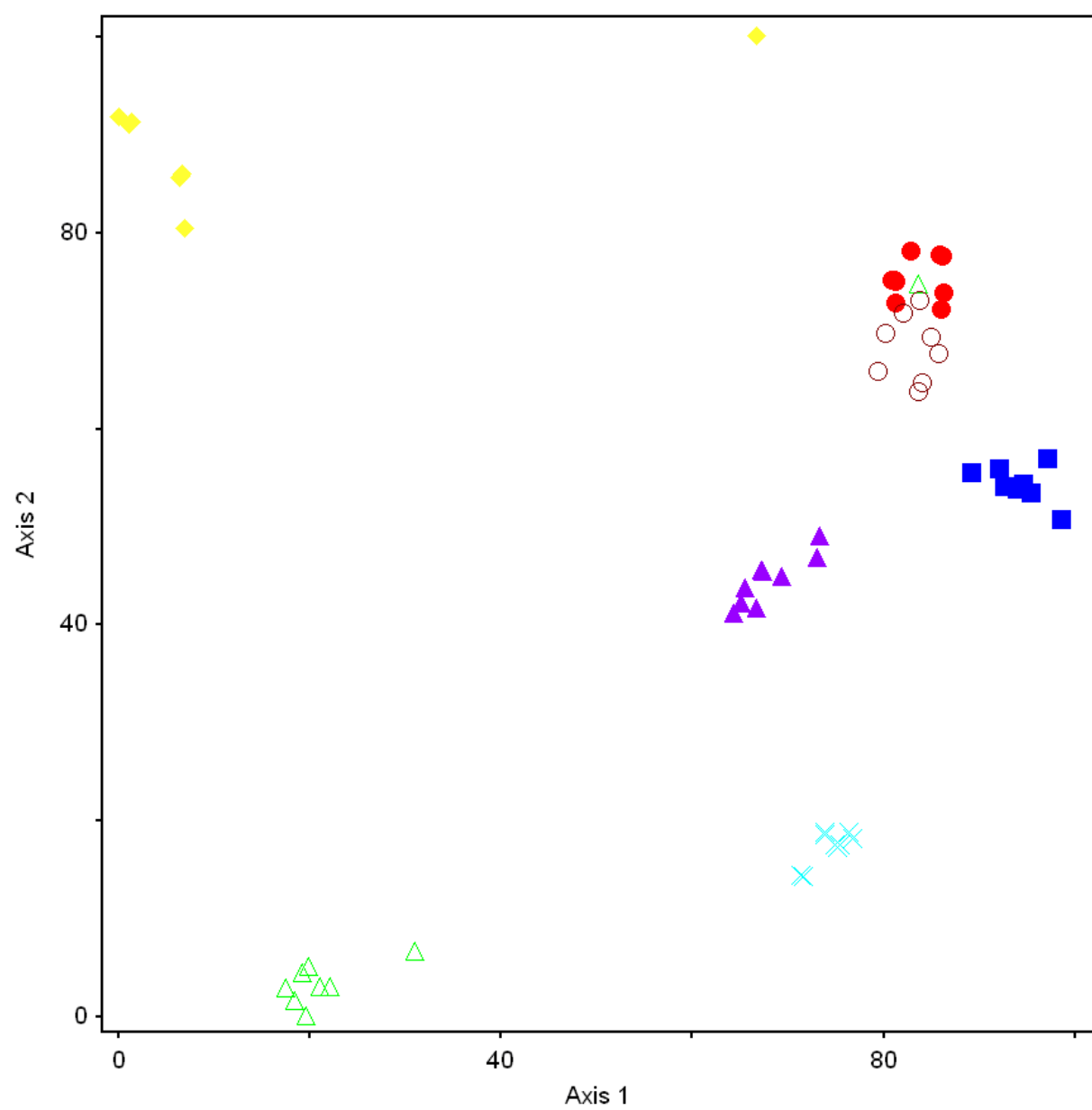**Fig. S2B.** Archaea weighted datasets according to clone libraries, RDP Classifier hierarchy files

Fig. S2C. Bacteria unweighted datasets according to clone libraries, RDP Classifier hierarchy files
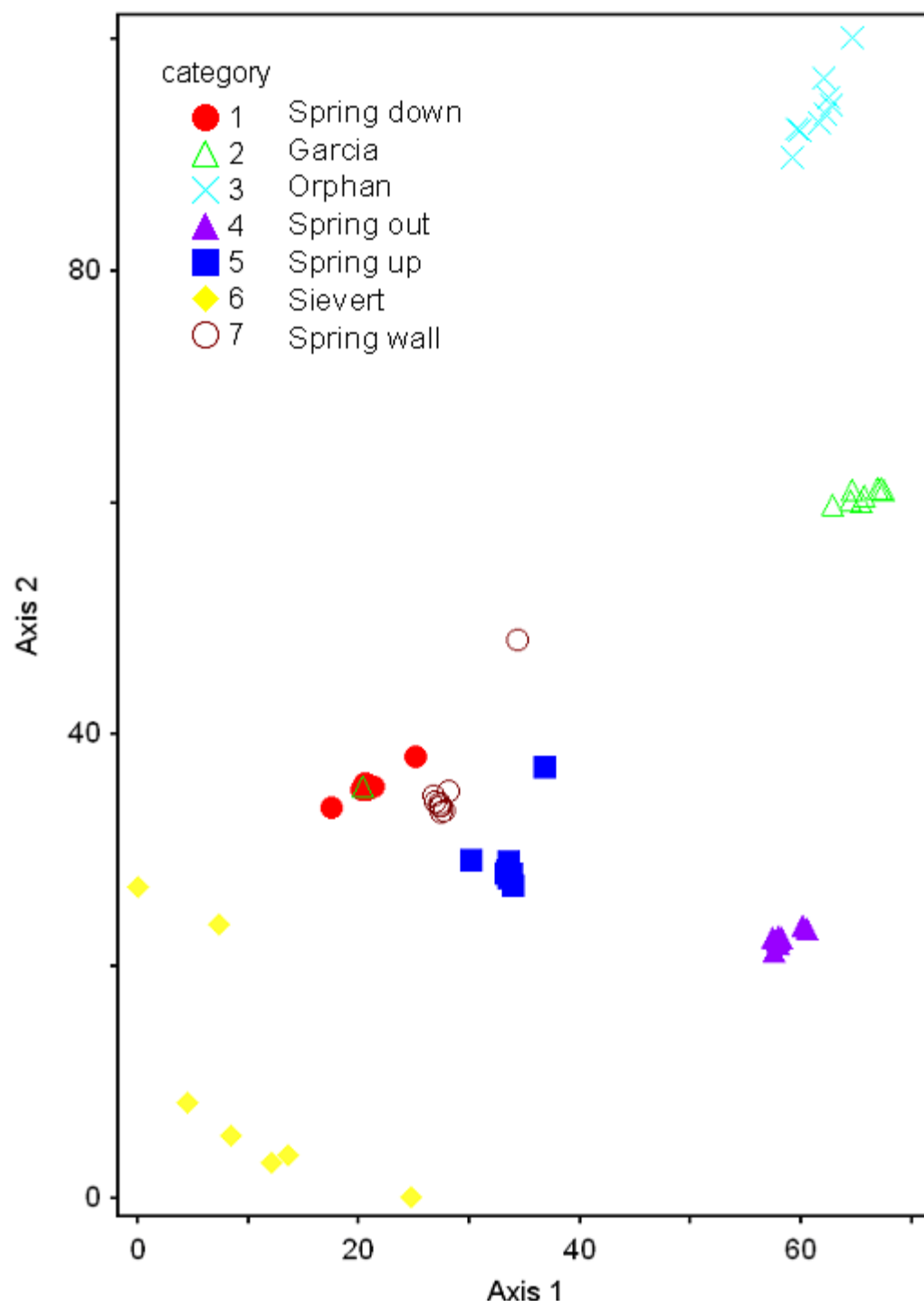
Fig. S2D. Bacteria weighted datasets according to clone libraries, RDP Classifier hierarchy files