

Supplementary Material

Supplementary Table 1 – DARTS-Sub clusters. Cluster: DART cluster index; Annotation: summary of the longest sequence annotation in the cluster; PDB: PDB id of the longest chain; %ID average ID measured on the SARA-Coffee MSA; N: Number of sequences; L: average length,; BP: fraction of nucleotides forming BP pairs. Bold: High BP fraction

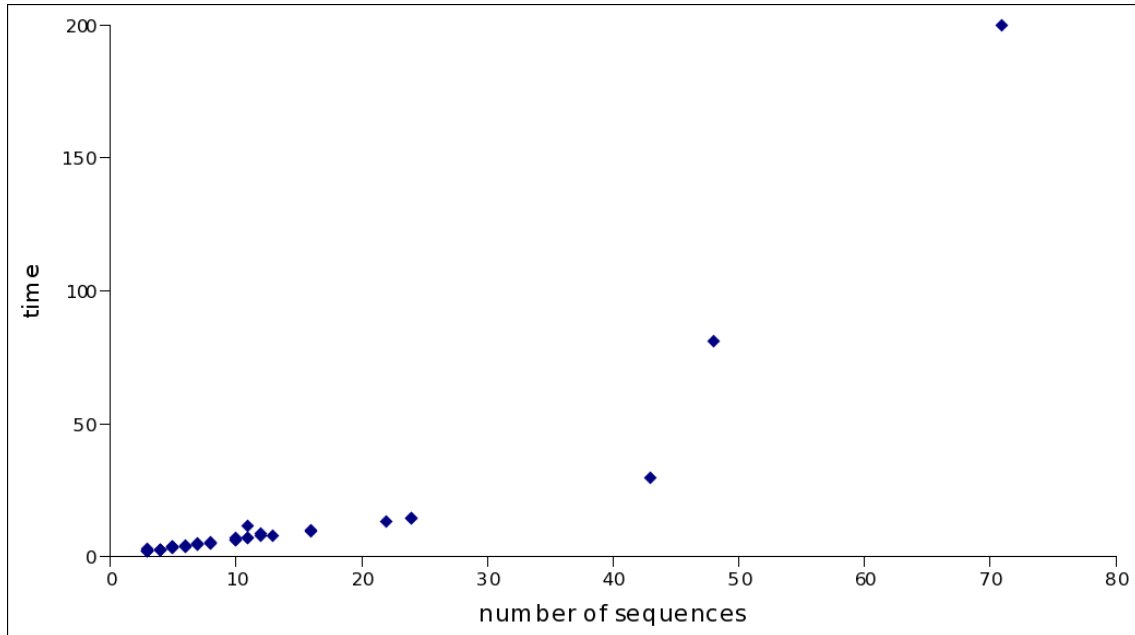
Cluster	Annotation	PDB	%ID	N	L	BP
13	rRNA e-loop	1D6K_A	66.14	4	31	0.90
22	Loop E Region of the 5S rRNA	1MNX_A	63.23	11	29	0.86
30	mRNA	1S03_A	61.34	12	39	0.86
77	Zinc finger	1UN6_E	85.91	3	59	0.86
15	ribozyme substrat	1E4P_A	40.56	16	21	0.84
12	self-splicing Intron	1KXK_A	67.73	8	47	0.82
36	RNASE P	1F6X_A	75.85	8	26	0.82
45	IRES	1KP7_A	31.71	3	25	0.82
19	Translational frameshift signal (stem loop)	1ZC5_A	61.52	10	31	0.78
3	RNA-binding domain	1EKZ_B	39.27	13	23	0.78
91	snRNA	2OZB_C	54.76	7	30	0.78
31	ribozyme	1X8W_D	80.81	11	213	0.76
70	PRE-tRNA	1A9L_A	65.23	5	36	0.76
17	REV RESPONSE ELEMENT	1ETF_A	42.25	6	28	0.74
21	tRNA	1WZ2_C	57.66	71	76	0.74
34	RNA hairpin	1QWB_A	86.36	4	23	0.74
79	5S rRNA	1Y69_9	72.07	48	122	0.74
38	TAR RNA	1LVJ_A	83.92	10	29	0.72
59	Ribonuclease P	1U9S_A	51.05	3	157	0.72
89	uPsi RNA	2IHX_B	71.93	5	61	0.72
32	RNA pseudoknot	1RNK_A	61.06	6	31	0.70
65	IRES	1P5N_A	32.81	3	27	0.70
78	ribozyme precursor	1VC7_B	99.21	12	75	0.70
86	riboswitch	2HOK_A	85.17	7	81	0.70
55	5S rRNA fragment	1MJI_C	49.33	3	31	0.68
68	A-SITE 16S	1BYJ_A	91.39	6	23	0.68
10	ribozyme	2LDZ_A	43.43	8	27	0.66
54	50S rRNA segment	2HEM_A	35.86	5	18	0.64
11	SRP RNA DOMAIN IV	1CQ5_A	78.87	11	36	0.62
49	MLV PSI site	1S9S_A	50.72	5	62	0.62
67	transfer-RNA domain	1P6V_B	62.16	5	30	0.58
35	tRNA	1J2B_C	44.62	10	49	0.56
47	S15-rRNA	1DK1_B	55.8	7	47	0.52
51	U6 ISL	1XHP_A	44.4	24	15	0.44
42	snR47 precursor	1T4L_A	36.52	24	16	0.40

88	h/ACA RNPs	2HVY_E	33.04	3	36	0.40
28	ribozyme	1YKV_D	51.56	22	22	0.30
29	Hammerhead ribozyme	1HMH_E	39.94	4	24	0.30
60	RNA-aptamer	1NTA_A	55.56	4	20	0.18
56	tetraloop-receptor RNA	2ADT_A	72.12	43	24	0.12
27	T7 RNA Polymerase	1QLN_T	30.76	16	16	0.06

Supplementary Table 2 – Composition of the BRAliDARTS

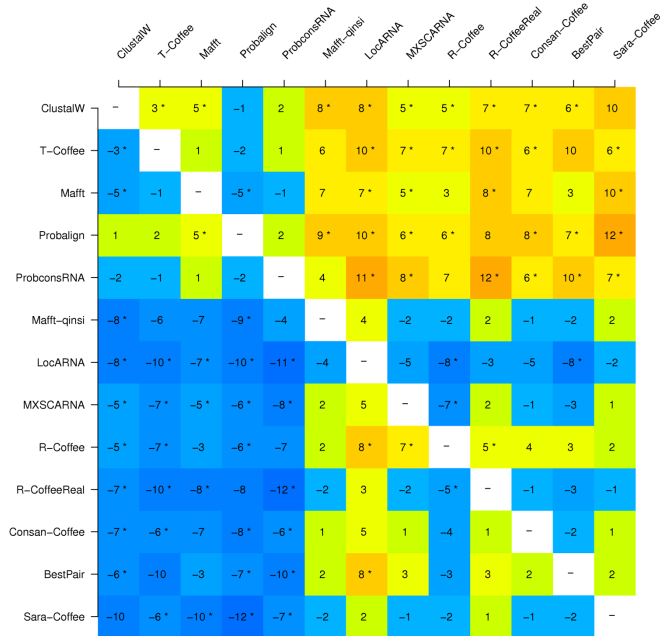
Type	# seqs	%
rRNA	117	24.1
tRNA	82	16.9
ribozyme	66	13.6
mRNA	19	3.9
virus	16	3.3
Responceelement	12	2.5
tar	10	2.1
RNA_adapter	9	1.9
Sep	9	1.9
snRNA	9	1.9
riboswitch	7	1.4
siRNA	4	0.8
SRE	3	0.6
SRP	2	0.4
snoRNA	1	0.2
other	94	19.3
DNA	26	5.3
SUM	486	100.0

Supplementary Figure 1 – SARA-Coffee running time for BRAlidARTS clusters. The x-axis shows the number of sequences included in each dataset. The vertical axis represents the single processor running time (user time plus sys time) expressed in seconds.

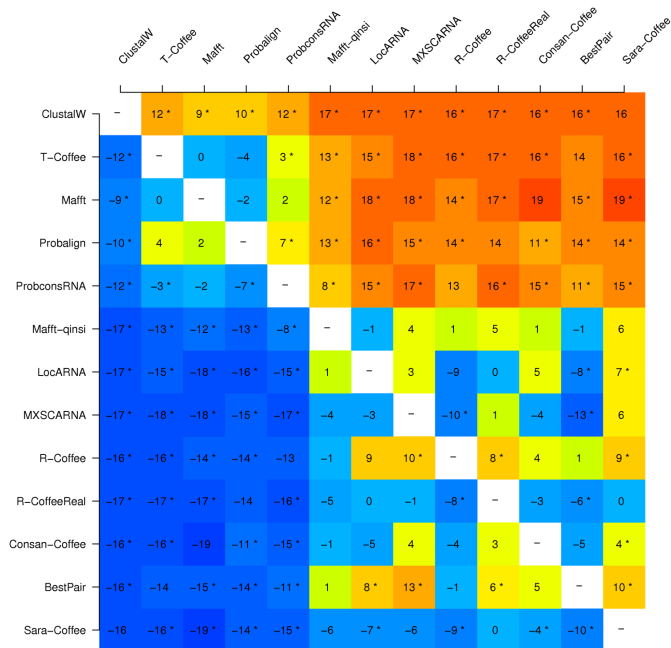


Supplementary Figure 2 – Results of 3SP score. The intersection between two methods denotes the difference between the number of times the method on the top wins over the method on the vertical axis (e.g. Sara-Coffee won on 11 datasets more than ClustalW in the case of high structured datasets). Stars denote a significant difference in the two distributions according to the Wilcoxon-Test with a p-value <0.05.

a) Low structured dataset

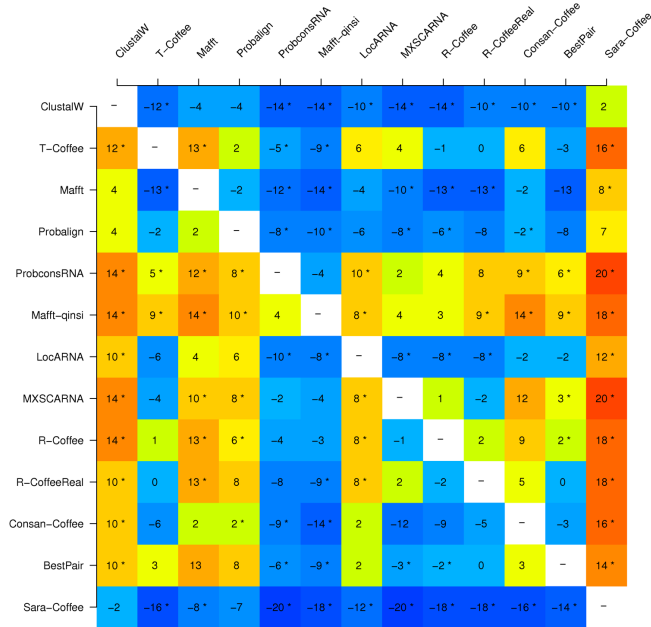


b) High structured dataset



Supplementary Figure 3 – Results of NiRMSD. The intersection between two methods denotes the difference between the number of times the method on the top wins over the method on the vertical axis (e.g. Sara-Coffee won on 11 datasets more than ClustalW in the case of high structured datasets). Stars denote a significant difference in the two distributions according to the Wilcoxon-Test with a p-value <0.05.

a) Low structured dataset



b) High structured

