Supplementary Materials

Table 1: Accuracy of COATi codon-triplet-mg, PRANK, MAFFT, ClustalOmega, and MACSE on 7719 simulated sequence pairs. Perfect alignments have $d_{seq} = 0$, best alignments have lowest d_{seq}, and imperfect alignments have $d_{seq} > 0$ when at least one aligner found a perfect alignment.

	tri-mg	PRANK	MAFFT	ClustalOmega	MACSE
d_{seq}	0.00214	0.02002	0.01390	0.02680	0.01351
Wilcoxon rank test	NA	0.00000	0.00000	0.00000	0.00000
Perfect alignments	4123.00000	4259.00000	4325.00000	2449.00000	3450.00000
Best alignments	5120.00000	4747.00000	4803.00000	2583.00000	3733.00000
Imperfect alignments	2064.00000	1928.00000	1862.00000	3738.00000	2737.00000
F1-score pos selection	0.98231	0.88429	0.86190	0.71268	0.81337
F1-score neg selection	0.99820	0.98838	0.98568	0.97068	0.98292

Table 2: Accuracy of COATi codon-triplet-ecm, PRANK, MAFFT, ClustalOmega, and MACSE on 7678 simulated sequence pairs. Perfect alignments have $d_{seq} = 0$, best alignments have lowest d_{seq}, and imperfect alignments have $d_{seq} > 0$ when at least one aligner found a perfect alignment.

	tri-ecm	PRANK	MAFFT	ClustalOmega	MACSE
d_{seq}	0.00243	0.02051	0.01464	0.02825	0.01367
Wilcoxon rank test	NA	0.00332	0.00003	0.00000	0.00000
Perfect alignments	3996.00000	4192.00000	4251.00000	2487.00000	3423.00000
Best alignments	4931.00000	4763.00000	4707.00000	2642.00000	3713.00000
Imperfect alignments	2109.00000	1913.00000	1854.00000	3618.00000	2682.00000
F1-score pos selection	0.97192	0.87719	0.84912	0.71327	0.80192
F1-score neg selection	0.99719	0.98792	0.98447	0.97102	0.98239

Table 3: Accuracy of COATi codon-marginal-mg, PRANK, MAFFT, ClustalOmega, and MACSE on 7661 simulated sequence pairs. Perfect alignments have $d_{seq} = 0$, best alignments have lowest d_{seq}, and imperfect alignments have $d_{seq} > 0$ when at least one aligner found a perfect alignment.

	mar-mg	PRANK	MAFFT	ClustalOmega	MACSE
d_{seq}	0.00200	0.01813	0.01370	0.02625	0.01309
Wilcoxon rank test	NA	0.17142	0.08244	0.00000	0.00000
Perfect alignments	3701.00000	4283.00000	4267.00000	2432.00000	3400.00000
Best alignments	4700.00000	4816.00000	4780.00000	2592.00000	3746.00000
Imperfect alignments	2378.00000	1796.00000	1812.00000	3647.00000	2679.00000
F1-score pos selection	0.96426	0.89164	0.85957	0.72496	0.82027
F1-score neg selection	0.99646	0.98924	0.98566	0.97174	0.98374

Table 4: Accuracy of COATi codon-marginal-ecm, PRANK, MAFFT, ClustalOmega, and MACSE on 7710 simulated sequence pairs. Perfect alignments have $d_{seq} = 0$, best alignments have lowest d_{seq}, and imperfect alignments have $d_{seq}>0$ when at least one aligner found a perfect alignment.

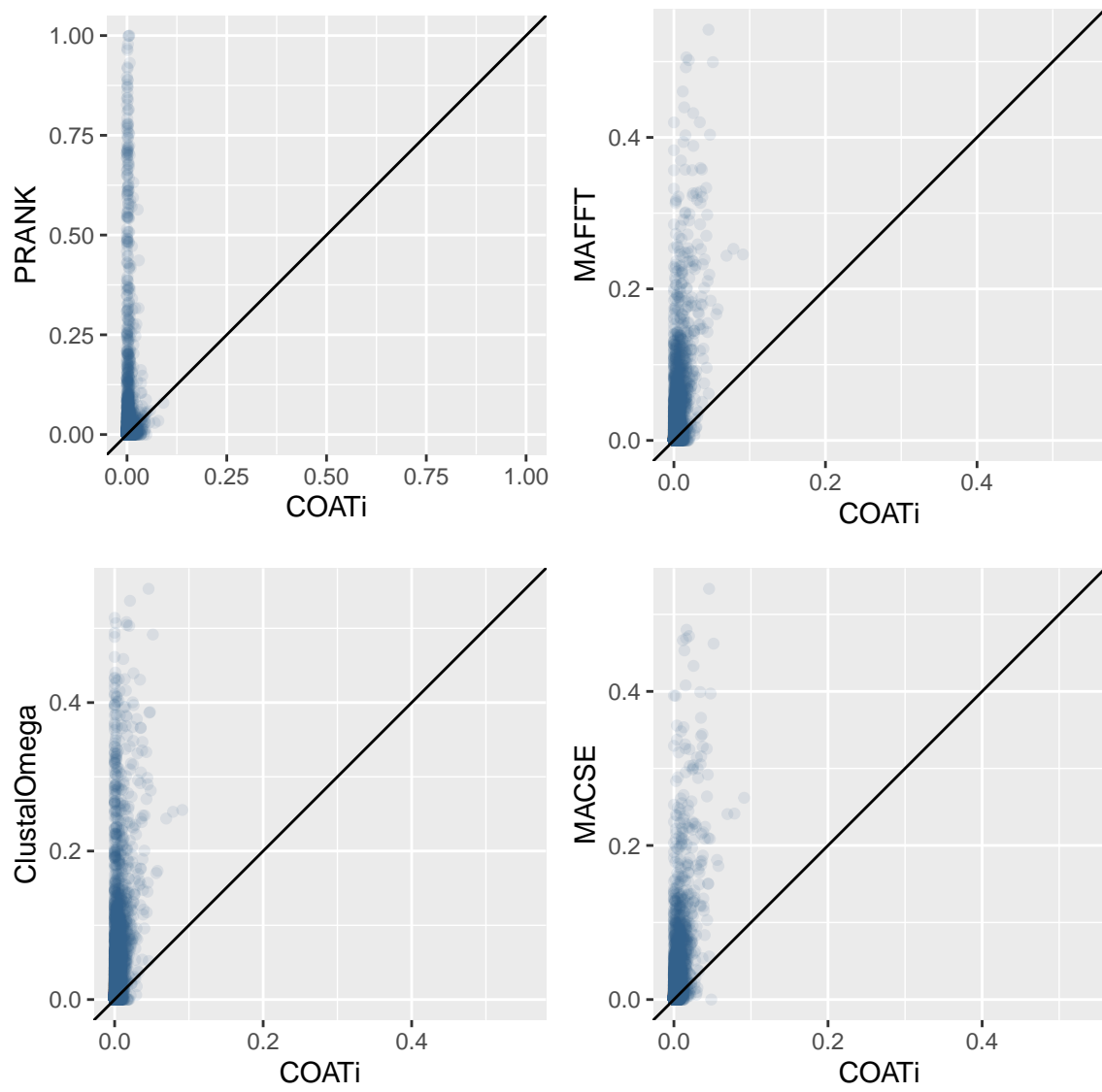|	mar-ecm	PRANK	MAFFT	ClustalOmega	MACSE
d_{seq}	0.00209	0.02089	0.01440	0.02809	0.01385
Wilcoxon rank test	NA	0.04707	0.08670	0.00000	0.00000
Perfect alignments	3589.00000	4235.00000	4203.00000	2333.00000	3410.00000
Best alignments	4567.00000	4771.00000	4749.00000	2512.00000	3738.00000
Imperfect alignments	2510.00000	1864.00000	1896.00000	3766.00000	2689.00000
F1-score pos selection	0.96987	0.88415	0.85379	0.71409	0.79405
F1-score neg selection	0.99691	0.98831	0.98475	0.97053	0.98131

Figure 1: Comparison of d_{seq} values between COATi codon-triplet-mg and PRANK, MAFFT, ClustalOmega, and MACSE.
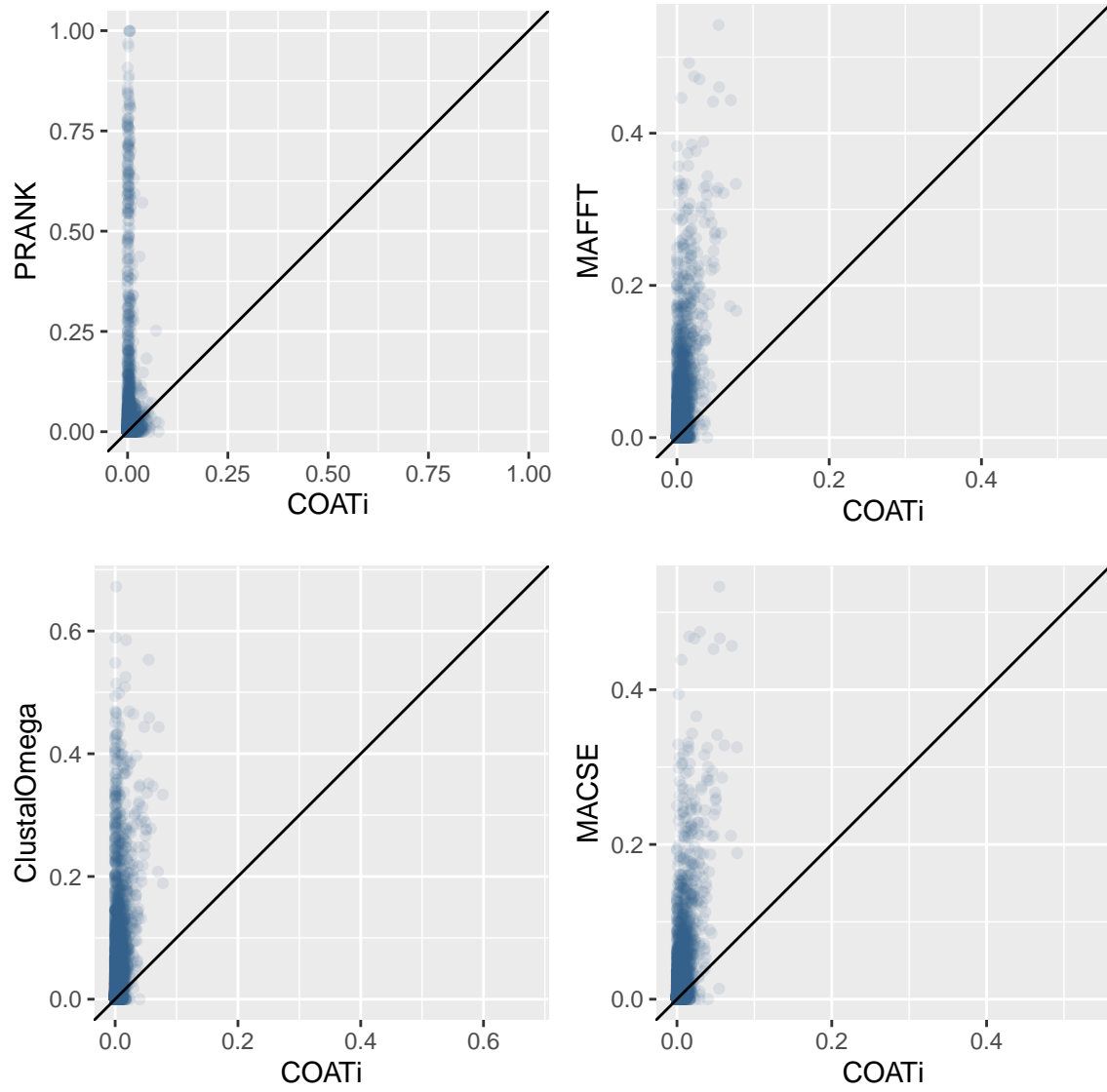
Figure 2: Comparison of d_{seq} values between COATi codon-triplet-ecm and PRANK, MAFFT, ClustalOmega, and MACSE.
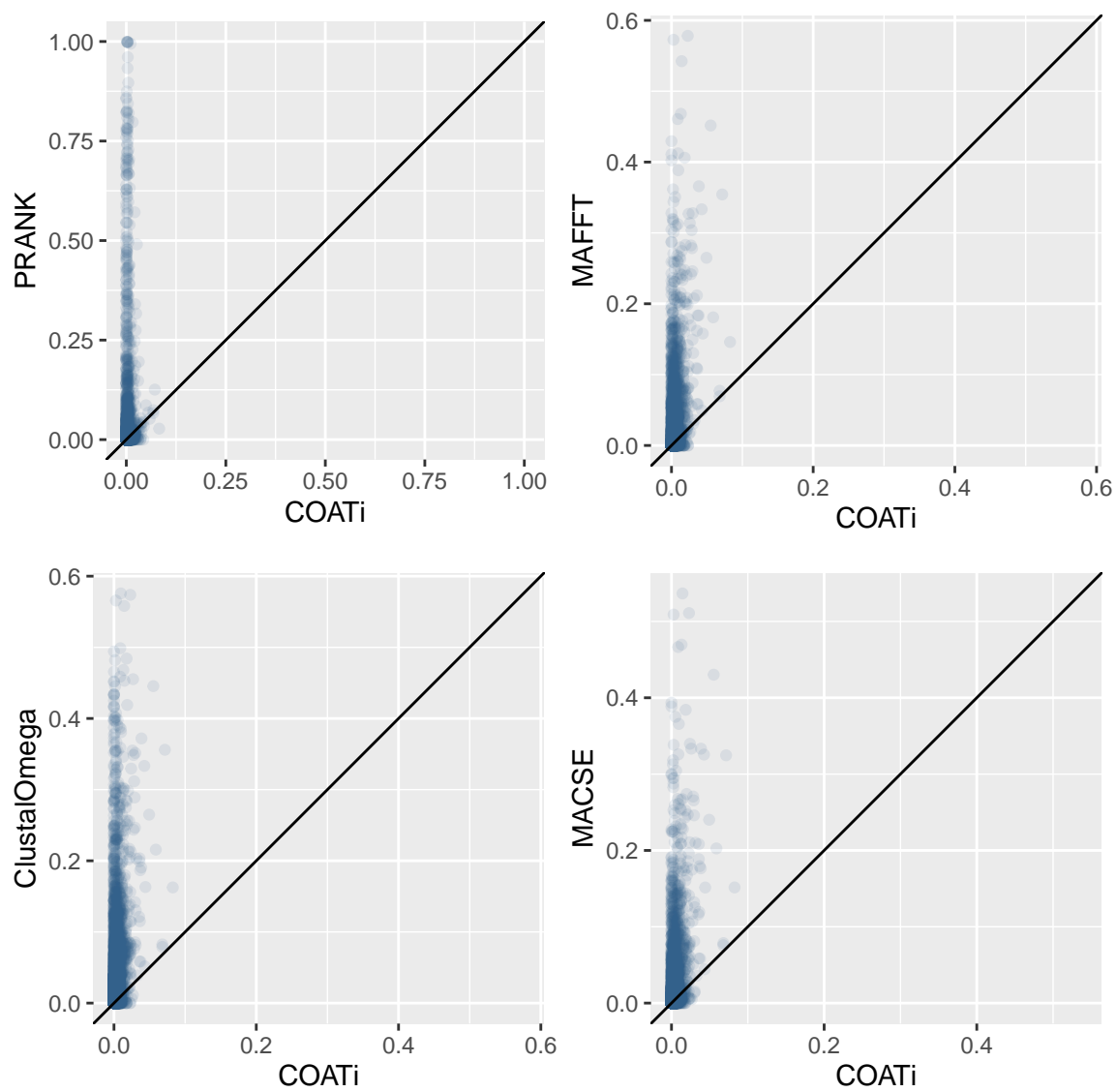
Figure 3: Comparison of d_{seq} values between COATi codon-marginal-mg and PRANK, MAFFT, ClustalOmega, and MACSE.
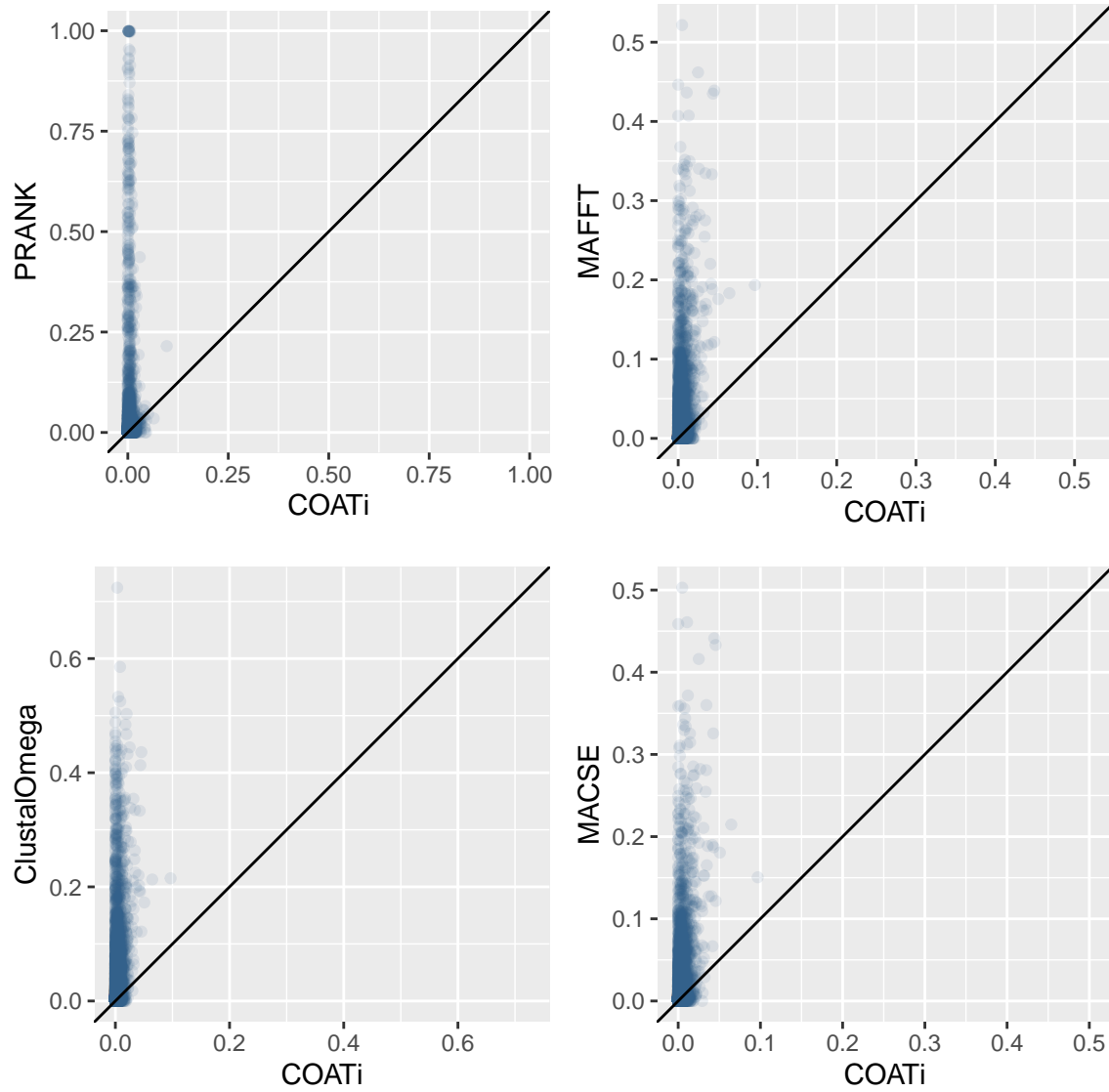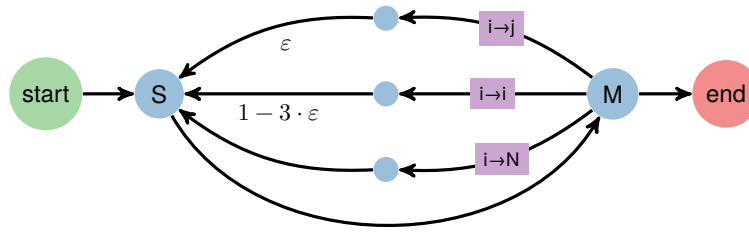
Figure 4: Comparison of d_{seq} values between COATi codon-marginal-ecm and PRANK, MAFFT, ClustalOmega, and MACSE.

Sequences

$i \rightarrow i$: matching input to intermediate base
$i \rightarrow j$: mismatching input to intermediate base
$i \rightarrow N$: input to intermediate ambiguous base N

Parameters

ε: base calling error weight

Figure 5: Base calling error FST. Arcs from M to S generate matches; however, here they can introduce single-nucleotide errors, which can generate stop codon artifacts.