Table 1: **Population diversity** (% of variants that are unique). Updated results are shown in blue. We see a noticeable improvement in diversity scores.

	Random search	Simulated annealing	MALA-*approx*	CMA-ES	PPDE (Potts only)	PPDE (Super. only)	PPDE (Potts)	PPDE (ESM2)	PPDE (Potts+ESM2)
PABP	32.8	28.9	28.9	0.8	**100.0**	71.9	**100.0**	81.2	**100.0**
UBE4B	7.0	4.7	6.2	3.1	**99.2**	46.9	93.8	68.8	82.0
GFP	9.4	3.9	9.4	92.2	63.3	68.8	86.7	**100.0**	75.0

	Log fitness↑ (Augmented EVmutation)			Evolutionary density↑ (MSA Transformer)			Exploration (mean±std # muts)		
Potts expert	PABP	UBE4B	GFP	PABP	UBE4B	GFP	PABP	UBE4B	GFP
PPDE	$1.32_{(2.42)}$	$1.32_{(3.45)}$	$0.13_{(0.55)}$	$2.09_{(14.55)}$	$-18.11_{(1.96)}$	$-21.85_{(-6.25)}$	$9.4_{\pm 0.9}$	$7.6_{\pm 1.8}$	$4.4_{\pm 1.6}$
Random search	$0.09_{(0.82)}$	$-0.19_{(0.34)}$	$-0.04_{(0.04)}$	$4.26_{(6.87)}$	$-1.09_{(2.46)}$	$-0.11_{(-0.11)}$	$1.3_{\pm 0.5}$	$1.1_{\pm 0.3}$	$1.0_{\pm 0.2}$
Sim. annealing	$0.09_{(0.44)}$	$-0.19_{(0.28)}$	$-0.04_{(0.10)}$	$3.55_{(8.63)}$	$-0.94_{(2.70)}$	$-5.89_{(-0.99)}$	$1.3_{\pm 0.5}$	$1.0_{\pm 0.2}$	$1.0_{\pm 0.1}$
MALA-*approx*	$0.09_{(0.56)}$	$-0.19_{(0.58)}$	$-0.04_{(0.10)}$	$2.25_{(5.14)}$	$-0.89_{(1.54)}$	$-6.74_{(-1.88)}$	$1.3_{\pm 0.5}$	$1.0_{\pm 0.2}$	$1.0_{\pm 0.2}$
CMA-ES	$1.37_{(1.37)}$	$2.54_{(2.54)}$	$-2.50_{(-0.15)}$	$3.47_{(3.47)}$	$-94.76_{(0.0)}$	$-62.43_{(0.0)}$	$17.0_{\pm 0}$	$15.5_{\pm 6.2}$	$10.2_{\pm 9.4}$
Unsupervised expert									
Potts only	$1.53_{(2.81)}$	$0.88_{(2.79)}$	$-0.14_{(0.69)}$	$1.76_{(18.67)}$	$-20.97_{(4.06)}$	$-12.83_{(-2.43)}$	$9.5_{\pm 0.7}$	$7.8_{\pm 1.7}$	$3.5_{\pm 2.3}$
Supervised only	$0.03_{(0.40)}$	$1.66_{(6.43)}$	$-0.25_{(0.12)}$	$-7.42_{(3.75)}$	$-6.17_{(-5.58)}$	$-13.85_{(-4.67)}$	$2.2_{\pm 1.1}$	$1.9_{\pm 1.2}$	$2.0_{\pm 1.0}$
ESM2	$0.14_{(0.74)}$	$2.14_{(8.43)}$	$-3.99_{(-2.13)}$	$-8.02_{(2.39)}$	$-18.54_{(-3.63)}$	$-92.66_{(-44.65)}$	$3.4_{\pm 2.1}$	$3.7_{\pm 2.5}$	$9.6_{\pm 1.0}$
Potts+ESM2	$0.88_{(2.05)}$	$2.50_{(4.50)}$	$0.08_{(0.63)}$	$-0.36_{(14.45)}$	$-24.78_{(-11.63)}$	$-22.82_{(-8.35)}$	$9.0_{\pm 1.0}$	$7.9_{\pm 1.4}$	$4.4_{\pm 1.9}$

Table 2: **$50^{\text{th}}_{(100^{\text{th}})}$ percentile scores**. Population size is 128. Updated PPDE results with a maximum number of 10 mutations enforced are shown in blue. Overall, we see improvements in diversity, log fitness, and exploration at the expense of decreases in evolutionary density. Throughout our experiments, we found that variants with higher numbers of mutations naturally tend to have lower MSA Transformer scores. We also observe that, in this case, combining Potts and ESM2 does not provide clear gains for PABP and GFP. The advantage of using an unsupervised expert trained on aligned sequences (the Potts expert) is more pronounced.