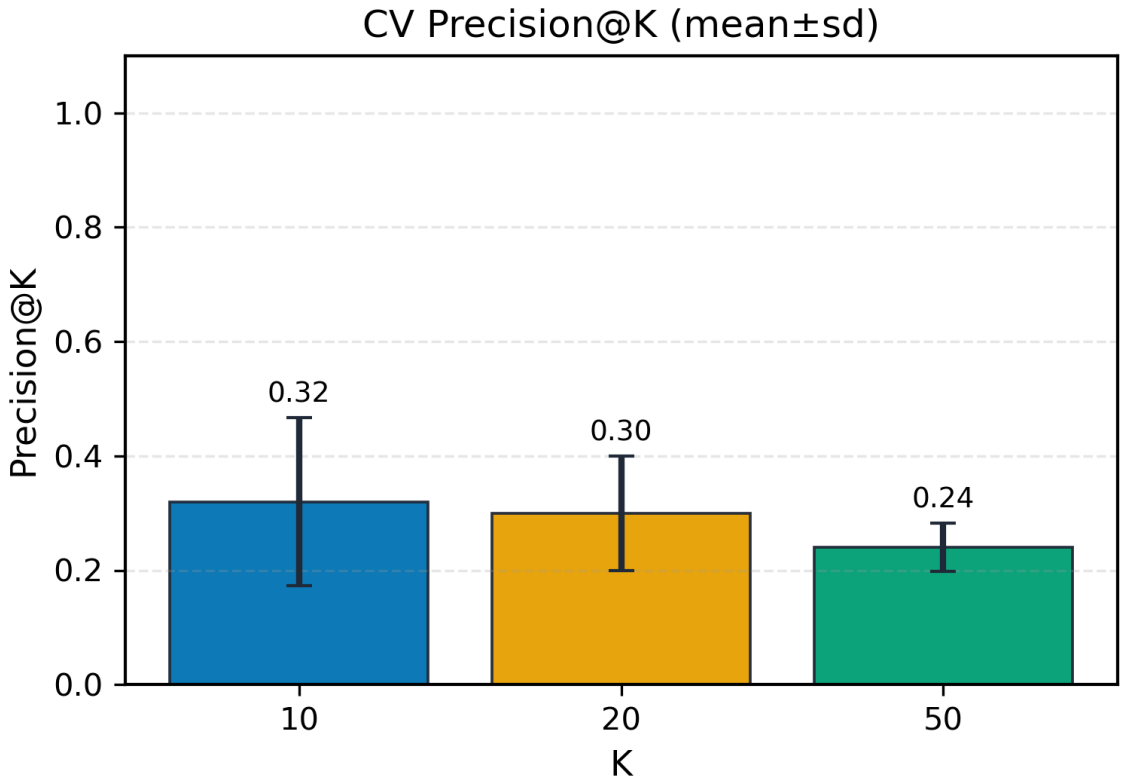
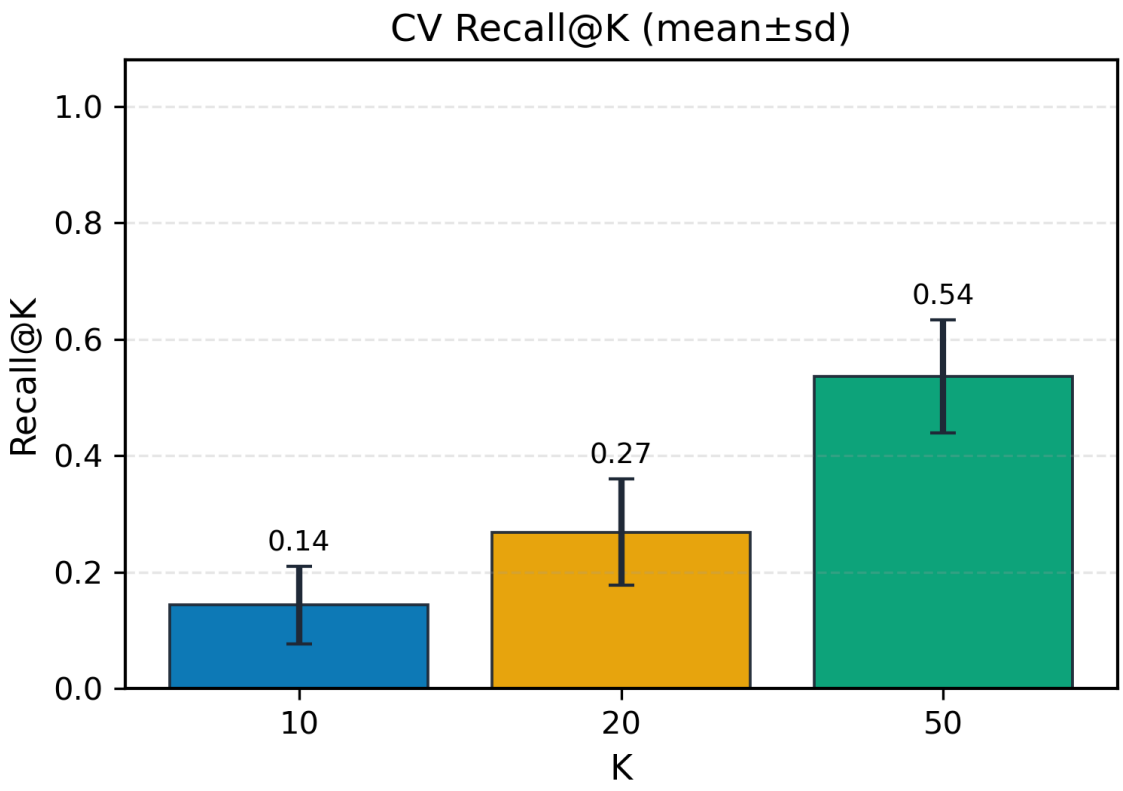
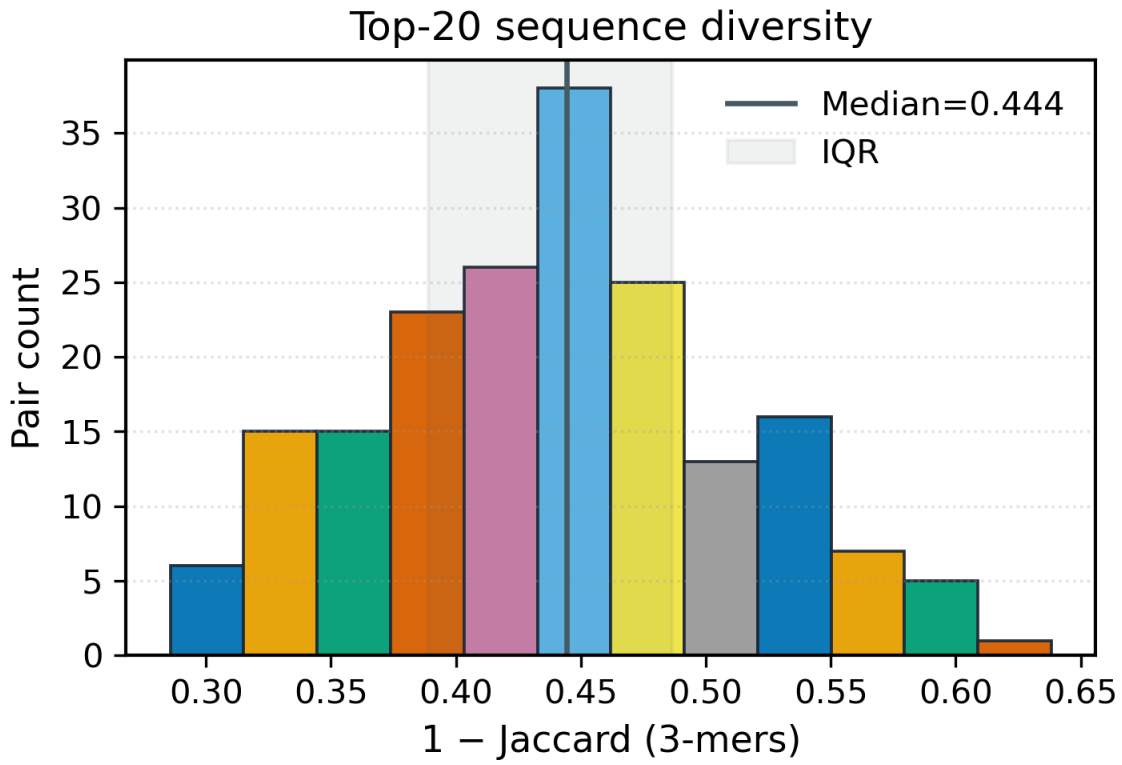
Supplementary Figures



Supplementary Figure S2a — 5-fold CV Precision@K (mean±sd). Means: 0.32 (K=10), 0.30 (K=20), 0.24 (K=50).



Supplementary Figure S2b — 5-fold CV Recall@K (mean±sd). Means: 0.14 (K=10), 0.27 (K=20), 0.54 (K=50).



Supplementary Figure S3 — Top-20 sequence diversity: 1 − Jaccard distance distribution on 3-mers. Median = 0.444; IQR shaded.