Cross-Contamination Risk from Sequencing Errors
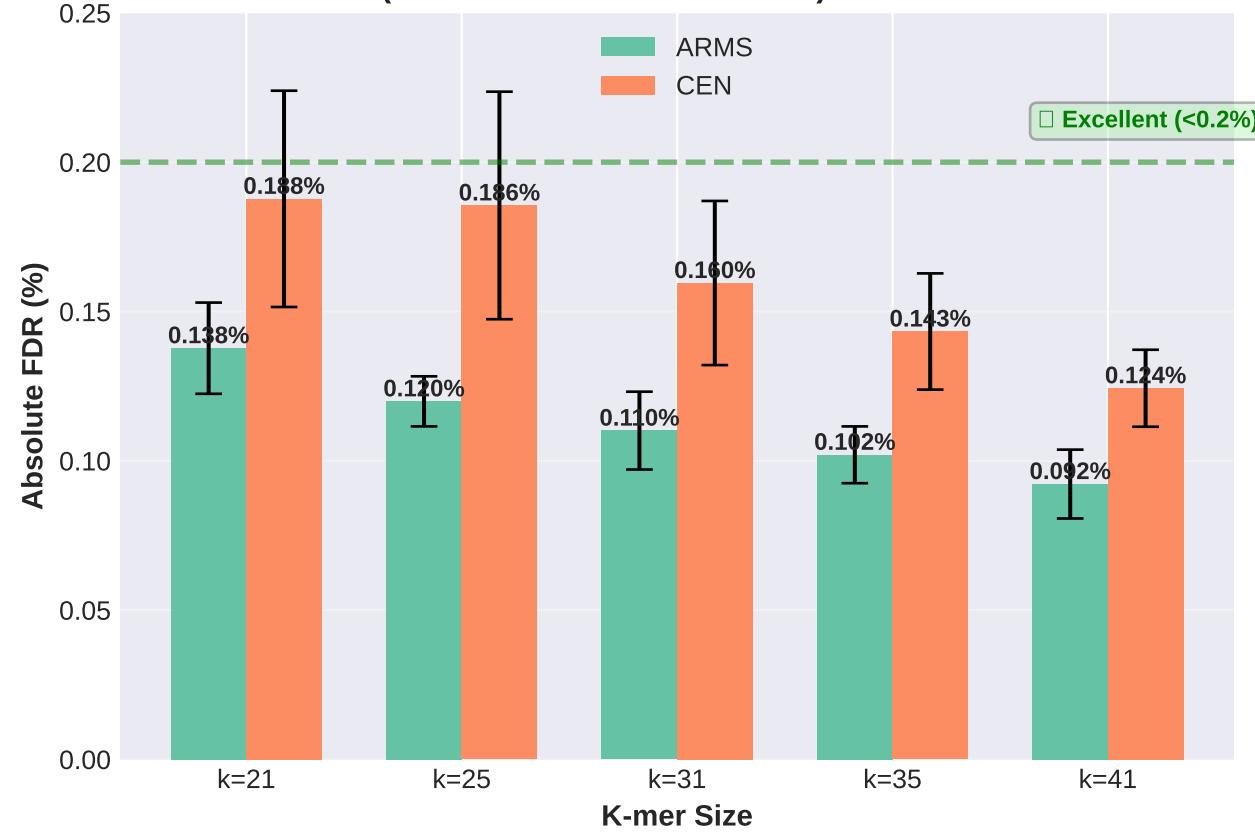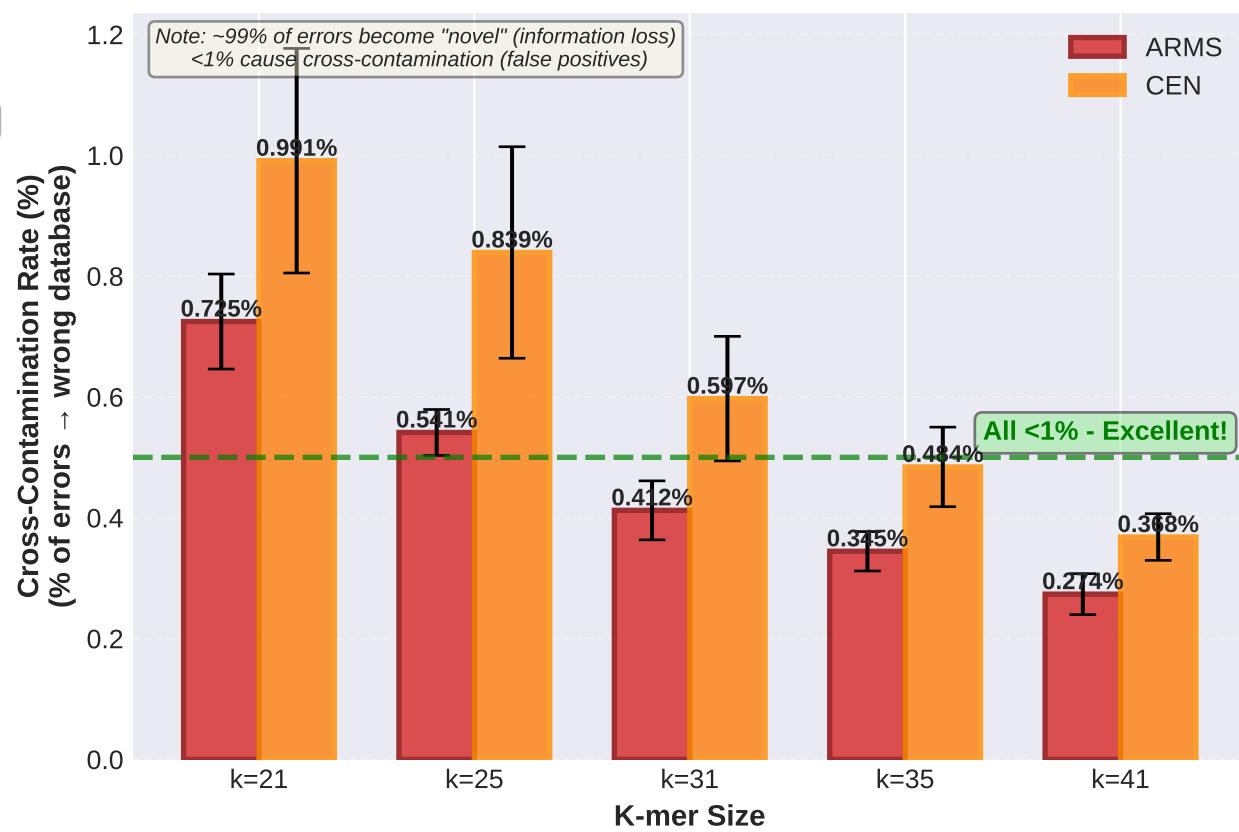**(1% per-base error rate, ONT-like)**
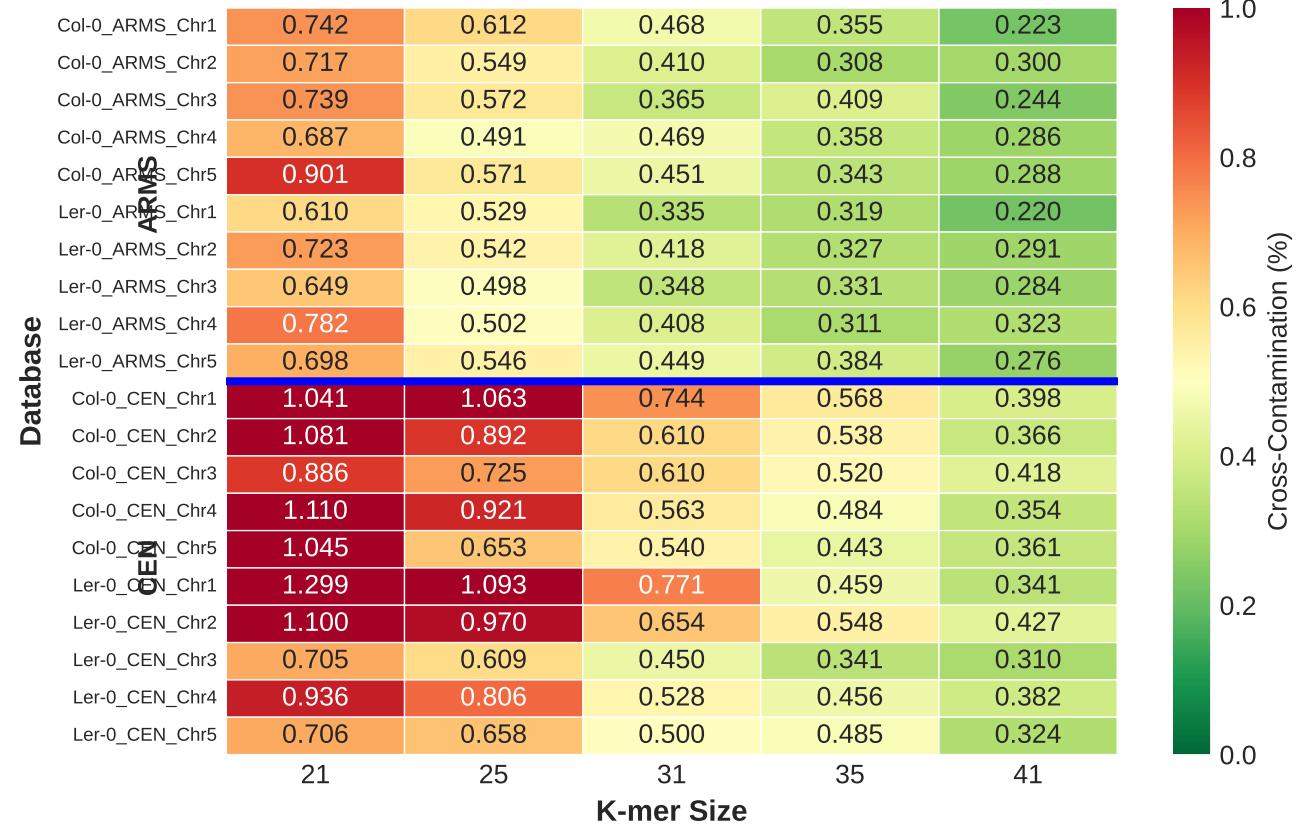
A. Absolute FDR (% of ALL k-mers that are FP)

ARMS / CEN

Excellent (<0.2%)

k=21: ARMS 0.138%, CEN 0.188%
k=25: ARMS 0.120%, CEN 0.186%
k=31: ARMS 0.110%, CEN 0.160%
k=35: ARMS 0.102%, CEN 0.143%
k=41: ARMS 0.092%, CEN 0.124%

Absolute FDR (%) vs K-mer Size

B. Cross-Contamination: FALSE POSITIVE Risk

Note: ~99% of errors become "novel" (information loss)
<1% cause cross-contamination (false positives)

ARMS / CEN

All <1% - Excellent!

k=21: ARMS 0.725%, CEN 0.991%
k=25: ARMS 0.541%, CEN 0.839%
k=31: ARMS 0.412%, CEN 0.597%
k=35: ARMS 0.345%, CEN 0.484%
k=41: ARMS 0.274%, CEN 0.368%

Cross-Contamination Rate (%) (% of errors → wrong database) vs K-mer Size

C. Per-Database Cross-Contamination (FP Risk)

Database	21	25	31	35	41
Col-0_ARMS_Chr1	0.742	0.612	0.468	0.355	0.223
Col-0_ARMS_Chr2	0.717	0.549	0.410	0.308	0.300
Col-0_ARMS_Chr3	0.739	0.572	0.365	0.409	0.244
Col-0_ARMS_Chr4	0.687	0.491	0.469	0.358	0.286
Col-0_ARMS_Chr5	0.901	0.571	0.451	0.343	0.288
Ler-0_ARMS_Chr1	0.610	0.529	0.335	0.319	0.220
Ler-0_ARMS_Chr2	0.723	0.542	0.418	0.327	0.291
Ler-0_ARMS_Chr3	0.649	0.498	0.348	0.331	0.284
Ler-0_ARMS_Chr4	0.782	0.502	0.408	0.311	0.323
Ler-0_ARMS_Chr5	0.698	0.546	0.449	0.384	0.276
Col-0_CEN_Chr1	1.041	1.063	0.744	0.568	0.398
Col-0_CEN_Chr2	1.081	0.892	0.610	0.538	0.366
Col-0_CEN_Chr3	0.886	0.725	0.610	0.520	0.418
Col-0_CEN_Chr4	1.110	0.921	0.563	0.484	0.354
Col-0_CEN_Chr5	1.045	0.653	0.540	0.443	0.361
Ler-0_CEN_Chr1	1.299	1.093	0.771	0.459	0.341
Ler-0_CEN_Chr2	1.100	0.970	0.654	0.548	0.427
Ler-0_CEN_Chr3	0.705	0.609	0.450	0.341	0.310
Ler-0_CEN_Chr4	0.936	0.806	0.528	0.456	0.382
Ler-0_CEN_Chr5	0.706	0.658	0.500	0.485	0.324

Cross-Contamination (%)

D. Error Fate: k=21 vs k=41 Comparison

k=41 CEN: 99.2% / 0.37
k=41 ARMS: 99.7% / 0.2
k=21 CEN: 98.4% / 0.99
k=21 ARMS: 99.2% / 0.72

Novel (lost) / Cross-contam (FP!)

% of K-mers WITH Errors