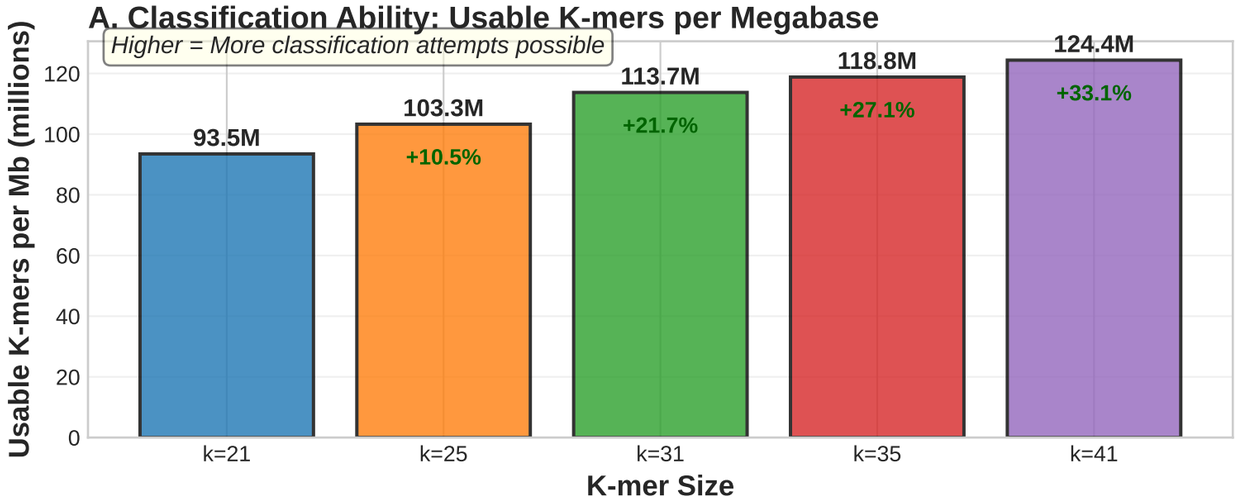
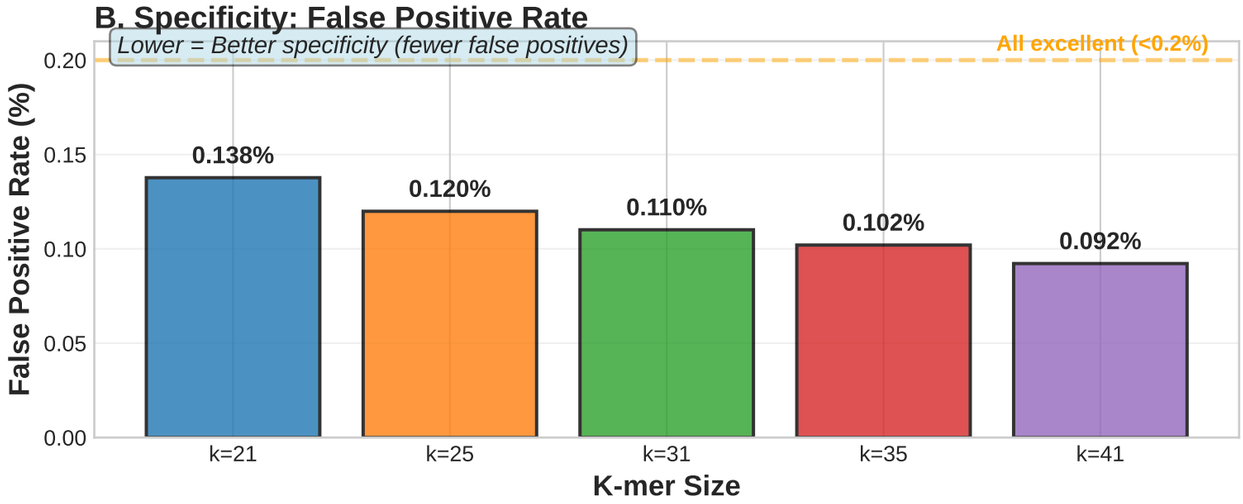
Objective K-mer Size Comparison: Understanding the Trade-offs
No arbitrary scoring - Make your own informed decision

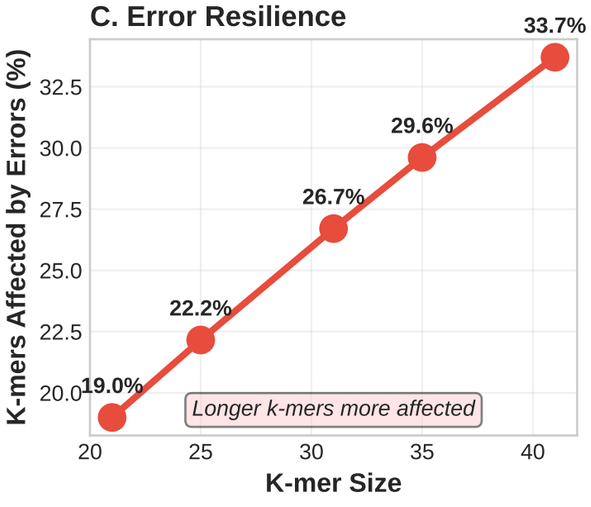
A. Classification Ability: Usable K-mers per Megabase



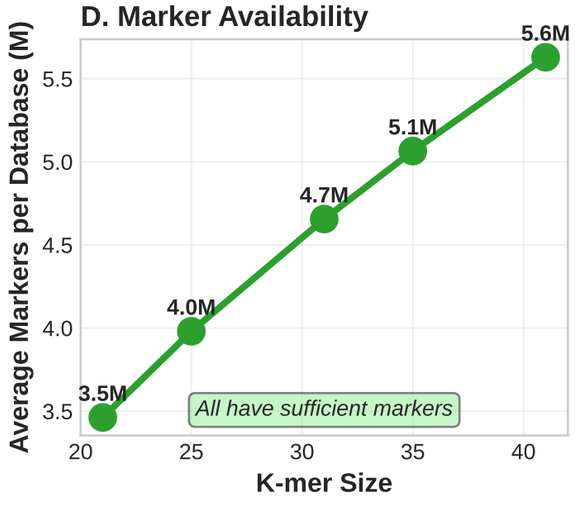
B. Specificity: False Positive Rate



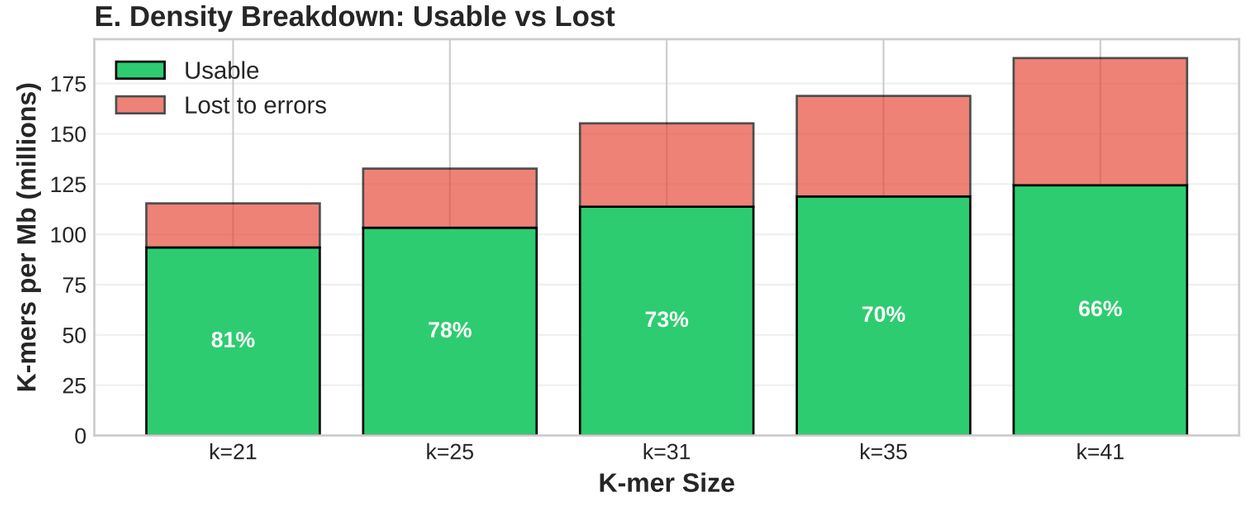
C. Error Resilience



D. Marker Availability



E. Density Breakdown: Usable vs Lost



F. Comprehensive Comparison: Understanding Your Options

K-mer	Usable K-mers/Mb	False Positive	Error Impact	Markers	Best For
	(Higher = Better)	*(Lower = Better)*	*(Lower = Better)*		
k=21	93.5M	0.138%	**19.0%**	3.5M	Very noisy data (>2% errors)
k=25	103.3M	0.120%	22.2%	4.0M	Balanced general use
k=31	113.7M	0.110%	26.7%	4.7M	Good balance + coverage
k=35	118.8M	0.102%	29.6%	5.1M	High coverage needs
k=41	**124.4M**	**0.092%**	33.7%	5.6M	Maximum resolution/specificity

DECISION GUIDE - No "one size fits all" answer!

Your choice depends on your priorities:

Prioritize ERROR RESILIENCE → k=21 or k=25
• Noisy ONT data (>2% per-base errors)
• Minimal false positives acceptable
• Can accept lower coverage

Prioritize BALANCE → k=31 □ GENERAL PURPOSE
• Standard ONT quality (~1% errors)
• Good coverage + reasonable specificity
• Most versatile choice

Prioritize MAXIMUM COVERAGE → k=35 or k=41
• High-quality ONT data (<1% errors)
• Need maximum resolution
• Best specificity possible
• Can tolerate more error impact