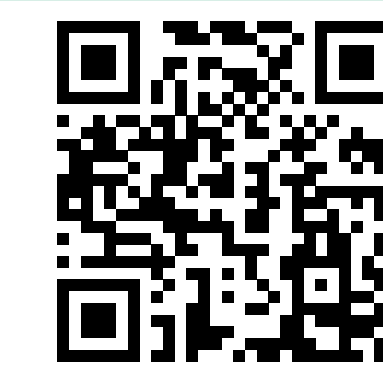


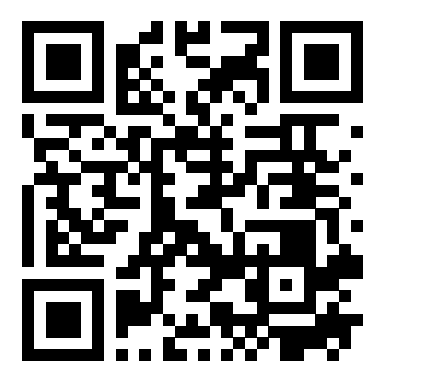
# Barbell Reveals and Resolves Demultiplexing and Trimming Issues in Nanopore Data

Rick Beeloo<sup>1</sup>, Ragnar Groot Koerkamp<sup>2</sup>, Xiu Jia<sup>3</sup>,  
Marian J. Broekhuizen-Stins<sup>1</sup>, Els M. Broens<sup>1</sup>,  
Aldert Zomer<sup>1</sup>, Bas E. Dutilh<sup>1,3</sup>

<sup>1</sup>Utrecht University | <sup>2</sup>Karlsruhe Institute of Technology | <sup>3</sup>Friedrich Schiller University Jena



github.com/rickbeeloo/  
barbell



Meet Rick

## Barcode Contamination in Public Database (Core NT)

- Rapid kits:** 103 matches of full tag (68 exact) across 67 assemblies of plasmids, mobile elements, and bacteria  
11x BC86 in *Photobacterium leiognathi* SV5.1
- Native kits:** 462 matches (270 exact) across 284 assemblies spanning viruses (SARS-CoV-2), bacteriophages, bacteria, parasites, fungi, and organelles.

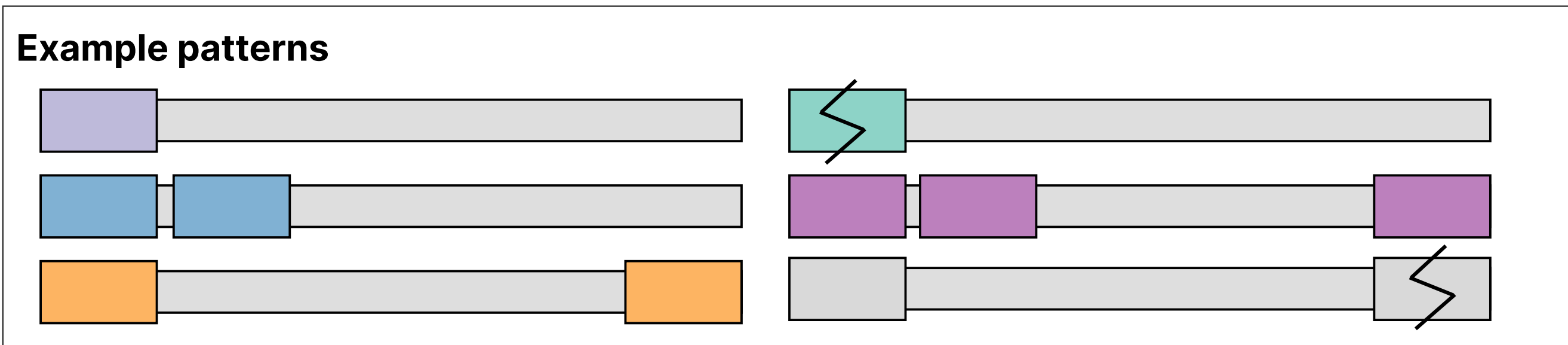
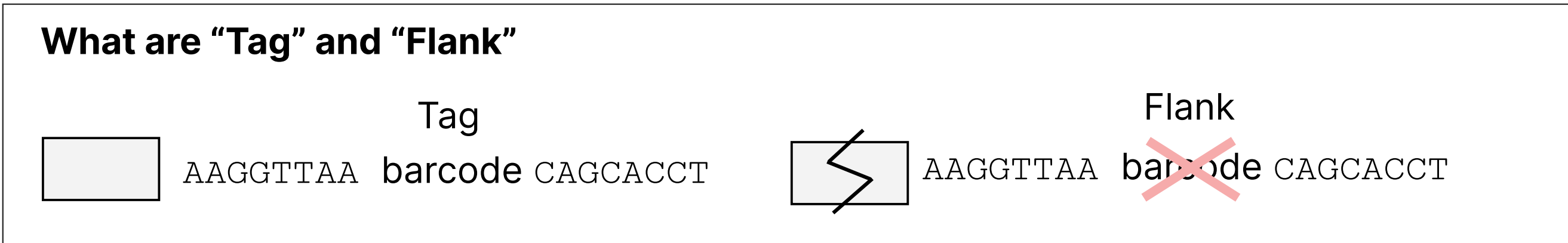
## Method: Approximate Matching Across All Reads

Barbell uses Sassy's approximate string matching to scan every read for a flank pattern:

AAGGTAA NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN CAGCACCT

Each flank match is tested against all barcodes in the kit. Barbell uses *overhang* alignment to detect partial tags at read ends.

Kit sequences	
Rapid kit	5' GCTTGGGTGTTAACC + barcode + GTTTTCGCATTTAT...CGTGCGCCGCTTCA 3'
Native kit	5' AAGGTAA + barcode + CAGCACCT 3'
	5' GGTGCTG + barcode + TTAACCTTAGCAAT 3'
	rc

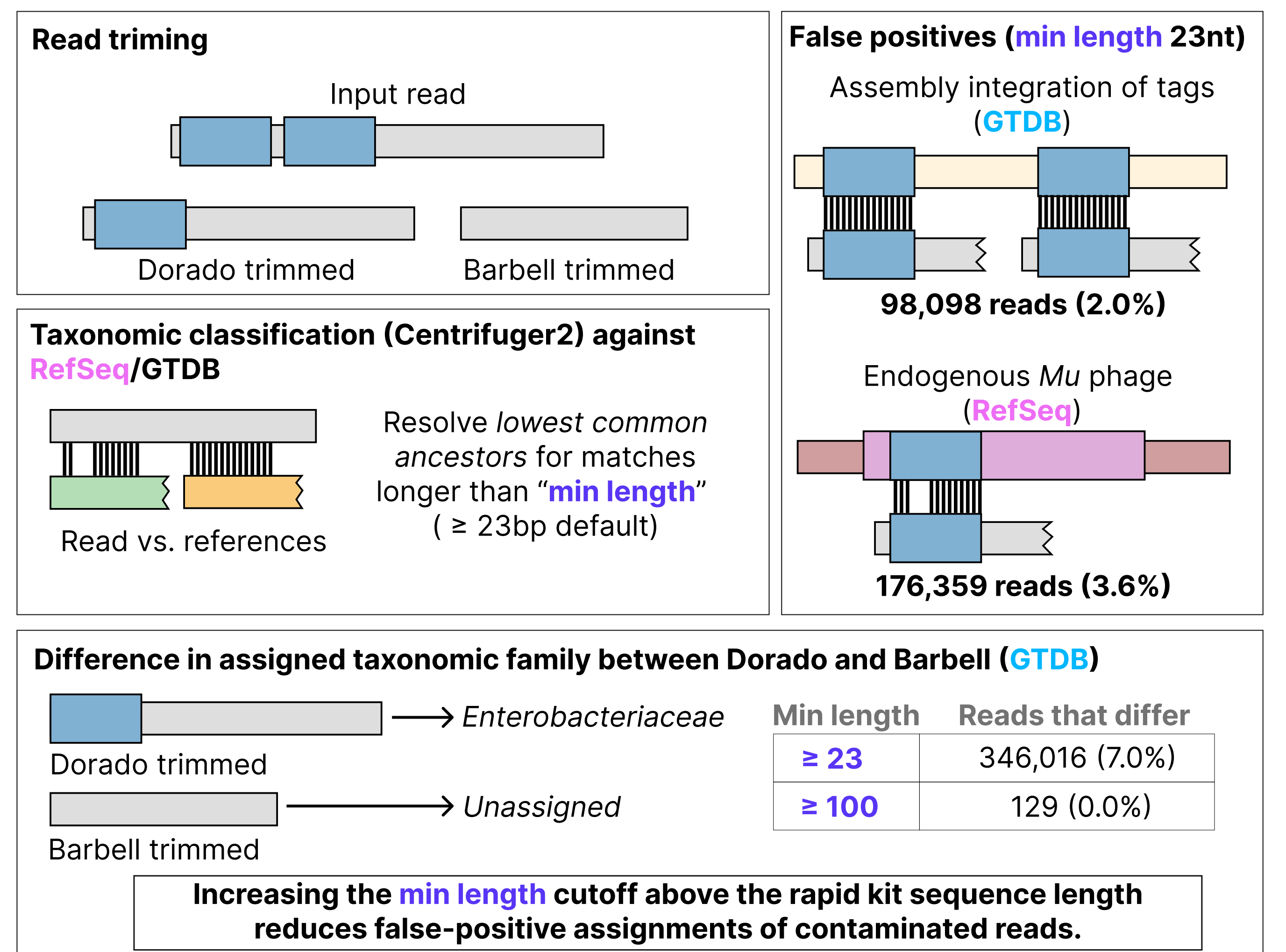


## Barbell Reduces Contamination

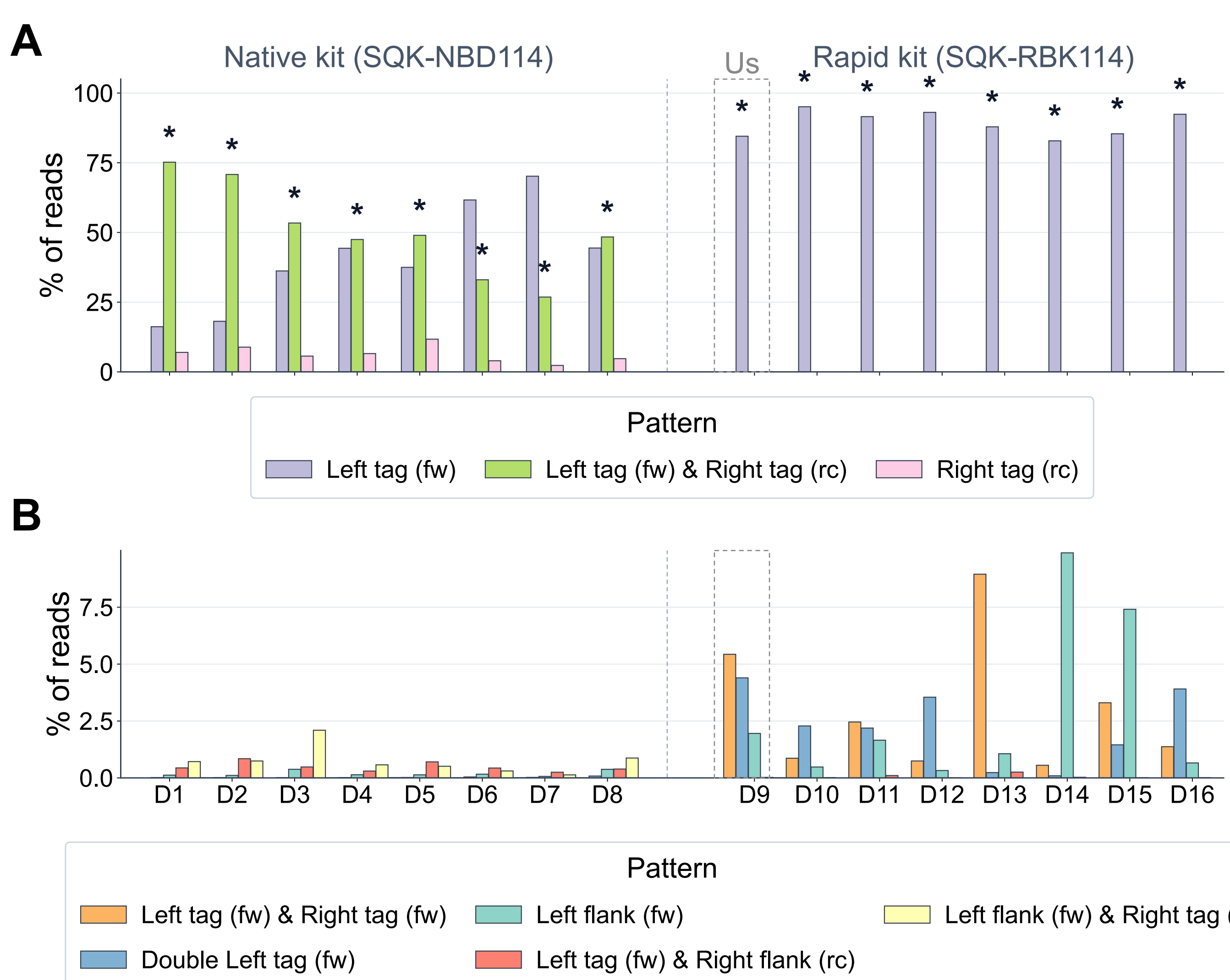
	Dorado	Flexiplex	Barbell
Reads assigned	94.1%	94.5%	85.6%
Contam. reads	464,518 (10.0%)	406,450 (8.7%)	166 (0.004%)
Runtime	6 min 50 s	1 min 2 s	4 min 20 s
Contam. assemblies	7/64	—	0/64

## Contamination Causes Taxonomic Misclassification

Barcode flanks left by Dorado create **artificial taxonomic signals**: rapid-barcode flanks share homology with *Mu*-transposon sequences found endogenously in *Enterobacteriaceae*.



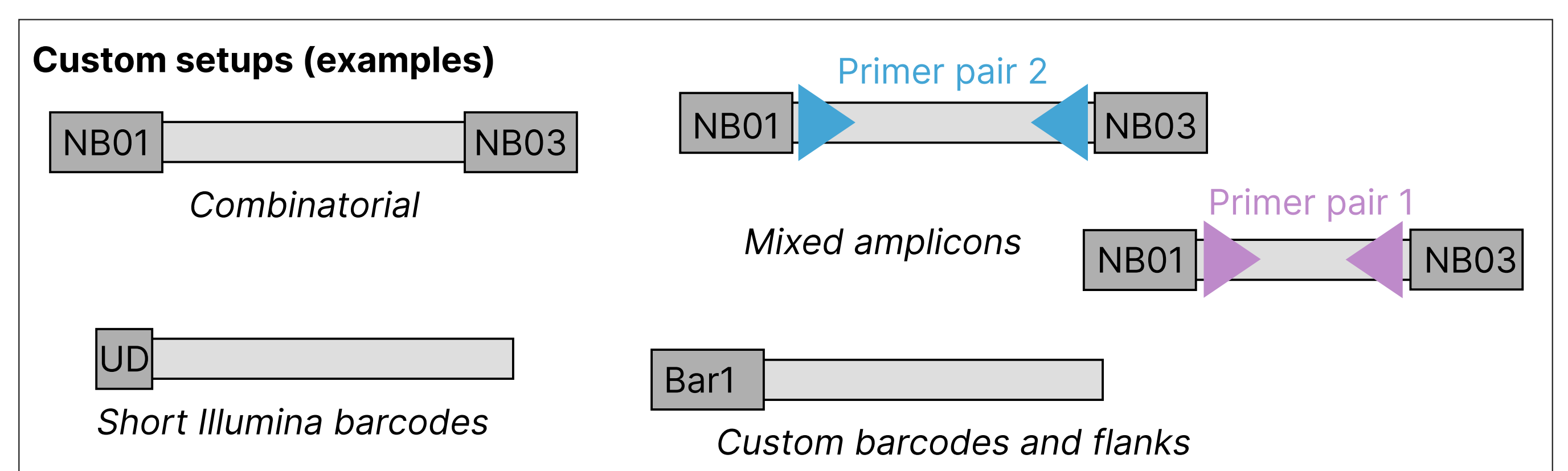
## Complex Barcode Patterns Across Datasets



Barcode attachment patterns in 8 rapid (SQK-RBK114) and 8 native (SQK-NBD114) SRA datasets. **(A)** Dominant (>10%) and **(B)** rare (<=10%) patterns. Native: the expected dual-end pattern accounts for only ~50% of reads. Rapid: the expected single left-tag exceeds 80%, yet complex multi-tag patterns are consistently present.

## Barbell Handles Custom Experimental Designs

Illumina barcodes, combinatorial barcodes, mixed primers: no problem!



**CLI:** barbell kit -k <kit> -i <reads> -o <out folder>

barbell kit infers flanks, sets cutoffs, trims reads, and reports the most frequent barcode-attachment patterns — one command, end to end.

1. Annotate	2. Filter	3. Trim
<p>Top 10 most common patterns</p> <p>Found 98 unique patterns</p> <p>Pattern 1: 46854 occurrences Ftag[fw, *, @left(0..250)]</p> <p>Pattern 2: 1913 occurrences Ftag[fw, *, @left(0..250)]_Ftag[fw, *, @left(0..250)]</p> <p>Pattern 3: 387 occurrences Ftag[fw, *, @left(0..250)]_Ftag[fw, *, @prev_left(0..250)]</p> <p>Pattern 4: 307 occurrences Ftag[fw, *, @left(0..250)]_Ftag[fw, *, @prev_left(0..250)]_Ftag[fw, *, @left(0..250)]</p> <p>Pattern 5: 156 occurrences Ftag[fw, *, @left(0..250)]_Ftag[fw, *, @prev_left(0..250)]_Ftag[fw, *, @left(0..250)]</p> <p>Pattern 6: 141 occurrences Ftag[fw, *, @left(0..250)]_Fflank[fw, *, @prev_left(0..250)]</p>	<p>Filtering reads...</p> <p>Total: 50225 reads 0s</p> <p>Kept: 49190 reads 0s</p> <p>Dropped: 1035 reads 0s</p>	<p>Trimming reads...</p> <p>Total: 52359 reads 1s</p> <p>Kept: 48676 reads 1s</p> <p>Kept split: 0 reads 1s</p> <p>Failed: 514 reads 1s</p>