Fragment Assembly Risk Scorer (FARS): Empirical Characterization of Split-Order Detection Boundaries for Benchtop DNA Synthesizers [1]

Hritika Chaturvedi 1
University of California,
Los Angeles

With
Apart Research

Abstract

Benchtop DNA synthesizers currently operate without mandatory sequence screening, enabling a dangerous attack vector: the split-order attack, in which a bad actor purchases multiple short DNA fragments that individually appear benign but collectively assemble into a sequence of concern. Existing tools, including SecureDNA and IBBIS commec, evaluate fragments one at a time. No deployed tool asks if these fragments, taken together, assemble into something dangerous.

I built FARS (Fragment Assembly Risk Scorer), a prototype on-device detector that scores orders by collective assembly potential — coverage, contiguity, and fragment count — rather than individual fragment identity. Tested against 960 simulated orders across three real 1918 H1N1 genomic segments from NCBI GenBank (AF117241, AF250356, AF116575), FARS achieves 100% detection of high- and medium-coverage split orders with zero false positives. On partial split orders, FARS detects 80% compared to 40% for an individual-sequence baseline modeled on IBBIS commec's methodology, doubling detection while eliminating false positives. At optimal threshold, FARS achieves 93.3% sensitivity, 100% specificity, and an F1 score of 0.966. The 20% that evade local detection precisely define the boundary where

[1] Research conducted at the AIxBio Hackathon, April 2026

shared cross-device infrastructure becomes necessary — directly quantifying the gap left by S.3741's mandated screening approach.

1. Introduction

Today, a bad actor targeting a benchtop DNA synthesizer faces one meaningful constraint: every fragment they order is screened individually. This paper shows that constraint is not enough — and quantifies exactly how much it falls short.

The split-order attack exploits a structural blind spot in current screening infrastructure. Multiple short DNA fragments are purchased separately, each appearing benign in isolation, then assembled off-device into a sequence of concern. Existing tools evaluate each sequence independently. None aggregate risk across an order history to detect assembly patterns. The IFP's report on Securing Benchtop DNA Synthesizers predicted this would matter; Edison, Toner & Esvelt (2026) confirmed the threat is real by demonstrating that unregulated fragments sufficient to reconstruct the 1918 influenza genome could be acquired from dozens of commercial providers with no individual purchase triggering an alert. Neither work answered the operational question: how much does assembly-awareness actually improve detection, and where does it fail?

This matters now because the Biosecurity Modernization and Innovation Act (S.3741, 2026) mandates homology-based screening but does not require assembly-aware detection and does not address benchtop devices. Kim (2026) identified this gap analytically. We quantify it empirically.

My main contributions are:

1. A working prototype of assembly-aware on-device split-order detection, evaluated against real publicly available NCBI GenBank reference sequences, with a deployable browser-based demo tool.
2. The first empirical head-to-head comparison of assembly-aware versus individual-sequence screening on real 1918 H1N1 genomic data, showing assembly-awareness doubles detection of realistic evasion attempts (80% vs. 40%) while eliminating false positives.
3. Precise quantification of the local detection boundary: orders below approximately 33% effective coverage with mutation rates above 11% consistently evade on-device screening — directly motivating shared cross-device infrastructure.

2. Related Work

(a) **SecureDNA** screens individual sequences using cryptographic DOPRF, detecting sequences as short as 30bp. Sherman et al. (NDSS 2026) conducted the first formal security analysis of SecureDNA, finding its mutual authentication protocol achieves only one-way authentication — motivating interest in on-device approaches not reliant solely on phone-home screening. SecureDNA does not aggregate risk across multiple orders from the same device.

(b) **IBBIS Common Mechanism (commec)** provides HMM-based biorisk screening with documented best performance above 150bp. Like SecureDNA, it evaluates each sequence independently without cross-order assembly analysis. Our comparison baseline models commec's key structural property — individual sequence evaluation with a 150bp performance threshold — using k-mer similarity rather than HMMs. This is not a direct benchmark of commec's actual detection rates, but a structural comparison isolating the contribution of assembly-awareness.

(c) **Edison, Toner & Esvelt (2026)** demonstrated empirically that the split-order threat is real, motivating fragment-level regulation. Our work provides the detection-side complement: how well can on-device screening catch such attacks, and where does it fail?

(d) **ABC-Bench (Kleinman et al., NeurIPS 2025)** evaluated frontier LLM capabilities on biosecurity tasks including screening evasion, finding leading models match or exceed expert performance. Our work is complementary: where ABC-Bench characterizes offensive AI capability, FARS characterizes defensive infrastructure gaps.

(e) When and why would someone use FARS over existing tools? FARS is designed for a gap existing tools do not address: aggregating assembly risk across multiple fragments in a single order. A practitioner deploying FARS gains something neither SecureDNA nor commec provides — the ability to flag an order where no single fragment triggers an alert but the combined set covers a dangerous sequence. The key insight FARS provides that did not exist before: a precise, empirically validated detection boundary for assembly-aware on-device screening, tested on real genomic reference data.

3. Methods

Scoring Logic

FARS computes three signals for each order simultaneously across all fragments:

- **Coverage:** Fraction of a reference sequence covered by the order's fragments, computed via sliding-window k-mer Jaccard similarity (k=8, step=3). A fragment counts as aligned if its best-window similarity exceeds 0.15.
- **Contiguity:** Whether aligned fragments form a connected assembly, measured as the largest contiguous covered block as a fraction of reference length. Gaps up to 50bp are tolerated to account for primer overhangs and cloning scars.
- **Fragment count penalty:** A small additional signal for unusually high fragment counts.

Risk score $= 0.50 \times$ coverage $+ 0.35 \times$ contiguity $+ 0.15 \times$ fragment penalty, producing a scalar from 0.0–1.0. Orders are classified as **CLEAR** (<0.35), **REVIEW** (0.35–0.60), or **ALERT** (≥ 0.60). FARS screens against all reference sequences and returns the highest score across all targets.

Comparison Baseline

To isolate assembly-awareness as a variable, we implemented an individual-sequence screening baseline modeled on IBBIS commec's published structural approach: each fragment is evaluated independently using sliding 150bp windows and k-mer similarity. An order is flagged only if any single fragment exceeds the similarity threshold. No cross-fragment aggregation. No assembly modeling. This captures the key structural property of current-generation screening tools.

Reference Sequences

Three publicly available 1918 H1N1 influenza segments were retrieved from NCBI GenBank via the BioPython Entrez API — standard biosecurity reference material, publicly available since 1999–2001:

Accession	Segment	Length	Strain
AF117241	Hemagglutinin (HA)	1,701 bp	A/South Carolina/1/18
AF250356	Neuraminidase (NA)	1,410 bp	A/Brevig Mission/1/18
AF116575	Nucleoprotein (NP)	1,220 bp	A/Brevig Mission/1/18

Table 1. Real NCBI GenBank reference sequences used for evaluation.

Dataset Construction

960 simulated orders were generated across four categories to span a realistic threat spectrum:

Category	N	Coverage	Mutation Rate	Represents
Benign	150	N/A	15–30%	Normal research purchases
Suspicious-High	270	80–95%	0–2%	Overt split-order attack
Suspicious-Medium	270	50–79%	4–8%	Moderate evasion attempt
Suspicious-Partial	270	25–49%	8–14%	Sophisticated evasion

Table 2. Dataset composition. Each suspicious tier was tested across all three genomic segments (90 orders per segment per tier).

Benign orders used fragments drawn from the same reference sequences with high mutation rates (15–30%) — ensuring they shared motifs with dangerous sequences but could not assemble, creating a realistic discrimination challenge rather than a trivial one.

Suspicious orders were generated by fragmenting each reference segment with random overlaps (10–30bp) and applying point mutations at the specified rates. Point mutation was chosen as the evasion model because it represents a well-documented, conservative baseline; more sophisticated evasion via codon optimization or AI-designed variants is acknowledged as a limitation.

One approach that was attempted and found insufficient: screening against the fictional 600bp reference sequence used in initial prototyping. Risk scores capped at 0.150 for all orders regardless of suspicion level, confirming that real genomic sequences are necessary for credible evaluation. Switching to real NCBI GenBank sequences immediately resolved the issue, with high-coverage suspicious orders scoring 0.946–0.997.

4. Results

Central Finding: Assembly-Awareness Doubles Detection on Realistic Evasion Attempts

The primary result is Table 3. On partial split orders — the realistic evasion scenario where each fragment is crafted to appear benign individually — FARS detects 80% compared to 40% for individual-sequence screening. That is 108 additional dangerous orders caught out of 270, with zero false positives versus a 2% false positive rate for the baseline.

Tier	N	FARS Flagged	IBBIS-Style	FARS Advantage
Benign	150	0 FP (0.0%)	3 FP (2.0%)	—
High (80–95%)	270	270 (100.0%)	270 (100.0%)	+0
Medium (50–79%)	270	270 (100.0%)	267 (98.9%)	+3
Partial (25–49%)	**270**	**216 (80.0%)**	**108 (40.0%)**	**+108 orders**
Total Suspicious	810	756 (93.3%)	645 (79.6%)	+111 orders

Table 3. Head-to-head comparison: FARS assembly-aware vs. IBBIS-style individual-sequence screening across 960 orders on real H1N1 genomic data. The partial tier (bold) is the primary finding.

The gap is largest exactly where it matters most. On high and medium coverage orders both approaches perform similarly. Assembly-awareness is critical precisely where individual fragments are designed to look benign — the partial split order scenario that real evasion attempts would most plausibly resemble.

Overall Performance

Threshold	Sensitivity	Specificity	Precision	F1 Score
0.25	98.9%	100.0%	100.0%	0.994
0.35 (optimal)	**93.3%**	**100.0%**	**100.0%**	**0.966**
0.60	68.5%	100.0%	100.0%	0.813

Table 4. FARS performance at three decision thresholds. Optimal F1 achieved at 0.35.

At the optimal threshold of 0.35, FARS achieves 93.3% sensitivity with 100% specificity and precision — zero false positives across all 150 benign orders on real genomic data.

REVIEW Queue Effectiveness

Of 270 partial split orders, 192 (71.1%) are escalated via REVIEW flags. The REVIEW queue achieves **100% precision** — every escalated order is genuinely suspicious, with zero false positives. A human analyst processing the REVIEW queue encounters no wasted effort on legitimate research orders. This graceful degradation — flagging what cannot be confirmed as ALERT for human review — is a deliberate architectural feature.

The Detection Boundary

All 54 completely missed orders came from the partial tier, with average coverage of 32.6% and average mutation rate of 11.4%. No high or medium coverage order was missed under any condition. Detection rates were identical across all three genomic segments (HA, NA, NP) despite differing lengths, suggesting the approach generalizes across sequence compositions.

5. Discussion and Limitations

Our results quantify something that was previously described only qualitatively. The IFP predicted local-only detection might be insufficient for split orders. Edison et al. showed the threat is real. We now know: local assembly-aware screening catches 93.3% of split-order attacks, compared to 79.6% for individual-sequence screening — a 13.7 percentage point gap at zero false-positive cost. The remaining 6.7% defines the boundary where shared infrastructure becomes necessary.

For S.3741 specifically: the bill mandates homology-based screening but not assembly-aware detection and does not address benchtop devices. Our results give policymakers a concrete empirical number for what that distinction costs. An amendment requiring assembly-aware screening methodology would improve split-order detection by 13.7 percentage points at the threshold that matters most.

The REVIEW escalation mechanism suggests a viable two-layer architecture: FARS as a fast on-device first-pass filter, with REVIEW-flagged orders escalated to a SecureDNA-style centralized cross-device database. Neither layer alone is sufficient. Together they address both the local detection gap and the shared infrastructure gap.

Limitations

Simulated dataset. All orders were computationally generated using point mutation models. Real evasion attempts might use codon optimization, synonymous substitutions, or AI-designed functional variants — as documented in Microsoft et al. (Science, 2025) — which could be more effective at evading k-mer screening while preserving biological function. Detection rates on real synthesis orders may differ.

Simplified IBBIS baseline. The comparison models commec's structural approach using k-mer similarity rather than HMMs. This is valid for isolating the contribution of assembly-awareness but should not be interpreted as a direct benchmark of commec's actual detection rates.

K-mer alignment limitations. Sliding-window k-mer Jaccard similarity is computationally efficient but less sensitive than Smith-Waterman alignment at high mutation rates. More robust alignment would improve partial order detection near the boundary.

Single-session assumption. FARS scores each order session independently. A sophisticated attacker could spread purchases across multiple sessions or devices, reducing per-session coverage below the local detection threshold — reinforcing the shared database motivation.

Reference database scope. FARS currently screens against three H1N1 segments. A production system requires a comprehensive sequence-of-concern database with implications for computational cost at scale.

Future Work

1. Temporal analysis: correlating fragment orders across multiple sessions from the same device
2. Validation against AI-designed synthetic variants as documented in Microsoft et al. (Science, 2025)
3. Replacement of k-mer alignment with seed-and-extend for higher mutation tolerance
4. Prototype shared cross-device database architecture motivated by the 33% coverage boundary finding
5. Collaboration with SecureDNA and IBBIS for validation against real screening databases

6. Conclusion

I set out to answer one question: how much does assembly-awareness improve split-order detection for benchtop DNA synthesizers, and where does it fail? The answer is now empirical. Assembly-aware scoring doubles detection of realistic evasion attempts compared to individual-sequence screening. It eliminates false positives. It fails precisely and predictably below 33% effective coverage at mutation rates above 11% — exactly the boundary where shared cross-device infrastructure must begin.

FARS is a prototype, not a production system. Its value is not only in what it catches but in what it measures: a precise, quantified threshold that gives biosecurity practitioners and policymakers concrete numbers for a tradeoff previously described only in qualitative terms.

Code and Data

- **Code repository**: https://github.com/hritikac25/ApartHack
- **Application:** https://hritikac25.github.io/ApartHack/demo.html

References

1. Edison, Toner & Esvelt (2026). "Assembling unregulated DNA segments bypasses synthesis screening: regulate fragments as select agents." Nature Communications.
2. IFP. Securing Benchtop DNA Synthesizers. Institute for Progress.
3. Kim, S. (2026). "AI Can Already Evade DNA Synthesis Screening. Congress's New Bill Doesn't Address That." March 2026.
4. Kleinman et al. (2025). ABC-Bench. NeurIPS 2025.
5. Microsoft et al. (2025). "Strengthening nucleic acid biosecurity screening against generative protein design tools." Science, October 2025.
6. OSTP (2024). Framework for Nucleic Acid Synthesis Screening. Effective April 29, 2025.
7. S.3741 (2026). Biosecurity Modernization and Innovation Act of 2026.
8. SecureDNA Documentation. https://securedna.org
9. Sherman et al. (2026). "Analysis of the Security Design, Engineering, and Implementation of the SecureDNA System." NDSS 2026.
10. IBBIS Common Mechanism (commec). https://ibbis.bio
11. Sentinel Bio (2026). "Why We're Doubling Down on Synthesis Screening." February 2026.

Appendix

Dual-use considerations. This paper characterizes detection capabilities and detection boundaries. The quantified boundary (33% coverage, 11% mutation rate) could theoretically inform evasion strategy. We note this information is implicit in the published literature on fragment-based attacks and does not provide meaningful additional uplift beyond what Edison et al. (2026) already established. FARS is a detector. The dataset generator produces fragments from publicly available sequences for research simulation only. No synthesis-ready information is produced.

Responsible disclosure. No vulnerabilities in existing screening systems were discovered. No real pathogen sequences were synthesized or used beyond their publicly available NCBI GenBank records. Future work extending FARS to comprehensive reference databases should be conducted in coordination with SecureDNA, IBBIS, and relevant government agencies.

Ethical considerations. All experiments used publicly available genomic sequences from NCBI GenBank. No novel dangerous sequences were generated. The research goal is entirely defensive: to characterize infrastructure gaps before benchtop synthesizers reach dangerous capability thresholds.

LLM Usage Statement

I used Claude to brainstorm approaches and help draft sections of code. All results and claims were independently verified.