

ASV Analysis Report

1. Executive Summary

This analysis processed 7,656 ASVs, identifying 1,954 (25.5%) as main ASVs and 2,120 (27.7%) as authenticated sequences. Key Findings: • Main/Candidate Classification achieved 100.0% F1 score • Authentication Analysis achieved 99.7% F1 score • Overall agreement with original labels: 97.8%

2. Main/Candidate Classification Analysis

2.1 Performance Metrics

Classification Performance: • Accuracy: 1.000 • Precision: 1.000 • Recall: 1.000 • F1 Score: 1.000 • ROC AUC: 1.000 Optimal Classification Threshold: 0.990

2.2 Feature Importance

Key Features: • log_read: 0.506 • rank_read: 0.490 • log_group_size: 0.004 • group_size_quantile: 0.000

3. Authentication Analysis

3.1 Authentication Performance

Overall Performance: • Accuracy: 0.998 • Precision: 1.000 • Recall: 0.993 • F1 Score: 0.997 • ROC AUC: 1.000 Authentication Thresholds: • Main ASVs: 0.374 • Candidate ASVs: 0.374

3.2 Results by ASV Type

Main ASVs: • Total: 1,954 • Authenticated: 1,954 • Authentication Rate: 25.5% • Agreement with Original Labels: 25.5%

Candidate ASVs: • Total: 5,702 • Authenticated: 166 • Authentication Rate: 2.2% • Agreement with Original Labels: 72.3%

4. Feature Analysis

4.1 Feature Statistics

Main Candidate Features: log_read: • Mean: 2.539 • Std: 1.923 • Min: 0.693 • Max: 7.577 rank_read: • Mean: 0.500 • Std: 0.287 • Min: 0.098 • Max: 1.000 log_group_size: • Mean: 3.821 • Std: 0.689 • Min: 2.047 • Max: 5.204 group_size_quantile: • Mean: 2.223 • Std: 0.951 • Min: 0.000 • Max: 3.000

Authentication Features: read_count: • Mean: 17.541 • Std: 22.208 • Min: 1.000 • Max: 59.500 total_read: • Mean: 170.915 • Std: 254.706 • Min: 4.000 • Max: 639.500 nearest_main_dist: • Mean: 0.114 • Std: 0.148 • Min: 0.000 • Max: 0.375 nearest_cand_dist: • Mean: 0.016 • Std: 0.021 • Min: 0.000 • Max: 0.058 dist_ratio: • Mean: 3.386 • Std: 4.440 • Min: 0.000 • Max: 12.609 log_group_size: • Mean: 3.821 • Std: 0.689 • Min: 2.047 • Max: 5.204

4.2 Feature Correlations

Strong Feature Correlations: • log_read - rank_read: 0.907 • log_read - read_count: 0.943 • rank_read - read_count: 0.851 • log_group_size - group_size_quantile: 0.888 • log_group_size - log_group_size: 1.000 • group_size_quantile - log_group_size: 0.888

5. Group Analysis

Group Analysis Summary: • Total Groups: 231 • Average ASVs per Group: 33.1 • Average Main ASV Percentage: 30.1% • Average Authentication Rate: 32.4% Group Size Distribution: • Minimum: 1 • Maximum: 181 • Median: 27