International Alliance for Biological Standardization

4th Conference on Next Generation Sequencing for Adventitious Virus Detection in Biologics for Humans and Animal

Frankfurt, Germany
December 3-5, 2024

Rapid alignment-free detection of adventitious agents using next-generation sequencing

Authors:
- **Tom J.B. de Man**
- **Brendan Redler**
- **Amber Overgard**
- **Afshin Sohrabi**

Merck KGaA

Next-Generation Sequencing (NGS) has revolutionized adventitious agent testing (AAT), becoming integral to the ICH Q5A(R2) guidelines due to its sensitivity and breadth of sample analysis. NGS technology facilitates detection of both known and novel adventitious agents that contaminate biopharmaceutical products, offering comprehensive understanding of sample composition. As reference genome databases and whole-genome sequencing output continue to grow, the demand for scalable and accurate computational methods increases, with traditional full-read local alignment approaches lagging in efficiency.

Our newly designed workflow employs a k-mer based analysis algorithm that addresses these issues as itis more application-appropriate for AAT metagenomic datasets. The workflow enables the rapid, accurate detection of viruses, bacteria, and fungi, identifying k-mers unique to respective taxa. By comparing test sample k-mers against a streamlined query database, consisting of diagnostic k-mers derived from public nucleic acid sequences, the method significantly cuts down search space and computation time. The workflow can screen out cell-line and other manufacturing process data to improve the reliability of results.

This algorithm is not only suitable for highly accurate, short Illumina sequencing reads (error rates <0.5%, lengths 100–250 bp) but also compatible with long-read data from platforms like Pacific Biosciences and Oxford Nanopore.

Benchmarking indicates that the pipeline can reduce false positives compared to existing AAT analysis methods and displays a detection limit of 1x10^{4} viral genome copies per mL across various backgrounds. Consistent output across multiple analysis runs affirm its GMP compliance as an AAT bioinformatics pipeline. These findings demonstrate performance on par with current bioinformatics methodologies, while offering enhanced efficiency in the processing of NGS AAT data.

This algorithm significantly improves the scalability and reliability of AAT assays and ensures the abilityto adapt to the next era of genomic data complexity.