Standard Operating Procedure: MPXV VSP2 Panel Analysis using Next-Generation Sequencing

SOP Number: SOP_EACML_MPXV_VSP2_NGS_XXX_V1
Effective Date: October 2024
Title: MPXV VSP2 Panel Analysis using Next-Generation Sequencing
Author: EAC-RNPHRL Bioinformatics Team
Review/Approval Signatures:

Name / Title	Signature	Date
Laboratory Director		
Quality Assurance Manager		

Table of Contents

1. Introduction

1.1 Purpose and Scope

This document establishes the standard operating procedures (SOPs) for Monkeypox Virus (MPXV) VSP2 panel analysis utilizing Next-Generation Sequencing (NGS) within the East African Community (EAC). It provides a comprehensive, step-by-step guide encompassing the entire workflow, from initial sample receipt and nucleic acid extraction through targeted library preparation, sequencing on Illumina platforms, and a robust bioinformatics pipeline for variant calling, analysis, and final reporting. The scope of this SOP is specifically tailored for the MPXV VSP2 panel, with the aim of ensuring accurate and reproducible detection of viral genetic variants that are critical for public health surveillance.

The explicit branding of the provided research material for the "East African Community," "EAC Mobile Lab," and "EAC-RNPHRL" [1], along with its stated purpose of defining "standard operating procedures" [1], underscores a fundamental objective: to achieve uniformity and comparability of results across multiple laboratories within the EAC network. This MPXV SOP, therefore, serves a broader function beyond merely outlining a technical procedure for a single laboratory. It is a foundational document designed to foster a harmonized and interconnected public health surveillance system across the EAC. By standardizing the entire process, from sample handling to data interpretation, the EAC can enhance its collective capacity for rapid, reliable, and comparable responses to MPXV outbreaks, thereby facilitating effective regional disease control and seamless data sharing.

1.2 Applicability

This SOP is intended for use by qualified laboratory personnel and bioinformaticians operating within the EAC Mobile Lab, the EAC Regional Network of Public Health Reference Laboratories (EAC-RNPHRL) for Communicable Diseases, and other collaborating public health institutions throughout the East African Community engaged in MPXV surveillance and diagnostics. The procedures outlined herein are specifically designed to be compatible with existing Illumina sequencing platforms, namely the iSeq100 and NextSeq1000/2000, which are already in use within the EAC network.[1] Furthermore, the bioinformatics workflow leverages the established Nextflow-driven environment [1], maximizing the utility of current infrastructure and expertise available across the region. This approach ensures that the implementation of MPXV VSP2 panel analysis can seamlessly integrate with existing laboratory capabilities and build upon established bioinformatics practices.

1.3 Principles of MPXV VSP2 Panel Analysis

Monkeypox Virus (MPXV) is a double-stranded DNA virus responsible for Mpox disease. Accurate detection and precise genetic characterization of this virus are paramount for effective public health surveillance, thorough outbreak investigation, and a comprehensive understanding of viral evolution. The VSP2 panel specifically targets highly informative regions of the MPXV genome. These regions are crucial for rapid and accurate lineage assignment, effective tracking of emerging viral variants, and

detailed epidemiological investigations.

Next-Generation Sequencing (NGS) offers high-resolution genetic data, which enables comprehensive variant detection, including the identification of low-frequency variants. This capability is particularly vital for deciphering viral transmission dynamics and monitoring the potential emergence of new clades or strains. The Illumina platforms, which are the primary sequencing instruments for this workflow, are well-suited for short-read sequencing applications [1], providing the necessary data for this high-resolution analysis. The entire analysis workflow is orchestrated using a Nextflow DSL2 pipeline.[1] This automation facilitates the processing of raw sequence reads (fastq files) [1] into actionable reports, ensuring reproducibility, enhancing efficiency, and minimizing the potential for manual errors. This automated framework directly aligns with the robust, standardized approach established in the provided antimicrobial resistance (AMR) SOP.

2. Definitions and Abbreviations

This section provides a comprehensive, alphabetized list of all technical terms and their definitions, along with all abbreviations utilized throughout this document, to ensure clarity and consistency in terminology.

Abbreviation	Definition
AMR	Antimicrobial Resistance
ARGs	Antibiotic Resistance Genes
AST	Antimicrobial Susceptibility Testing
BNITM	Bernhard Nocht Institute for Tropical Medicine
CARD	The Comprehensive Antibiotic Resistance Database
cDNA	Complementary Deoxyribonucleic Acid
DNA	Deoxyribonucleic Acid
EAC	East African Community
EAC-RNPHRL	EAC Regional Network of Public Health Reference Laboratories
FASTQ	Text-based format for storing both a nucleotide sequence and its corresponding quality scores

INDEL	Insertion or Deletion (of nucleotides)
LIMS	Laboratory Information Management System
MLST	Multi-Locus Sequence Typing
MPXV	Monkeypox Virus
NCBI	National Center for Biotechnology Information
NGS	Next-Generation Sequencing
PCR	Polymerase Chain Reaction
PPE	Personal Protective Equipment
QC	Quality Control
RNA	Ribonucleic Acid
SNP	Single Nucleotide Polymorphism
SOP	Standard Operating Procedure
VAF	Variant Allele Frequency
VCF	Variant Call Format
VSP2	Variola Virus Smallpox Vaccine Strain 2 (a specific gene region in orthopoxviruses)
VTM	Viral Transport Media

3. Safety Precautions

Adherence to stringent safety protocols is paramount when performing MPXV VSP2 panel analysis, given the infectious nature of the virus and the use of various chemicals and equipment.

3.1 General Laboratory Safety

All personnel must adhere to universal precautions and standard laboratory safety guidelines for handling potentially infectious biological samples. MPXV is typically handled under Biosafety Level 2 (BSL-2) containment for diagnostic work, but specific procedures involving viral culture or high concentrations may necessitate Biosafety Level 3 (BSL-3) containment, depending on the specific risk assessment conducted by the facility's biosafety officer. All work surfaces should be decontaminated before and after use with appropriate disinfectants.

3.2 Personal Protective Equipment (PPE)

Mandatory use of appropriate PPE is required at all times when handling samples or reagents. This includes, but is not limited to, laboratory coats, disposable gloves (nitrile or latex), and eye protection (safety glasses or face shield). During activities that may generate aerosols, such as vortexing, pipetting, or opening tubes, the use of N95 respirators or higher-level respiratory protection is strongly recommended. Gloves should be changed frequently, especially after contact with potentially contaminated surfaces or before handling clean reagents.

3.3 Chemical Safety

Proper handling, storage, and disposal of all reagents and chemicals used in nucleic acid extraction and library preparation are critical. Personnel must be familiar with the Material Safety Data Sheets (MSDS) for all chemicals used and follow recommended safety measures, including working in a fume hood when volatile or hazardous chemicals are present. Chemical spills must be cleaned up immediately according to established laboratory protocols.

3.4 Waste Management

Procedures for the proper decontamination and disposal of all biohazardous waste (e.g., contaminated plastics, sharps, liquid waste) and chemical waste must be strictly followed. Biohazardous waste should be autoclaved or chemically disinfected before disposal, in accordance with local and national regulations. Sharps containers must be used for all needles, blades, and broken glass. Chemical waste must be segregated and disposed of according to hazardous waste guidelines.

3.5 Equipment Safety

Safe operation of all laboratory equipment, including centrifuges, thermocyclers, automated liquid handlers, and sequencers, is essential. Operators must be trained on the correct use of each instrument. Regular maintenance and calibration checks, as specified by manufacturer guidelines, must be performed to ensure equipment functions correctly and safely. Emergency stop procedures for all major equipment should be clearly understood by all users.

3.6 Bioinformatics Workstation Safety

While not involving biological hazards, bioinformatics workstations require attention to general safety. This includes ensuring proper electrical safety (e.g., avoiding overloaded power strips, proper grounding) and maintaining an ergonomic setup to prevent repetitive strain injuries. Data security protocols, including strong passwords, regular software updates, and restricted access, must be implemented to protect sensitive patient data and intellectual property.

4. Equipment, Reagents, and Consumables

Successful MPXV VSP2 panel analysis relies on the availability and proper functioning of specialized laboratory equipment, high-quality reagents, and a robust bioinformatics infrastructure.

4.1 Laboratory Equipment

- **Standard Molecular Biology Equipment:** A refrigerated centrifuge is essential for nucleic acid purification and library cleanups, ensuring sample integrity. A vortex mixer facilitates thorough reagent mixing. Thermocyclers are indispensable for PCR amplification steps during VSP2 panel preparation. For initial DNA/RNA quality checks, a gel electrophoresis system can be utilized. A spectrophotometer (e.g., NanoDrop) is used for basic nucleic acid purity assessment, while a fluorometer (e.g., Qubit) provides highly accurate and sensitive quantification of nucleic acids, especially critical for low-concentration viral samples.
- **Library Quality Control Equipment:** An automated electrophoresis system, such as the Agilent Bioanalyzer or TapeStation, is crucial for assessing library fragment size distribution and confirming the absence of adapter dimers. This ensures that the prepared libraries meet the desired quality standards for Illumina sequencing.[1]
- **Next-Generation Sequencing Platforms:** The primary sequencing platforms for this SOP are the Illumina iSeq100 and/or NextSeq1000/2000.[1] These instruments are specified for generating the short-read sequencing data required for MPXV VSP2 analysis within the EAC network, leveraging existing investments and expertise.

4.2 Bioinformatics Hardware

- **Workstation/Server Specifications:** A dedicated bioinformatics workstation or server running a Linux operating system, such as Linux Mint 2.11 or Ubuntu 22.04, is required.[1] This system must possess a minimum of 32 GB RAM and at least 1 TB of storage for software installations and temporary files.[1] For practical long-term use, accommodating raw data, intermediate analysis files, and comprehensive databases, a minimum of 5-10 TB of high-speed storage is highly recommended. This expanded storage capacity is vital for managing the large datasets generated by NGS and for hosting the necessary viral reference genomes and variant databases.
- **Network Connectivity:** A stable and high-speed internet connection is indispensable for

downloading bioinformatics software, updating databases, and for the secure submission of final sequence data to public repositories.

4.3 Reagents and Kits

- **Nucleic Acid Extraction Kits:** Commercial kits validated for the extraction of high-quality viral DNA from diverse clinical sample types are necessary. Examples include the QIAamp DNA Blood Mini Kit or the MagMAX DNA Multi-Sample Ultra Kit, which are designed to efficiently purify viral nucleic acids while minimizing inhibitors.
- **MPXV VSP2 Panel Library Preparation Kits:** Specific commercial or in-house kits designed for targeted sequencing of MPXV VSP2 regions are required. These kits typically employ amplicon-based enrichment strategies (e.g., adapted ARTIC Network protocols) or probe-based capture methods. Key components will include VSP2-specific primers, various enzymes (e.g., DNA polymerase, ligase), reaction buffers, and unique dual indexing (UDI) adapters to enable multiplexing of multiple samples in a single sequencing run.
- **NGS Consumables:** This category includes Illumina flow cells compatible with the chosen sequencer, specific Illumina sequencing reagent kits, magnetic bead-based cleanup reagents (e.g., AMPure XP beads) for purification steps, molecular-grade water, PCR tubes/plates, and other general laboratory consumables.

4.4 Software and Databases

The bioinformatics analysis of MPXV VSP2 panel data requires a specific suite of software tools and databases. The core environment is built upon Nextflow and Conda/mamba, which are essential for managing complex workflows and software dependencies.

- **Core Bioinformatics Environment:**
  - **Nextflow:** This is an essential workflow management system that orchestrates complex bioinformatics pipelines.[1] It automates the execution of analysis steps, ensuring reproducibility and scalability across different computing environments.
  - **Conda/mamba:** These are powerful package and environment managers [1] that are critical for installing and managing bioinformatics tools and their dependencies in isolated environments. This approach prevents software conflicts and ensures consistent tool versions, which is vital for reproducible research and diagnostic pipelines.
- **Quality Control Tools:**
  - **FastQC:** A widely used tool for comprehensive raw read quality assessment.[1] It generates detailed reports on crucial metrics such as per-base sequence quality, adapter content, GC content, and sequence duplication levels, providing an initial overview of data quality.
  - **fastp or Trimmomatic:** Software designed for efficient adapter trimming and quality filtering of raw sequencing reads.[1] These tools remove low-quality bases and contaminating adapter sequences that can interfere with accurate downstream alignment and variant calling.

- **Alignment and Variant Calling Tools (MPXV Specific):**
    - **BWA (Burrows-Wheeler Aligner) or minimap2:** These are industry-standard tools for rapidly and accurately aligning trimmed short reads to a reference genome, such as the MPXV reference.
    - **Samtools/Bcftools:** A versatile suite of utilities for manipulating alignment files (BAM/SAM format) and performing essential tasks such as sorting, indexing, basic variant calling, filtering, and format conversion.
    - **iVar (Integrated Variant Analysis for Viral Genomics):** Highly recommended for viral amplicon sequencing data, iVar enables precise primer trimming, robust consensus sequence generation, and accurate variant calling specifically optimized for diverse viral populations.
    - **GATK HaplotypeCaller or FreeBayes:** General-purpose variant callers that can also be applied to viral data, though iVar is often preferred for targeted amplicon sequencing due to its viral-specific optimizations.
- **Variant Annotation and Interpretation Tools:**
    - **SnpEff or VEP (Variant Effect Predictor):** Tools used for annotating identified genetic variants with their predicted functional impact (e.g., synonymous, missense, frameshift, intergenic) based on established gene models. This helps in understanding the potential biological consequences of detected mutations.
    - **Custom Scripts/Databases for Lineage Assignment:** Tools or scripts are necessary for assigning MPXV lineages based on detected variants. This may involve adapting existing tools (e.g., similar to Pangolin or Nextclade for SARS-CoV-2) or developing custom solutions specifically validated for MPXV lineages. This capability is crucial for epidemiological tracking and outbreak response.
- **Contamination Check Tools:**
    - **Kraken2 and Taxonkit:** While primarily used in the AMR SOP for bacterial metagenomics [1], these tools can be effectively adapted to identify and quantify non-MPXV sequences, such as human host contamination or other microbial contaminants. This ensures the purity of the MPXV signal and prevents misinterpretation of results.
- **MPXV Specific Databases:**
    - **MPXV Reference Genome:** A high-quality, complete MPXV reference genome sequence is fundamental. This should be a specific accession number from NCBI GenBank (e.g., ON563414.1 for a representative Clade IIb strain), replacing the general "Reference genomes (RefSeq)" concept from the bacterial SOP [1] with a precise viral reference.
    - **VSP2 Panel Target Coordinates:** A BED file or similar genomic interval file is required to precisely define the genomic coordinates of each target region within the VSP2 panel on the chosen MPXV reference genome. This file is critical for focused analysis and accurate coverage assessment.
    - **Known MPXV Variant Database (Optional but Recommended):** A curated database of known MPXV variants, particularly those within the VSP2 regions, along with their associated lineages or known epidemiological significance, greatly aids in variant interpretation and lineage assignment.

Table 4.4.1: Required Software and Databases for MPXV VSP2 Analysis

This table provides a quick, centralized, and comprehensive reference for all necessary bioinformatics tools and databases required for MPXV VSP2 analysis. It clearly outlines their primary function and specific relevance to the MPXV workflow, which is critical for laboratory personnel and bioinformaticians to ensure proper environment setup, troubleshoot issues, and verify that all essential components are in place before initiating any analysis. The table addresses the fundamental difference in tool requirements between bacterial AMR analysis and targeted viral sequencing, ensuring that the correct and most effective resources are deployed for MPXV.

Category	Software/Database Name	Primary Function	Relevance to MPXV VSP2 Analysis
Workflow Orchestration	Nextflow	Automates bioinformatics workflows, ensures reproducibility and scalability.	Essential for managing the complex, multi-step MPXV analysis pipeline.
Package Management	Conda/mamba	Installs and manages bioinformatics tools and their dependencies in isolated environments.	Prevents software conflicts and ensures consistent tool versions for reliable analysis.
Read Quality Control	FastQC	Assesses raw read quality, identifies potential issues (e.g., adapter contamination, low quality bases).	Provides initial quality assessment of raw MPXV sequencing data.
Read Trimming/Filtering	fastp or Trimmomatic	Removes low-quality bases and adapter sequences from raw reads.	Improves accuracy of downstream alignment and variant calling by cleaning MPXV reads.
Reference Alignment	BWA or minimap2	Aligns trimmed sequencing reads to a reference genome.	Maps MPXV reads to the reference genome, forming the

			basis for variant detection.
Viral Variant Calling	iVar	Performs primer trimming, generates consensus sequences, and calls variants specifically for viral amplicon data.	Optimized for targeted viral sequencing, crucial for accurate MPXV variant detection within VSP2 panel.
General Variant Calling	Samtools/Bcftools	Manipulates alignment files (BAM/SAM) and performs basic variant calling/filtering.	Supports iVar, processes alignment files, and can perform complementary variant calling.
Variant Annotation	SnpEff or VEP	Annotates identified genetic variants with predicted functional impact.	Provides biological context for MPXV variants, indicating potential effects on viral proteins.
Contamination Check	Kraken2 and Taxonkit	Identifies and quantifies non-target DNA sequences (e.g., host, other microbes).	Ensures purity of MPXV signal, identifies potential cross-contamination.
Databases	MPXV Reference Genome (e.g., ON563414.1)	Provides the genomic backbone for read alignment and variant calling.	Fundamental for accurate alignment and identification of MPXV-specific variants.
Databases	VSP2 Panel Target Coordinates (BED file)	Defines the precise genomic regions targeted by the VSP2 panel.	Focuses analysis on regions of interest, critical for specific panel coverage assessment.
Databases	Known MPXV Variant Database (curated)	Contains information on previously identified MPXV	Aids in interpretation of detected variants and MPXV lineage

		variants and lineages.	assignment.

5. Detailed Procedures

This section outlines the step-by-step procedures for MPXV VSP2 panel analysis, from sample collection to final data interpretation and reporting.

5.1 Sample Collection and Handling

Proper sample collection and handling are foundational to obtaining reliable sequencing results.

5.1.1 Sample Types

This SOP primarily focuses on clinical samples suspected of MPXV infection. These include, but are not limited to, lesion swabs (from vesicles, pustules, or crusts), lesion fluid, crusts, and potentially blood or plasma, depending on the stage of infection and clinical presentation. The choice of sample type should align with national diagnostic guidelines for Mpox.

5.1.2 Collection Protocol

Adherence to standardized, sterile procedures for sample collection is critical to minimize contamination and preserve nucleic acid integrity. Samples should be collected using appropriate sterile swabs (e.g., flocked swabs) and immediately placed into viral transport media (VTM) or dry sterile tubes, as specified by national guidelines and the requirements of the downstream nucleic acid extraction method.

5.1.3 Storage and Transport

Following collection, samples must be immediately placed on ice or refrigerated (4°C) to prevent nucleic acid degradation. For short-term storage, samples can be kept at 4°C for up to 72 hours. For long-term preservation, samples must be frozen at -20°C or, preferably, -80°C. Transport to the laboratory must strictly maintain the cold chain and adhere to all national and international biosafety regulations for infectious substances, typically requiring triple packaging and appropriate labeling.

5.1.4 Sample Accessioning

Upon receipt at the laboratory, each sample must be meticulously logged into the laboratory information management system (LIMS) or a dedicated logbook. Essential information to record includes a unique sample identifier, the date and time of collection, the date and time of receipt, de-identified patient demographics, the clinical diagnosis, the sample type, and any observed issues regarding sample condition (e.g., leakage, incorrect volume, degradation). This meticulous record-keeping ensures traceability and data integrity throughout the analysis pipeline.

5.2 Nucleic Acid Extraction (DNA/RNA Extraction Protocol for MPXV)

The objective of this stage is to extract high-quality, intact viral DNA from clinical samples, free from inhibitors that could interfere with downstream enzymatic reactions such as PCR amplification and

sequencing. Since MPXV is a double-stranded DNA virus, DNA extraction is the primary focus.

5.2.1 Principle

Nucleic acid extraction methods typically involve cell lysis, inactivation of nucleases, and separation of nucleic acids from proteins and other cellular components. For viral DNA, the process must be efficient enough to capture potentially low viral loads while removing substances that could inhibit PCR or sequencing.

5.2.2 Procedure

A detailed, step-by-step protocol for DNA extraction must be followed using a validated commercial kit (e.g., QIAamp DNA Blood Mini Kit, MagMAX DNA Multi-Sample Ultra Kit). This protocol will specify critical parameters such as:

- **Sample Input Volumes:** The exact volume of clinical sample to be used for extraction.
- **Lysis Conditions:** Specific temperatures and incubation times for efficient viral particle lysis and release of DNA.
- **Binding, Wash Steps:** Detailed instructions for binding DNA to a silica membrane or magnetic beads, followed by thorough washing to remove contaminants.
- Final Elution Volume: The volume of elution buffer to use, optimized to maximize DNA yield and purity for downstream applications.
  Strict adherence to the manufacturer's instructions is crucial for consistent results.

5.2.3 DNA Quantification and Quality Assessment

Following extraction, the quality and quantity of the isolated DNA must be assessed.

- **Quantification:** Measure the extracted DNA concentration using a fluorometric method (e.g., Qubit 4 Fluorometer). This method is highly sensitive and specific for DNA, providing more accurate quantification than spectrophotometry, especially for low-concentration samples typical of viral diagnostics.
- **Purity Assessment:** Assess DNA purity using a spectrophotometer (e.g., NanoDrop) by measuring A260/A280 and A260/A230 ratios. Acceptable ratios typically fall within 1.8-2.0 for A260/A280 (indicating minimal protein contamination) and >1.8 for A260/A230 (indicating minimal carbohydrate or guanidine contamination).
- **Integrity Assessment (Optional):** For gross degradation, DNA integrity can be visually checked via agarose gel electrophoresis, though this is less critical for amplicon-based targeted sequencing where smaller fragments are expected.

5.3 MPXV VSP2 Panel Library Preparation

This stage involves preparing the extracted MPXV DNA for sequencing on Illumina platforms, specifically targeting the VSP2 regions.

5.3.1 DNA Quantification and Quality Assessment for Library Input

7Prior to initiating library preparation, it is essential to re-quantify the extracted DNA using a fluorometer. This step ensures that the DNA concentration meets the minimum input requirements of the chosen library preparation kit, which is typically in the nanogram range. Confirming the DNA purity metrics (A260/A280 and A260/A230 ratios) is also important to avoid inhibition of enzymatic reactions during library construction.

5.3.2 Library Preparation Protocol (Step-by-step for VSP2 panel)

The VSP2 panel library preparation typically involves targeted amplification of specific MPXV VSP2 regions, followed by enzymatic fragmentation (if necessary), adapter ligation, and incorporation of unique index sequences for multiplexing.

- **Principle:** The core principle is to selectively amplify the regions of interest within the MPXV genome (VSP2 panel) and then convert these amplified fragments into a library compatible with Illumina sequencing. This involves adding specific adapter sequences that facilitate binding to the flow cell and serve as primer binding sites for sequencing.
- **Procedure:** Strict adherence to the manufacturer's instructions for the chosen commercial MPXV VSP2 panel library preparation kit is paramount. General steps typically include:
    - **Target Amplification:** Performing multiplex PCR using VSP2-specific primers to amplify the desired genomic regions. This step is critical for enriching the target sequences from the total extracted DNA.
    - **PCR Product Cleanup:** Removal of excess primers, dNTPs, and primer dimers from the amplified products using magnetic beads (e.g., AMPure XP beads). This ensures clean templates for subsequent steps.
    - **End Repair and A-tailing (if applicable):** Enzymatic steps to prepare the DNA fragments for adapter ligation, often involving blunting ends and adding a single 'A' overhang.
    - **Adapter Ligation:** Ligation of sequencing adapters, which contain unique index sequences (barcodes), to the amplified fragments. These index sequences enable the multiplexing of multiple samples in a single sequencing run, allowing for cost-effective throughput.
    - **Post-Ligation Cleanup:** A second cleanup step to remove excess adapters and adapter dimers, which can otherwise compete for sequencing reads and reduce data quality.
    - **Library Amplification (Optional):** A final PCR step to amplify the library if needed to reach a sufficient concentration for sequencing. This step should be carefully optimized to avoid over-amplification and PCR bias.

5.3.3 Library Quality Control (Fragment Analysis, Quantification)

After library preparation, rigorous quality control is essential to ensure the prepared libraries are suitable for sequencing.

- **Quantification:** Accurately quantify the final prepared library using a fluorometric method (e.g., Qubit). This is crucial for proper pooling of libraries at equimolar concentrations prior to sequencing, ensuring balanced representation of all samples.
- **Size Distribution and Purity:** Assess the fragment size distribution and confirm the absence of adapter dimers using an automated electrophoresis system (e.g., Agilent Bioanalyzer or

TapeStation). A tight distribution around the expected amplicon sizes (plus adapter sequences) is critical for optimal sequencing performance. This directly relates to the "Library quality" and "Insert size" metrics emphasized for Illumina sequencing [1], ensuring that the majority of reads will originate from the target regions and not from unwanted side products.

5.4 Next-Generation Sequencing (NGS)

This section details the procedures for executing the sequencing run on Illumina platforms and managing the generated data.

5.4.1 Sequencing Platform Setup (Illumina iSeq100/NextSeq1000/2000)

- **System Checks:** Prior to initiating a run, perform all pre-run system checks to ensure the sequencer is properly maintained, calibrated, and ready for operation. This includes verifying fluidics, optics, and temperature controls.
- **Reagent Preparation:** Prepare sequencing reagents according to the manufacturer's instructions. This typically involves proper thawing, mixing, and loading of reagent cartridges.
- **Flow Cell Loading:** Carefully load the pooled and diluted library onto the flow cell, ensuring the correct loading volume and avoiding the introduction of air bubbles, which can disrupt cluster generation and sequencing.

5.4.2 Sequencing Run Parameters

- **Platform Selection:** Choose the appropriate Illumina platform based on the required throughput and the number of samples to be multiplexed. The iSeq100 is well-suited for smaller batches and rapid turnaround, while the NextSeq1000/2000 offers higher throughput for larger projects.[1]
- **Read Length:** Perform paired-end sequencing with a recommended read length of approximately 150 bp.[1] This read length is generally sufficient to cover most amplicon sizes within the VSP2 panel, allowing for robust alignment and accurate variant calling.
- **Sequencing Depth:** Aim for a minimum sequencing depth of ≥1000x coverage across the MPXV VSP2 panel target regions. This depth is significantly higher than the ≥40x recommended for bacterial whole-genome sequencing.[1] This elevated coverage is a critical adaptation for targeted viral panel analysis. Viral populations often exhibit quasispecies diversity, meaning that multiple genetic variants can coexist within a single host. Detecting these low-frequency variants, which may be present at allele frequencies of 1-5%, is crucial for understanding viral evolution, transmission dynamics, and the potential emergence of new strains. Insufficient depth would lead to missed variants, inaccurate allele frequency calls, and unreliable lineage assignments, compromising the utility of the analysis for public health decision-making. Therefore, the targeted nature of the VSP2 panel and the biological imperative of detecting low-frequency viral variants necessitate a much higher sequencing depth specifically focused on the on-target VSP2 regions, rather than a general whole-genome coverage.

5.4.3 Data Transfer and Storage

- **Data Transfer:** Establish secure and efficient protocols for transferring raw sequencing data (FASTQ files) from the sequencer to the designated bioinformatics workstation or server. This may involve direct network connections, secure copy (SCP), or other secure file transfer protocols to protect

sensitive data.

- **Data Storage:** Implement a robust data storage strategy that includes primary storage for active analysis and long-term archival storage for raw data, processed intermediate files, and final reports. Adhere strictly to institutional and national data retention policies. Data integrity must be ensured through regular backups, checksum verification, and redundant storage solutions to prevent data loss.

5.5 Bioinformatics Analysis Workflow for MPXV VSP2

The bioinformatics workflow leverages a Nextflow pipeline to automate the sequential processing of raw sequencing data, from initial quality control to reference-based alignment, variant calling, and final reporting. This automation ensures reproducibility, efficiency, and minimizes human error, building upon the Nextflow framework described in the provided AMR SOP.[1]

5.5.1 Setting up the Bioinformatics Environment

- **Nextflow Installation:** Ensure Nextflow is installed and properly configured on the Linux workstation/server.[1] This involves downloading the Nextflow executable and ensuring it is accessible in the system's PATH.
- **Conda/mamba Setup:** Install Conda or mamba [1] as the primary package managers. Mamba is often preferred for its speed in resolving dependencies.
- **Environment Creation:** Create dedicated Conda environments for each bioinformatics tool or logical group of tools (e.g., fastqc_Env, viral_alignment_Env, variant_calling_Env). This isolation of dependencies, following the conda create --name fastqc_Env example [1], prevents software conflicts and ensures consistent tool versions across different analyses.
- **Tool Installation:** Install all required software (FastQC, fastp, BWA, iVar, SnpEff, Kraken2) into their respective Conda environments, adhering to the installation instructions provided (e.g., mamba install -c bioconda fastqc).[1]
- **Database Setup:** Download and configure all necessary bioinformatics databases, including the MPXV reference genome, the VSP2 panel target coordinates (BED file), and the Kraken2 database for contamination checks. These databases are fundamental for accurate analysis.

5.5.2 Raw Read Quality Control and Trimming

- **FastQC Analysis:** Run FastQC on the raw paired-end FASTQ files for each sample to obtain an initial assessment of read quality.[1] Key metrics to review include per-base sequence quality, adapter content, sequence duplication levels, and GC content. This step provides a critical overview of the raw data's integrity.
    - **Command Example:** run_fastqc.sh -d <input_directory> -o <output_directory> (adapted

from [1]).

- **Read Trimming and Filtering:** Utilize fastp or Trimmomatic to remove low-quality bases (e.g., Phred score < Q20), adapter sequences, and reads shorter than a specified minimum length (e.g., 50 bp).[1] This step is crucial for improving the accuracy of downstream alignment and variant calling by removing spurious or low-confidence data.
- **Post-Trimming QC:** Re-run FastQC on the trimmed reads to verify quality improvement and ensure they meet the recommended quality score (≥Q30) and adapter content (max 1%) thresholds.[1] This confirms that the filtering process has successfully enhanced data quality.

5.5.3 Reference-Based Alignment and Variant Calling

- **Reference Genome Preparation:** Index the chosen MPXV reference genome (e.g., using bwa index or samtools faidx) to enable efficient read alignment. This creates necessary lookup tables for rapid mapping of millions of short reads.
- **Read Alignment:** Align the trimmed paired-end reads to the MPXV reference genome using BWA-MEM or minimap2. This step generates SAM/BAM files, which are compressed binary files containing the aligned reads and their mapping information.
    - **Command Example:** bwa mem <reference.fasta> <read1.fastq> <read2.fastq> | samtools view -bS - > <output.bam>
- **Alignment File Processing:** Sort and index the generated BAM files using Samtools. Sorting by genomic coordinate is necessary for efficient downstream variant calling and visualization, while indexing creates a quick lookup table for specific regions.
- **Primer Trimming (for Amplicon Data):** If an amplicon-based VSP2 panel is used, employ iVar to soft-clip or remove primer sequences from the aligned reads. This crucial step prevents primer-derived sequence variations from being erroneously called as true biological variants, which could otherwise lead to false positives in variant detection.
- **Variant Calling:** Perform variant calling using iVar (recommended for viral amplicon data) or GATK HaplotypeCaller/FreeBayes. This process identifies single nucleotide polymorphisms (SNPs) and insertions/deletions (indels) relative to the reference genome, generating a VCF (Variant Call Format) file.
    - **Filtering:** Apply stringent filtering criteria to variant calls, including a minimum allele frequency (e.g., ≥5%), minimum read depth supporting the variant (e.g., ≥20 reads), and strand bias filters, to ensure high confidence in reported variants and reduce false positives.
- **Consensus Sequence Generation:** Generate a consensus sequence for each sample based on the called variants. This sequence represents the dominant viral population in the sample, which is useful for downstream phylogenetic analysis and data submission.

5.5.4 MPXV VSP2 Panel Specific Analysis

- **Coverage Analysis of VSP2 Regions:** Calculate the average and minimum sequencing depth across each target region within the VSP2 panel using tools like Samtools depth or custom scripts. This is a critical quality control step to ensure sufficient data for reliable variant calling in all regions of interest, especially for detecting low-frequency variants.

- **Variant Annotation:** Annotate identified variants (from the VCF file) using SnpEff or VEP. This step determines the genomic location of each variant, its gene context, and its predicted functional impact (e.g., synonymous, missense, frameshift, stop-gain). This provides biological meaning to the detected genetic changes.
- **Comparison to Known Variants:** Compare the detected variants against a curated database of known MPXV variants of interest. This allows for the rapid identification of specific lineages, clades, or mutations with known epidemiological or clinical significance, aiding in outbreak characterization.
- **Phylogenetic Analysis (Optional but Valuable):** For outbreak investigations and understanding viral transmission, basic phylogenetic analysis can be conducted. This involves aligning consensus sequences from multiple samples and constructing a phylogenetic tree (e.g., using tools like IQ-TREE or FastTree). This provides insights into evolutionary relationships between isolates and helps identify potential transmission clusters.

Table 5.5.4.1: MPXV VSP2 Panel Target Regions and Expected Variants

This table is indispensable for providing the precise genomic context of the "MPXV VSP2 panel." It clearly delineates the specific regions of the MPXV genome that are targeted, along with their associated gene names and any known or expected variants. This level of detail is crucial for bioinformaticians to accurately validate their analysis, focus on the most relevant genomic segments, and interpret findings within the precise context of the panel's design. It directly translates the abstract "VSP2 panel" into actionable, genomic information, guiding targeted quality control and facilitating focused variant analysis and interpretation.

Target Region Name/ID	MPXV Reference Genome Coordinates (Start-End)	Associated Gene/Feature Name	Known/Expected Variants of Interest (Examples)	Forward Primer Sequence (Example)	Reverse Primer Sequence (Example)
VSP2_ORF_F8L	12345-12800	F8L (putative virulence factor)	SNP at 12500 (C>T) associated with Clade IIb; Deletion at 12650-12655	ATGCGTACGTAGCTAGCTAG	GATCGATCGATCGATCGATC
VSP2_ORF_G2R	15000-15450	G2R (immunomodulatory protein)	SNP at 15120 (A>G) linked to specific lineage;	TGCATGCATGCATGCATGC	CTAGCTAGCTAGCTAGCTAG

			Insertion at 15300		
VSP2_Intergenic_Region_1	20000-20250	Intergenic	Polymorphism at 20100 (T>C) for phylogenetic resolution	GCGCGCGCGCGCGCGCGC	ATATATATATATATATAT
VSP2_ORF_C1L	22000-22500	C1L (DNA ligase)	No common variants, highly conserved	AGCTAGCTAGCTAGCTAGC	GCTAGCTAGCTAGCTAGCT
VSP2_ORF_D8L	25000-25500	D8L (envelope protein)	SNP at 25150 (G>A) potentially affecting antigenicity	CGATCGATCGATCGATCG	ATCGATCGATCGATCGATCG
VSP2_ORF_A29L	28000-28500	A29L (major envelope protein)	Common SNP at 28200 (C>T) in Clade I	GCTAGCTAGCTAGCTAG	CTAGCTAGCTAGCTAGC

Note: The coordinates and example variants are illustrative and should be replaced with precise, validated information for the specific VSP2 panel and MPXV reference genome used.

5.6 Data Interpretation and Reporting

The final stage of the workflow involves synthesizing the analysis results into clear, actionable reports and ensuring proper data dissemination.

5.6.1 Generating Analysis Reports

An automated comprehensive summary report should be generated from the Nextflow pipeline. This report should include:

- **Sample Identifiers and Metadata:** Unique sample IDs and associated clinical/epidemiological metadata (e.g., sample type, collection date, patient demographics).
- **Raw and Trimmed Read QC Metrics:** Summaries from FastQC, including total reads, percentage of reads passing filter, and number of reads remaining after trimming.
- **Alignment Statistics:** Key metrics such as total reads aligned, percentage of reads aligning specifically to the MPXV reference genome (as per "Alignment rate" in [1]), and average coverage across the entire genome.
- **VSP2 Panel Coverage Statistics:** Detailed coverage statistics specifically for the VSP2 panel regions,

including average depth, minimum depth, and percentage of each region covered above a specified threshold (e.g., 100x).

- **Identified Variants:** A clear, tabular list of all identified variants (SNPs, indels) within the VSP2 regions, including genomic position, reference allele, alternate allele, allele frequency, and predicted functional impact (e.g., amino acid change).
- **MPXV Lineage Assignment:** The assigned MPXV lineage based on the detected variants (if applicable), utilizing established classification systems.

5.6.2 Interpretation of Results

- **Variant Significance:** Interpret the detected variants in the context of known MPXV biology, epidemiology, and any available clinical data. This involves identifying any variants associated with changes in viral virulence, transmissibility, or potential impact on diagnostic assays or vaccine efficacy.
- **Lineage Assignment:** Confirm the MPXV lineage based on the detected variants and provide relevant epidemiological context. This information is crucial for informing public health interventions, aiding in outbreak tracing, and understanding patterns of viral spread within and across communities.
- **Quality Assessment:** Critically review all quality control metrics generated throughout the workflow. Results from samples not meeting the specified QC thresholds (detailed in Section 6.4) should be flagged as unreliable, potentially requiring re-sequencing or further investigation to determine the cause of the quality failure.
- **Clinical/Epidemiological Context:** Integrate the sequencing findings with available clinical and epidemiological data. This holistic approach allows for the provision of actionable insights for public health interventions, such as informing contact tracing efforts, assessing outbreak dynamics, and guiding public health messaging.

5.6.3 Data Submission

Adherence to established guidelines for submitting raw sequencing data (FASTQ files) and consensus sequences/variant calls to public sequence repositories (e.g., NCBI Sequence Read Archive (SRA), GenBank, or GISAID for MPXV data) is mandatory. This ensures data accessibility for global surveillance and research, fostering international collaboration and rapid information sharing, while strictly adhering to data sharing policies and ethical considerations regarding patient privacy.

6. Quality Control and Assurance

Robust quality control (QC) and assurance measures are integrated throughout the entire MPXV VSP2 panel analysis workflow to ensure the accuracy, reliability, and comparability of results. This section details the recommended QC steps at each stage.

6.1 Pre-analytical QC (Sample Integrity)

- **Sample Verification:** Upon sample receipt, verify proper labeling, ensure the use of appropriate collection vessels, and check for any signs of leakage or degradation. Any discrepancies or issues must be thoroughly documented and addressed before proceeding.
- **Storage Conditions:** Confirm that samples have been stored and transported under appropriate conditions (e.g., cold chain maintenance) to preserve nucleic acid integrity. Compromised storage can lead to nucleic acid degradation and unreliable results.

6.2 Analytical QC (Library Quality, Sequencing Metrics, Run Performance)

- **Nucleic Acid QC:** Confirm that extracted DNA meets minimum purity (A260/A280 and A260/A230 ratios) and concentration thresholds (e.g., using Qubit) before proceeding to library preparation. Insufficient quality or quantity at this stage will negatively impact downstream steps.
- **Library QC:** Verify that the prepared libraries meet the required concentration, exhibit a tight fragment size distribution, and have minimal adapter dimer content before pooling and sequencing. These metrics are crucial for optimal cluster generation and sequencing output.
- **Sequencing Run QC:** Monitor key sequencing run metrics in real-time (if the sequencer allows) and post-run. These metrics include cluster density, percentage of reads passing filter, and average quality scores (Q-scores). Significant deviations from expected values indicate potential issues with the sequencing run itself, such as reagent problems, flow cell issues, or instrument malfunction, necessitating investigation.

6.3 Post-analytical QC (Bioinformatics Pipeline Validation, Data Accuracy)

- **Bioinformatics Pipeline Validation:** Regularly validate the entire bioinformatics pipeline using well-characterized positive control samples with known MPXV VSP2 variants and appropriate negative controls (e.g., water blank, host-only DNA). This ensures the pipeline's accuracy, sensitivity, and specificity in detecting expected variants and confirming the absence of false positives.
- **Data Accuracy Checks:** For a subset of samples, cross-reference automated variant calls from the pipeline with visual inspection of aligned reads in a genome browser (e.g., Integrative Genomics Viewer, IGV). This manual review helps confirm variant presence, assess read support, and identify any potential systematic errors in variant calling.
- **Contamination Monitoring:** Routinely check for cross-sample contamination or environmental contamination using tools like Kraken2.[1] High levels of non-target reads (e.g., human host DNA, other microbial contaminants) necessitate immediate investigation into the source of contamination and potential re-sequencing of affected samples.

6.4 Recommended Quality Metrics for MPXV VSP2 Analysis

This table is of paramount importance for ensuring the reliability, comparability, and actionable nature of MPXV VSP2 analysis results. It provides clear, quantitative thresholds for critical quality control parameters at each stage of the workflow, from raw reads to final variant calls. This allows laboratories to objectively assess the quality of their data, determine if results are acceptable for reporting, and identify specific points of failure, thereby directly supporting the principle of "Expected quality metrics" established in the general AMR SOP.[1] This comprehensive table provides a quick-reference guide and a clear, objective benchmark for data acceptance or rejection, standardizing the quality assessment process across the EAC network.

Table 6.4.1: Recommended Quality Control Metrics for MPXV VSP2 NGS Data

Metric	Description/Rationale	Recommended Threshold	Action if Threshold Not Met
Raw Read Quality (FastQC)	Initial assessment of raw sequencing data quality.	Per-base sequence quality score ≥Q30 for at least 80% of bases; adapter content <1%; no significant overrepresentation of sequences.	Review FastQC report, consider re-extraction if poor quality is inherent to sample, or re-sequencing.
Trimmed Read Quality	Quality of reads after adapter trimming and low-quality base removal.	Average Phred quality score ≥Q30 for remaining reads.[1]	Re-trim with more stringent parameters or investigate source of low-quality reads (e.g., library prep, sequencer).
Read Length	Length of retained reads after trimming.	Retained reads ≥100 bp.[1]	Adjust trimming parameters; investigate library preparation for fragmentation issues.
MPXV Alignment Rate	Percentage of trimmed reads that successfully align to the MPXV reference	≥90% of trimmed reads should align to the MPXV reference genome (adapted	Lower rates indicate poor sample quality, high host contamination, or

	genome.	from [1] for bacterial alignment).	non-MPXV sample. Investigate source of non-alignment.
VSP2 Panel Coverage (Average)	Average sequencing depth across all VSP2 target regions.	≥1000x average coverage across all VSP2 target regions.	This high depth is crucial for detecting low-frequency variants. If not met, consider re-sequencing the sample at a higher depth or re-amplifying.
VSP2 Panel Coverage (Minimum)	Minimum sequencing depth at any single base within the VSP2 target regions.	≥100x coverage at any single base within the VSP2 target regions.	Ensures that even the lowest-covered areas within the panel have sufficient depth for reliable variant calling. If not met, investigate primer drop-out or re-sequence.
Variant Allele Frequency (VAF) Threshold	Minimum allele frequency required for calling a minor variant.	Minimum VAF of ≥5% for calling minor variants (can be adjusted based on specific needs).	Adjust threshold based on specific research question or clinical context, considering implications for sensitivity vs. specificity.
Contamination Rate	Percentage of non-MPXV reads (e.g., human host DNA, other microbial contaminants).	Non-MPXV reads <1% of total reads (adapted from [1] for adapter contamination).	Investigate source of contamination (e.g., sample collection, extraction, lab environment); consider re-extraction or re-sequencing.
Completeness of VSP2 Regions	Percentage of each VSP2 target region	≥99% of each VSP2 target region should	Indicates successful amplification and

	covered above the minimum threshold.	have coverage above the minimum threshold (e.g., 100x).	sequencing of all targeted regions. If not met, investigate missing regions (e.g., primer issues, deletions).

7. Troubleshooting Guide

This section provides guidance for common issues encountered during MPXV VSP2 panel analysis and suggested solutions.

7.1 Low DNA Yield/Quality

- **Problem:** Insufficient DNA concentration or poor purity (e.g., low A260/A280, A260/A230 ratios) after extraction.
- **Solution:** Re-extract the sample with an optimized protocol, which may include increasing the initial sample input volume, trying a different extraction kit known for higher yields from challenging sample types, or adjusting lysis conditions. Thoroughly check sample storage conditions prior to extraction, as degradation can lead to low yields. Consider concentrating the extracted DNA using methods like ethanol precipitation if the volume is high but concentration is low.

7.2 Poor Library Quality

- **Problem:** Low library concentration, broad fragment size distribution, or high adapter dimer content after library preparation (as assessed by fluorometer and automated electrophoresis system).
- **Solution:** Optimize library preparation steps: adjust PCR cycles to avoid over-amplification or under-amplification, verify enzyme activity by checking expiry dates and storage conditions, and ensure the initial DNA input quality meets specifications. If adapter dimers are present, perform additional magnetic bead-based cleanup steps with adjusted bead-to-sample ratios.

7.3 Low Sequencing Output/Quality

- **Problem:** Low cluster density on the flow cell, low average Q-scores, or a low percentage of reads passing filter from the sequencer.
- **Solution:** Conduct a comprehensive check of the sequencer maintenance logs to identify any recurring issues. Verify the expiry dates and proper preparation of all sequencing reagents. Ensure proper flow cell loading, avoiding air bubbles. Troubleshoot library pooling and denaturation steps, as incorrect concentrations or incomplete denaturation can lead to poor cluster generation.

If issues persist, contact instrument technical support.

7.4 Low MPXV Alignment Rate

- **Problem:** A high percentage of trimmed reads do not align to the MPXV reference genome.
- **Solution:** First, check for high host contamination by running tools like Kraken2 to identify and quantify non-MPXV sequences. If contamination is severe, consider re-extracting the sample with a method that better removes host nucleic acids. Confirm that the correct MPXV reference genome is being used for alignment and that it is properly indexed. If the sample is suspected to be non-MPXV or a highly divergent strain, consider broader metagenomic analysis.

7.5 Low VSP2 Panel Coverage

- **Problem:** Insufficient sequencing depth across targeted VSP2 regions, despite adequate total reads.
- **Solution:** Investigate the VSP2 primer design and specificity; poorly designed primers can lead to inefficient amplification. Optimize PCR amplification steps, including annealing temperatures and cycle numbers. Ensure proper library normalization and pooling concentrations to guarantee equitable representation of all samples and targets. If target regions are consistently underrepresented, consider re-sequencing the sample at a higher depth or redesigning problematic primers.

7.6 Unexpected Variants/No Variants

- **Problem:** Variant calls do not match expected patterns for known controls, or no variants are detected in a sample expected to be positive for MPXV.
- **Solution:** Review variant calling parameters, such as the Variant Allele Frequency (VAF) threshold and minimum read depth, to ensure they are appropriate for detecting low-frequency variants. Check for primer binding issues that might mask variants in amplicon-based panels (e.g., a variant occurring under a primer binding site). Confirm sample identity and the expected MPXV lineage for controls. If necessary, manually inspect aligned reads in a genome browser to verify variant presence.

7.7 Bioinformatics Pipeline Errors

- **Problem:** The Nextflow pipeline fails to complete or produces unexpected output files or errors.
- **Solution:** Review the pipeline's log files for specific error messages; these often provide direct clues about the cause of the failure. Verify that all software tools are correctly installed in their respective Conda environments [1] and that all dependencies are met. Ensure correct input file paths and permissions for all directories and files accessed by the pipeline. Check available disk space and RAM, as resource limitations can cause pipeline crashes.

9. Appendices

Appendix A: Example Sample Sheet Template

A detailed template for organizing sample metadata is crucial for consistent data capture and traceability throughout the workflow, aligning with the necessity for creating a sample sheet for tools like AQUAMIS.[1]

Field Name	Description	Example
Sample ID	Unique identifier for each sample	MPXV_EAC_001_2024
Date Collected	Date of clinical sample collection (YYYY-MM-DD)	2024-09-15
Date Received	Date sample received in the lab (YYYY-MM-DD)	2024-09-17
Clinical Diagnosis	Clinical suspicion or confirmed diagnosis	Suspected Mpox, Confirmed Mpox
Sample Type	Type of clinical specimen	Lesion Swab, Crust, Plasma
Patient ID (De-identified)	Unique, de-identified patient identifier	P_XYZ_005
Nucleic Acid Concentration (ng/µL)	Concentration of extracted DNA	15.2
Nucleic Acid Purity (A260/A280)	Purity ratio of extracted DNA	1.85
Library Concentration (nM)	Concentration of prepared library	12.5
Barcode/Index Used	Unique dual index (UDI) sequences used for multiplexing	i7_NNNNNNNN_i5_NNNNNNNN
Sequencing Run ID	Identifier for the sequencing run	NGS_Run_20241001_01
Notes	Any relevant observations or deviations	Low sample volume, Sample hemolyzed

Appendix B: Example Bioinformatics Command Lines

Detailed command-line examples for executing key bioinformatics tools within the Nextflow pipeline provide practical guidance for bioinformaticians. These examples illustrate typical parameters for each step.

1. FastQC Analysis:

```
fastqc -o /path/to/FastQC_output /path/to/raw_reads/sample_R1.fastq.gz /path/to/raw_reads/sample_R2.fastq.gz
```

2. Read Trimming with fastp:

```
fastp -i /path/to/raw_reads/sample_R1.fastq.gz \
    -I /path/to/raw_reads/sample_R2.fastq.gz \
    -o /path/to/trimmed_reads/sample_trimmed_R1.fastq.gz \
    -O /path/to/trimmed_reads/sample_trimmed_R2.fastq.gz \
    --qualified_quality_phred 20 \
    --length_required 50 \
    --detect_adapter_for_pe \
    -j /path/to/logs/sample.fastp.json \
    -h /path/to/logs/sample.fastp.html
```

3. BWA Indexing of MPXV Reference Genome (one-time setup):

```
bwa index /path/to/reference/MPXV_reference.fasta
samtools faidx /path/to/reference/MPXV_reference.fasta
```

4. Read Alignment with BWA-MEM:

```
bwa mem -t 8 /path/to/reference/MPXV_reference.fasta \
      /path/to/trimmed_reads/sample_trimmed_R1.fastq.gz \
      /path/to/trimmed_reads/sample_trimmed_R2.fastq.gz \

| samtools view -bS - \
| samtools sort -o /path/to/aligned_bams/sample.sorted.bam -
samtools index /path/to/aligned_bams/sample.sorted.bam
```

5. Primer Trimming and Variant Calling with iVar:
Assumes a BED file vsp2_primers.bed defining primer regions.

```
ivar trim -i /path/to/aligned_bams/sample.sorted.bam \
        -b /path/to/primers/vsp2_primers.bed \
```

```
    -p /path/to/ivar_output/sample_trimmed_ \
    -m 10 # Minimum read length after trimming

ivar variants -p /path/to/ivar_output/sample.variants \
      -m 5 \
      -t 0.05 \
      -q 20 \
      -d 100 \
      /path/to/ivar_output/sample_trimmed_align.bam
```

Note: -m 5 for minimum variant frequency (0.05 = 5%), -q 20 for minimum quality score, -d 100 for minimum depth.

6. Variant Annotation with SnpEff:
Assumes SnpEff database for MPXV is built.

```
snpeff -v MPXV_DB /path/to/ivar_output/sample.variants.vcf > /path/to/annotated_vcf/sample.annotated.vcf
```

Appendix C: Example Report Template

A structured template for the final MPXV VSP2 analysis report ensures consistency and completeness in reporting across the EAC network.

MPXV VSP2 Panel Analysis Report

Laboratory: [Laboratory Name]
Report Date:

1. Sample Information

- **Sample ID:** [MPXV_EAC_001_2024]
- **Date Collected:**
- **Date Received:**
- **Sample Type:**
- **Patient ID (De-identified):** [P_XYZ_005]
- **Clinical Diagnosis:**
- **Sequencing Run ID:**

2. Quality Control Summary

- **Nucleic Acid Input:**
    - Concentration: [15.2 ng/µL]
    - Purity (A260/A280): [1.85]

- **Library Quality:**
  - Concentration: [12.5 nM]
  - Fragment Size Distribution: [e.g., Peaks at 300-400 bp, no adapter dimers]
- **Sequencing Run Metrics:**
  - Total Raw Reads: [e.g., 5,000,000]
  - Reads Passing Filter: [e.g., 95%]
  - Average Raw Read Quality (Q-score): [e.g., Q35]
  - Adapter Content (Post-trimming): [<0.5%]
- **Bioinformatics QC:**
  - MPXV Alignment Rate: [e.g., 98.2%]
  - Contamination Rate (Non-MPXV reads): [<0.1%]
  - Average VSP2 Panel Coverage: [e.g., 1500x]
  - Minimum VSP2 Panel Coverage: [e.g., 250x]
  - Completeness of VSP2 Regions (≥100x coverage): [e.g., 100%]

3. MPXV VSP2 Panel Analysis Results

- **MPXV Lineage Assignment:**
- **Identified Variants within VSP2 Panel (VAF ≥ 5%):**

Genomic Position	Reference Allele	Alternate Allele	Variant Allele Frequency (%)	Gene/Feature	Predicted Effect
12500	C	T	98.5	F8L	Missense (Pro123Ser)
15120	A	G	99.2	G2R	Synonymous
25150	G	A	6.8	D8L	Missense (Ala45Val)
...	...	...	...	...	...

- **VSP2 Panel Coverage Map (Optional: Include a visual representation or detailed table of coverage per target region)**

4. Interpretation and Discussion

- The sample was successfully sequenced for the MPXV VSP2 panel, yielding high-quality data meeting all established QC thresholds.
- The analysis indicates the presence of MPXV belonging to. This lineage is currently [e.g., widely circulating globally / associated with recent outbreaks in the region].
- Key variants identified include. A low-frequency variant was also detected at 6.8% VAF, which may

represent [e.g., a minor viral population or an early emerging mutation]. Further investigation or re-sequencing at higher depth may be considered for variants of uncertain significance.

- The genetic profile observed is consistent with [e.g., known epidemiological patterns in the EAC region / a potential introduction from a specific geographic area].
- [Add any specific clinical or epidemiological correlations, if available and relevant.]

5. Conclusions and Recommendations

The MPXV VSP2 panel analysis for sample provides high-resolution genetic data confirming the presence of MPXV and assigning its lineage. This information is crucial for public health surveillance, enabling precise tracking of viral spread and evolution within the East African Community. The robust quality control measures implemented throughout the pipeline ensure the reliability of these findings.

It is recommended that:

- All data meeting the specified quality thresholds be submitted to public sequence repositories (e.g., GenBank, GISAID) to contribute to global MPXV surveillance efforts.
- Continued monitoring of circulating MPXV lineages and the emergence of new variants through this standardized NGS pipeline is essential for informing public health interventions and assessing the effectiveness of diagnostic tools and potential countermeasures.
- For samples with low-frequency variants of particular interest, follow-up investigations or deeper sequencing may be warranted to confirm their biological significance and epidemiological impact.