

South Africa - Vukuzazi HIV polymerase sequences:2018-2020.

Sweetness H Dube

Report generated on: July 17, 2023

Visit our data catalog at: <https://data.ahri.org/index.php>

Identification

SURVEY ID NUMBER

AHRI.Vukuzazi.PANGEASequences.2023.v1

TITLE

Vukuzazi HIV polymerase sequences:2018-2020.

COUNTRY

Name	Country code
South Africa	ZA

ABSTRACT

Vukuzazi ("Wake up and know yourself " in isiZulu) was established in 2018 to offer a community-based health phenotyping and comprehensive bio-sampling to all resident adult (=15 years) members of the Africa Health Research Institute demographic health surveillance population, while building upon the existing wealth of demographic and HIV information collected over the previous 20 years. The objective of the platform was to determine the prevalence and overlap of infectious diseases and NCDs in the population 20 years into the HIV epidemic. Additionally, the platform aimed to create a data, image and biorepository that could be used to understand the host, pathogen, social and environmental determinants of specific states of health and disease in the population.

The purpose of the genomic HIV data was to understand HIV transmission patterns among HIV infected individuals in rural KwaZulu Natal. To address this question, consensus polymerase sequences were derived from HIV positive samples with viral load greater than 200 copies/ml.

KIND OF DATA

HIV Genomic Data, Viral Load

UNIT OF ANALYSIS

Each sequence derived from a single specimen

Version

VERSION DESCRIPTION

v1.0.0

Scope

TOPICS

Topic	Vocabulary	URI
HIV-1;phylogeny;population surveillance; Epidemics; HIV infections; Africa	Africa Health Research Institute	www.ahri.org

KEYWORDS

Keyword	Vocabulary	URI
HIV polymerase sequence; phylodynamics; phylogeny; HIV-1	Africa Health Research Institute	www.ahri.org

Coverage

GEOGRAPHIC COVERAGE

uMkhanyakude district in northern KwaZulu-Natal

UNIVERSE

HIV positive individuals enrolled in Vukuzazi study with a VL >200 copies/ml

Producers and sponsors

PRIMARY INVESTIGATORS

Name	Affiliation
Prof. Willem Hanekom	Africa Health Research Institute
Prof. Thumbi Ndung'u	Africa Health Research Institute
Dr. Kobus Herbst	Africa Health Research Institute

PRODUCERS

Name
Africa Health Research Institute

FUNDING AGENCY/SPONSOR

Name	Abbreviation	Role
Wellcome Trust	WT	Core funding

OTHER IDENTIFICATIONS/ACKNOWLEDGMENTS

Name	Affiliation	Role
Sweetness Dube	AHRI	Data Documentation
Dickman Gareta	AHRI	Head of Research Data Management
Dr. Anne Derache	AHRI	Clinical Research Coordinator
Dr. Resign Gunda	AHRI	Vukuzazi Programme Manager
PANGEA consortium	Big Data Institute Oxford	Sequencing platform

Sampling

SAMPLING PROCEDURE

HIV positive individuals enrolled in Vukuzazi study between 2018 and 2020. Individuals blood were attempted for sequencing if they:

1. Had a viral load done.

Viral load was above 200 copies/ml

Data Collection

DATES OF DATA COLLECTION

Start	End
2018-01-01	2020-03-30

Data Processing

DATA EDITING

Consensus sequences were derived from full-length HIV Next Generation Sequencing data generated by the PANGEA-HIV 2

(phylogenetics and Networks for Generalised Epidemics in Africa). Consensus sequences are presented in the standard fasta file format.

Access policy

ACCESS CONDITIONS

Access to the data requires accurate completion of the online data access application form accessible on the AHRI Data repository(<<https://data.ahri.org/>>). Data users are required to abide by the data use conditions stipulated on the application for access to the data. Failure to do so may result in their data access privileges being revoked by the Data Custodian. In order to recognise the effort and intellectual contributions of AHRI investigators in producing and curating the data, users of AHRI data must acknowledge the source of the data and abide by the terms and conditions under which the data is accessed and must cite the dataset in publication using the citation provided as part of this documentation. All analytical datasets published on the AHRI Data Repository are assigned digital object identifier (DOIs) and the DOIs can be found on the Data Repository under Study Description tab - Access policy. AHRI data users are required to always cite the dataset using the relevant DOI.

CITATION REQUIREMENTS

Hanekom, W., Ndung'u, T., & Herbst, K. (2023). Vukuzazi HIV polymerase sequences:2018-2020. [Data set]. Africa Health Research Institute.

DOI:<https://doi.org/10.23664/AHRI.Vukuzazi.PANGEASequences.2023>

Metadata production

DDI DOCUMENT ID

DDI.AHRI.Vukuzazi.PANGEASequences.2023.v1

PRODUCERS

Name	Abbreviation
Africa Health Research Institute	AHRI

Data Dictionary

Data file	Cases	Variables
AHRI.Vukuzazi.PANGEASequence.2023.v1	1112	7

Data file: AHRI.Vukuzazi.PANGEASequences.2023.v1

Cases: 1112

Variables: 7

Variables

ID	Name	Label	Question
V299	IndividualId	Vukuzazi Individual ID	
V300	SpecimenId	Specimen ID	
V301	AliquotId	Aliquot ID	
V302	MappedNum	Mapped Number	
V303	LengthRelaxed	Length Relaxed	
V304	ShiverConsensus	Shiver Consensus	
V305	PANGEAId	PANGEA ID	

Total: 7

INDIVIDUALID: Vukuzazi Individual ID**Data file: AHRI.Vukuzazi.PANGEASequences.2023.v1****Overview**

Valid: 1112 Invalid: 0
 Type: Discrete Width: 12 Range: - Format: character

SPECIMENID: Specimen ID**Data file: AHRI.Vukuzazi.PANGEASequences.2023.v1****Overview**

Valid: 1112 Invalid: 0
 Type: Discrete Width: 14 Range: - Format: character

ALIQUOTID: Aliquot ID**Data file: AHRI.Vukuzazi.PANGEASequences.2023.v1****Overview**

Valid: 1112 Invalid: 0
 Type: Discrete Width: 13 Range: - Format: character

MAPPEDNUM: Mapped Number**Data file: AHRI.Vukuzazi.PANGEASequences.2023.v1****Overview**

Valid: 1108 Invalid: 0
 Type: Discrete Width: 9 Range: - Format: character

LENGTHRELAXED: Length Relaxed**Data file: AHRI.Vukuzazi.PANGEASequences.2023.v1****Overview**

Valid: 1098 Invalid: 0
 Type: Discrete Width: 13 Range: - Format: character

SHIVERCONSENSUS: Shiver Consensus**Data file: AHRI.Vukuzazi.PANGEASequences.2023.v1****Overview**

Valid: 1098
 Type: Discrete Width: 255 Range: - Format: character

PANGAID: PANGA ID

Data file: AHRI.Vukuzazi.PANGASequence.2023.v1

Overview

Valid: 1112 Invalid: 0

Type: Discrete Width: 13 Range: - Format: character
