

South Africa - PANGEA1 HIV Clinical Data

Report generated on: February 22, 2019

Visit our data catalog at: <https://data.africacentre.ac.za/index.php>

Overview

Identification

ID NUMBER

AHRI.PANGA1.HIV.Clinical.Data.2019.v1

Version

VERSION DESCRIPTION

V1.0.0

Overview

ABSTRACT

The purpose of this study was address one key question: What is the contribution of external sources of HIV in driving the epidemic within a community. High rates of cross-community transmission calls into question the efficacy of increasing local coverage of antiretroviral therapy to prevent new infections. To address this question, 1068 polymerase sequences were derived from sampling conducted within the Africa Health Research Institute's demographic surveillance area between 2010 and 2013. AHRI genotypes were analysed against a large national dataset of homologous sequences in a novel phylodynamic framework to infer the overall transmission dynamics within the community. This allowed us to enumerate the relative contribution of local transmission versus external introductions to the overall HIV incidence in the community. Major variables within the dataset to answer this question are: (1) the genotype; (2) the community's location from which the genotype was derived; and (3) the date of sampling.

KIND OF DATA

HIV Genomic Data

UNITS OF ANALYSIS

Each sequences derived from a single specimen.

TOPICS

Topic	Vocabulary	URI
HIV-1; Incidence; Phylogeny; Epidemics; Population Surveillance; Rural Population; HIV Infections; Africa	Africa Health Research Institute	AHRI

KEYWORDS

HIV polymerase sequence; date of sampling; phylodynamics, phylogeny, HIV-1

Coverage

GEOGRAPHIC COVERAGE

Treatment as Prevention trial area of the Africa Health Research Institute.

UNIVERSE

HIV positive participants enrolled in the Treatment as Prevention 12449 ANRS trial, who linked to care

Producers and Sponsors

PRIMARY INVESTIGATOR(S)

Name	Affiliation
De Oliveira, Tulio	KwaZulu-Natal Research Innovation Sequencing Platform
Pillay, Deenan	Africa Health Research Institute

Name	Affiliation

OTHER PRODUCER(S)

Name	Affiliation	Role
Africa Health Research Institute		

FUNDING

Name	Abbreviation	Role
South African Medical Research Council	SAMRC	Genotyping funding source

OTHER ACKNOWLEDGEMENTS

Name	Affiliation	Role
Wilkinson, Eduan	KwaZulu-Natal Research Innovation Sequencing Platform	Cleaned, aligned and help analyse the sequence data.

Metadata Production

METADATA PRODUCED BY

Name	Abbreviation	Affiliation	Role
Africa Health Research Institute	AHRI		

DDI DOCUMENT ID

DDI.AHRI.PANGA1.HIV.Clinical.Data.2019.v1

Sampling

Sampling Procedure

HIV positive individuals within the demographic surveillance area of the Africa Health Research Institute from 2010 to 2013. Individuals blood spots were only attempted sequencing if they:

A) Had a viral load done

B) Viral load was above 1,550 copies/ml

C) The laboratory got approximately 60% success rate on sequencing them, they should have the list that they attempt to sequence and the ones that were successful.

Questionnaires

No content available

Data Collection

Data Collection Dates

Start	End	Cycle
2010-01-01	2013-12-04	N/A

Data Processing

Data Editing

Sequences were generated by the Durban based laboratory of AHRI. Sequences were aligned with one another in ClustalW and are presented in the standard fasta file format.

Data Appraisal

No content available

File Description

Variable List

Hexagons

Content

Cases 1021

Variable(s) 5

Structure
Type:
Keys: ()

Version

Producer

Missing Data

Variables

ID	Name	Label	Type	Format	Question
V87	Id		contin	numeric	
V88	Uid		discrete	character	
V89	Perimeter		discrete	character	
V90	Centroid		discrete	character	
V91	t		discrete	character	

PangeaIndividualsFinalTasP

Content

Cases 3966

Variable(s) 24

Structure Type:
Keys: ()

Version

Producer

Missing Data

Variables

ID	Name	Label	Type	Format	Question
V92	IndividualId	Unique Individual Identifier	contin	numeric	
V93	SampleId	Unique Sample Identifier	discrete	character	
V94	Source	Source Project	discrete	character	
V95	pangea_id	PANGA Unique Identifier	discrete	character	
V96	dob	Date of Birth	discrete	character	
V97	Age	Age	contin	numeric	
V98	Gender	Gender	discrete	character	
V99	sample_date	Sample Collection Date	discrete	character	
V100	ACDIS_Id	ACDIS Unique Identifier	contin	numeric	
V101	ARTEMIS_Id	ARTEMIS Unique Identifier	contin	numeric	
V102	ACCDB_Id	ACCDB Unique Identifier	contin	numeric	
V103	TasP_Id	TasP Unique Identifier	contin	numeric	
V104	LatestNegative	Last HIV Negative Date	discrete	character	
V105	EarliestPositive	Earliest HIV Positive Date	discrete	character	
V106	EarliestKnownART	Earliest Known ART Start Date	discrete	character	
V107	LastKnownART	Last Known ART Start Date	discrete	character	
V108	LastKnownARTRegimen	Last Known ART Regimen	discrete	character	
V109	LastKnownAdhereScore	Last Known ART Adherence VAS Score	contin	numeric	
V110	ViralLoad	Viral Load	contin	numeric	
V111	ViralLoadBelowLDL	Was Viral Load below LDL?	discrete	numeric	
V112	CD4Count	CD4 Count - Analysed by LAB on Sample	contin	numeric	
V113	CD4Count_PIMA	CD4 Count - Analysed by Clinic Staff on PIMA	contin	numeric	
V114	CD4Pct	CD4 Percentage - Analysed by LAB on Sample	contin	numeric	
V115	HexagonId	Hexagon Unique Identifier	contin	numeric	

(Id)

File: Hexagons

Overview

Type: Continuous	Valid cases: 1021
Format: numeric	Invalid: 0
Width: 12	Minimum: 1
Decimals: 0	Maximum: 1021
Range: 1-1021	Mean: 511
	Standard deviation: 294.9

(Uid)

File: Hexagons

Overview

Type: Discrete	Valid cases: 1021
Format: character	Invalid: 0
Width: 36	

(Perimeter)

File: Hexagons

Overview

Type: Discrete	Valid cases: 1021
Format: character	
Width: 244	

(Centroid)

File: Hexagons

Overview

Type: Discrete	Valid cases: 1021
Format: character	
Width: 244	

(t)

File: Hexagons

Overview

Type: Discrete	Valid cases: 1021
Format: character	
Width: 244	

Unique Individual Identifier (IndividualId)

File: PangeaIndividualsFinalTasP

Overview

Type: Continuous	Valid cases: 3966
Format: numeric	Invalid: 0
Width: 12	Minimum: 1
Decimals: 0	Maximum: 200695
Range: 1-200695	Mean: 35127.4
	Standard deviation: 58416.4

Unique Sample Identifier (SampleId)

File: PangeaIndividualsFinalTasP

Overview

Type: Discrete	Valid cases: 3966
Format: character	
Width: 244	

Source Project (Source)

File: PangeaIndividualsFinalTasP

Overview

Type: Discrete	Valid cases: 3966
Format: character	
Width: 244	

PANGEA Unique Identifier (pangea_id)

File: PangeaIndividualsFinalTasP

Overview

Type: Discrete	Valid cases: 3966
Format: character	
Width: 244	

Date of Birth (dob)

File: PangeaIndividualsFinalTasP

Overview

Type: Discrete	Valid cases: 3966
Format: character	Minimum: NaN
Width: 11	Maximum: NaN

Age (Age)

File: PangeaIndividualsFinalTasP

Overview

Age (Age)

File: PangeaIndividualsFinalTasP

Type: Continuous
 Format: numeric
 Width: 12
 Decimals: 0
 Range: 14-114

Valid cases: 3966
 Invalid: 0
 Minimum: 14
 Maximum: 114
 Mean: 36.4
 Standard deviation: 12.3

Gender (Gender)

File: PangeaIndividualsFinalTasP

Overview

Type: Discrete
 Format: character
 Width: 244

Valid cases: 2590

Sample Collection Date (sample_date)

File: PangeaIndividualsFinalTasP

Overview

Type: Discrete
 Format: character
 Width: 11

Valid cases: 3966
 Minimum: NaN
 Maximum: NaN

ACDIS Unique Identifier (ACDIS_Id)

File: PangeaIndividualsFinalTasP

Overview

Type: Continuous
 Format: numeric
 Width: 12
 Decimals: 0
 Range: 84-218561

Valid cases: 2785
 Invalid: 1181
 Minimum: 84
 Maximum: 218561
 Mean: 160690.8
 Standard deviation: 53859.4

ARTEMIS Unique Identifier (ARTEMIS_Id)

File: PangeaIndividualsFinalTasP

Overview

Type: Continuous
 Format: numeric
 Width: 12
 Decimals: 0
 Range: 41-200695

Valid cases: 1367
 Invalid: 2599
 Minimum: 41
 Maximum: 200695
 Mean: 92656.5
 Standard deviation: 74427.2

ACCDB Unique Identifier (ACCDB_Id)

File: PangeaIndividualsFinalTasP

Overview

ACCDB Unique Identifier (ACCDB_Id)

File: PangeaIndividualsFinalTasP

Type: Continuous
 Format: numeric
 Width: 12
 Decimals: 0
 Range: 4-56761

Valid cases: 2626
 Invalid: 1340
 Minimum: 4
 Maximum: 56761
 Mean: 32619
 Standard deviation: 18233.3

TasP Unique Identifier (TasP_Id)

File: PangeaIndividualsFinalTasP

Overview

Type: Continuous
 Format: numeric
 Width: 12
 Decimals: 0
 Range: 9-100102

Valid cases: 2590
 Invalid: 1376
 Minimum: 9
 Maximum: 100102
 Mean: 11710.9
 Standard deviation: 8333.6

Last HIV Negative Date (LatestNegative)

File: PangeaIndividualsFinalTasP

Overview

Type: Discrete
 Format: character
 Width: 11

Valid cases: 304
 Minimum: NaN
 Maximum: NaN

Earliest HIV Positive Date (EarliestPositive)

File: PangeaIndividualsFinalTasP

Overview

Type: Discrete
 Format: character
 Width: 11

Valid cases: 3955
 Minimum: NaN
 Maximum: NaN

Earliest Known ART Start Date (EarliestKnownART)

File: PangeaIndividualsFinalTasP

Overview

Type: Discrete
 Format: character
 Width: 11

Valid cases: 3661
 Minimum: NaN
 Maximum: NaN

Last Known ART Start Date (LastKnownART)

File: PangeaIndividualsFinalTasP

Overview

Type: Discrete
 Format: character
 Width: 11

Valid cases: 2256
 Minimum: NaN
 Maximum: NaN

Last Known ART Regimen (LastKnownARTRegimen)

File: PangeaIndividualsFinalTasP

Overview

Type: Discrete
Format: character
Width: 244

Valid cases: 2256

Last Known ART Adherence VAS Score (LastKnownAdhereScore)

File: PangeaIndividualsFinalTasP

Overview

Type: Continuous
Format: numeric
Width: 12
Decimals: 0
Range: 1-100

Valid cases: 2102
Invalid: 1864
Minimum: 1
Maximum: 100
Mean: 96.8
Standard deviation: 8.9

Viral Load (ViralLoad)

File: PangeaIndividualsFinalTasP

Overview

Type: Continuous
Format: numeric
Width: 12
Decimals: 0
Range: 1-6277000

Valid cases: 2527
Invalid: 1439
Minimum: 1
Maximum: 6277000
Mean: 165876.3
Standard deviation: 438237.8

Was Viral Load below LDL? (ViralLoadBelowLDL)

File: PangeaIndividualsFinalTasP

Overview

Type: Discrete
Format: numeric
Width: 2
Decimals: 0
Range: 0-1

Valid cases: 2526
Invalid: 1440

CD4 Count - Analysed by LAB on Sample (CD4Count)

File: PangeaIndividualsFinalTasP

Overview

Type: Continuous
Format: numeric
Width: 12
Decimals: 0
Range: 2-782

Valid cases: 170
Invalid: 3796
Minimum: 2
Maximum: 782
Mean: 270.5
Standard deviation: 225.8

CD4 Count - Analysed by Clinic Staff on PIMA (CD4Count_PIMA)

File: PangeaIndividualsFinalTasP

Overview

Type: Continuous	Valid cases: 2128
Format: numeric	Invalid: 1838
Width: 12	Minimum: 4
Decimals: 0	Maximum: 3134
Range: 4-3134	Mean: 416.9
	Standard deviation: 253.1

CD4 Percentage - Analysed by LAB on Sample (CD4Pct)

File: PangeaIndividualsFinalTasP

Overview

Type: Continuous	Valid cases: 170
Format: numeric	Invalid: 3796
Width: 12	Minimum: 2
Decimals: 0	Maximum: 37
Range: 2-37	Mean: 19.5
	Standard deviation: 8

Hexagon Unique Identifier (HexagonId)

File: PangeaIndividualsFinalTasP

Overview

Type: Continuous	Valid cases: 2589
Format: numeric	Invalid: 1377
Width: 12	Minimum: 412
Decimals: 0	Maximum: 1013
Range: 412-1013	Mean: 846.5
	Standard deviation: 103.6

Documentation

Technical documents

Documentation:PANGEA HIV Clinical Data

Title Documentation:PANGEA HIV Clinical Data
Author(s) Sweetness H Dube
Date 21/02/2019
Country South Africa
Language English
Filename PANGEA HIV Clinical Data.pdf

DDI:PANGEA1 HIV Clinical Data

Title DDI:PANGEA1 HIV Clinical Data
Author(s) Sweetness H Dube
Date 21/02/2019
Country South Africa
Language English
Filename AHRI.PANGEA1.HIV.Clinical.Data.2019.v1.pdf
