

South Africa - PANGEA1 Demographic and Clinical Data.2019.v1

Sweetness H Dube

Report generated on: April 18, 2019

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

Overview

Identification

ID NUMBER

AHRI.PANGA1.Data.2019.v1

Version

VERSION DESCRIPTION

V1.0.0

Overview

ABSTRACT

PANGA is an international consortium of researchers based in Africa, the US and the UK studying transmission dynamics in HIV epidemics in sub-Saharan Africa. The goal of the consortium is to analyse HIV phylogenetic and demographic data to identify individual and population-level factors that drive the epidemic, analyse the dynamics of the epidemic and translate these findings into information that can be used to more effectively target interventions. PANGA encompasses four analysis themes: (1) molecular epidemiology and mathematical modelling, (2) phylodynamics, (3) mobility and migration, and (4) clinical science, drug resistance and ethics. Major variables within the dataset to answer this question are: (1) the genotype; (2) the participant's geolocation from which the genotype was derived; and (3) the date of sampling, (4) all relevant clinical data and (5) all relevant demographic data.

KIND OF DATA

Clinical and Demographic data

UNITS OF ANALYSIS

Each sequences derived from a single specimen, but some sequences may have been repeated on the same specimen, or on another time point, depending on the quality of the sequence.

TOPICS

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

KEYWORDS

HIV full-length sequences; date of sampling; phylodynamics, phylogeny, HIV-1, ART history

Coverage

GEOGRAPHIC COVERAGE

South Africa

UNIVERSE

HIV genome extracted and sequenced from participants infected with HIV

Producers and Sponsors

PRIMARY INVESTIGATOR(S)

Name	Affiliation
Pillay, Deenan	Africa Health Research Institute

OTHER PRODUCER(S)

Name	Affiliation	Role
Africa Health Research Institute		

FUNDING

Name	Abbreviation	Role
PANGA consortium, funded by the Bill & Melinda Gates Foundation	BMGF	Genotyping funding source

OTHER ACKNOWLEDGEMENTS

Name	Affiliation	Role
Derache, Anne	Africa Health Research Institute	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.Data.2019.v1

Sampling

Sampling Procedure

HIV positive individuals within the demographic surveillance area of the Africa Health Research Institute from 2010 to 2016. Full-length HIV deep sequencing was attempted on samples that:

A) Had a blood sample taken (not DBS)

B) Had a viral load >1000 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	2016-12-04	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

AHRI.PANGA1.ART.2019.v1

Content

Cases 71421

Variable(s) 7

Structure Type:
Keys: ()

Version

Producer

Missing Data

Variables

ID	Name	Label	Type	Format	Question
V116	Pangeald	Pangea unique Identifier of Individual	discrete	character	
V117	ARTStartDate	Date when antiretroviral therapy was initiated	discrete	character	
V118	ARTEncounterDate	ART encounter or visit date	discrete	character	
V119	ARTRegimen	ART regimen	discrete	character	
V120	ARTEndEventDate	Date of ART end event	discrete	character	
V121	ARTEndEventReasons	Reasons for ART end event occurring	discrete	character	
V122	ARTEndEventType	ARTEndEventType	discrete	numeric	

AHRI.PANGA1.INDIVIDUALS.2019.v1

Content

Cases 3890

Variable(s) 15

Structure Type:
Keys: ()

Version

Producer

Missing Data

Variables

ID	Name	Label	Type	Format	Question
V123	Pangeald	Pangea unique Identifier of Individual	discrete	character	
V124	SampleId	Sample identifier	discrete	character	
V125	SampleSource	A study a sample was taken from	discrete	numeric	
V126	Source		discrete	character	
V127	DoB	Individual's Date of Birth	discrete	character	
V128	Age	Individual's age	contin	numeric	
V129	Sex	Gender	discrete	numeric	
V130	SampleDate	Date when sample was taken	discrete	character	
V131	ACDIS_Id	Surveillance unique Identifier of Individual	contin	numeric	
V132	ARTEMIS_Id	ARTEMIS unique Identifier of Individual	contin	numeric	
V133	ACCDB_Id	Tier.Net unique Identifier of Individual	contin	numeric	
V134	TasP_Id	TasP unique Identifier of Individual	contin	numeric	
V135	LatestNegative	Date of last negative	discrete	character	
V136	EarliestPositive	Date of earliest positive	discrete	character	
V137	EarliestKnownART	Date of earliest known antiretroviral therapy	discrete	character	

AHRI.PANGA1.LAB.2019.v1

Content

Cases 19197

Variable(s) 6

Structure Type:
Keys: ()

Version

Producer

Missing Data

Variables

ID	Name	Label	Type	Format	Question
V138	Pangeald	Pangea unique Identifier of Individual	discrete	character	
V139	SampleId	Sample identifier	discrete	character	
V140	ResultDate	Date when sample was processed	discrete	character	
V141	CD4Count	CD4 Count value	contin	numeric	
V142	CD4Percentage	CD4 Percentage value	contin	numeric	
V143	Virallload	Viral load value	contin	numeric	

Pangea unique Identifier of Individual (PangeaId)

File: AHRI.PANGEA1.ART.2019.v1

Overview

Type: Discrete	Valid cases: 71421
Format: character	Invalid: 0
Width: 20	

Date when antiretroviral therapy was initiated (ARTStartDate)

File: AHRI.PANGEA1.ART.2019.v1

Overview

Type: Discrete	Valid cases: 70901
Format: character	Minimum: NaN
Width: 11	Maximum: NaN

ART encounter or visit date (ARTEncounterDate)

File: AHRI.PANGEA1.ART.2019.v1

Overview

Type: Discrete	Valid cases: 70901
Format: character	Minimum: NaN
Width: 11	Maximum: NaN

ART regimen (ARTRegimen)

File: AHRI.PANGEA1.ART.2019.v1

Overview

Type: Discrete	Valid cases: 63949
Format: character	
Width: 244	

Date of ART end event (ARTEndEventDate)

File: AHRI.PANGEA1.ART.2019.v1

Overview

Type: Discrete	Valid cases: 6440
Format: character	Minimum: NaN
Width: 11	Maximum: NaN

Reasons for ART end event occurring (ARTEndEventReasons)

File: AHRI.PANGEA1.ART.2019.v1

Overview

Type: Discrete	Valid cases: 93
Format: character	
Width: 244	

ARTEndEventType (ARTEndEventType)

File: AHRI.PANGA1.ART.2019.v1

Overview

Type: Discrete
Format: numeric
Width: 12
Decimals: 0
Range: 1-5

Valid cases: 5874
Invalid: 65547

Pangea unique Identifier of Individual (PangeaId)

File: AHRI.PANGEA1.INDIVIDUALS.2019.v1

Overview

Type: Discrete

Valid cases: 3890

Format: character

Width: 244

Sample identifier (SampleId)

File: AHRI.PANGEA1.INDIVIDUALS.2019.v1

Overview

Type: Discrete

Valid cases: 3890

Format: character

Width: 244

A study a sample was taken from (SampleSource)

File: AHRI.PANGEA1.INDIVIDUALS.2019.v1

Overview

Type: Discrete

Valid cases: 3890

Format: numeric

Invalid: 0

Width: 12

Decimals: 0

Range: 1-4

(Source)

File: AHRI.PANGEA1.INDIVIDUALS.2019.v1

Overview

Type: Discrete

Valid cases: 3890

Format: character

Width: 244

Individual's Date of Birth (DoB)

File: AHRI.PANGEA1.INDIVIDUALS.2019.v1

Overview

Type: Discrete

Valid cases: 3890

Format: character

Minimum: NaN

Width: 11

Maximum: NaN

Individual's age (Age)

File: AHRI.PANGEA1.INDIVIDUALS.2019.v1

Overview

Individual's age (Age)

File: AHRI.PANGA1.INDIVIDUALS.2019.v1

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

Valid cases: 3890
 Invalid: 0
 Minimum: 14
 Maximum: 114
 Mean: 36.4
 Standard deviation: 12.3

Gender (Sex)

File: AHRI.PANGA1.INDIVIDUALS.2019.v1

Overview

Type: Discrete
 Format: numeric
 Width: 12
 Decimals: 0
 Range: 1-9

Valid cases: 3889
 Invalid: 1

Date when sample was taken (SampleDate)

File: AHRI.PANGA1.INDIVIDUALS.2019.v1

Overview

Type: Discrete
 Format: character
 Width: 11

Valid cases: 3890
 Minimum: NaN
 Maximum: NaN

Surveillance unique Identifier of Individual (ACDIS_Id)

File: AHRI.PANGA1.INDIVIDUALS.2019.v1

Overview

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

Valid cases: 2736
 Invalid: 1154
 Minimum: 84
 Maximum: 218561

ARTEMIS unique Identifier of Individual (ARTEMIS_Id)

File: AHRI.PANGA1.INDIVIDUALS.2019.v1

Overview

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

Valid cases: 1343
 Invalid: 2547
 Minimum: 41
 Maximum: 200695

Tier.Net unique Identifier of Individual (ACCDB_Id)

File: AHRI.PANGA1.INDIVIDUALS.2019.v1

Overview

Tier.Net unique Identifier of Individual (ACCDB_Id)

File: AHRI.PANGA1.INDIVIDUALS.2019.v1

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

Valid cases: 2574
 Invalid: 1316
 Minimum: 4
 Maximum: 56761

TasP unique Identifier of Individual (TasP_Id)

File: AHRI.PANGA1.INDIVIDUALS.2019.v1

Overview

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

Valid cases: 2538
 Invalid: 1352
 Minimum: 9
 Maximum: 100102
 Mean: 11738.4
 Standard deviation: 8344.1

Date of last negative (LatestNegative)

File: AHRI.PANGA1.INDIVIDUALS.2019.v1

Overview

Type: Discrete
 Format: character
 Width: 11

Valid cases: 304
 Minimum: NaN
 Maximum: NaN

Date of earliest positive (EarliestPositive)

File: AHRI.PANGA1.INDIVIDUALS.2019.v1

Overview

Type: Discrete
 Format: character
 Width: 11

Valid cases: 3879
 Minimum: NaN
 Maximum: NaN

Date of earliest known antiretroviral therapy (EarliestKnownART)

File: AHRI.PANGA1.INDIVIDUALS.2019.v1

Overview

Type: Discrete
 Format: character
 Width: 11

Valid cases: 3585
 Minimum: NaN
 Maximum: NaN

Pangea unique Identifier of Individual (PangeaId)

File: AHRI.PANGEA1.LAB.2019.v1

Overview

Type: Discrete	Valid cases: 19197
Format: character	Invalid: 0
Width: 20	

Sample identifier (SampleId)

File: AHRI.PANGEA1.LAB.2019.v1

Overview

Type: Discrete	Valid cases: 19197
Format: character	Invalid: 0
Width: 20	

Date when sample was processed (ResultDate)

File: AHRI.PANGEA1.LAB.2019.v1

Overview

Type: Discrete	Valid cases: 19197
Format: character	Minimum: NaN
Width: 11	Maximum: NaN

CD4 Count value (CD4Count)

File: AHRI.PANGEA1.LAB.2019.v1

Overview

Type: Continuous	Valid cases: 13822
Format: numeric	Invalid: 5375
Width: 10	Minimum: 0
Decimals: 0	Maximum: 4495
Range: 0-4495	
Invalid: 10001	

CD4 Percentage value (CD4Percentage)

File: AHRI.PANGEA1.LAB.2019.v1

Overview

Type: Continuous	Valid cases: 4934
Format: numeric	Invalid: 14263
Width: 10	Minimum: 0
Decimals: 0	Maximum: 66
Range: 0-66	
Invalid: 101	

Viral load value (Viralload)

File: AHRI.PANGEA1.LAB.2019.v1

Overview

Viral load value (Viralload)

File: AHRI.PANGEA1.LAB.2019.v1

Type: Continuous
Format: numeric
Width: 10
Decimals: 0
Range: 0-11000000
Invalid: 100000001

Valid cases: 15189
Invalid: 4008
Minimum: 0
Maximum: 11000000

Documentation

Technical documents

Documentation:PANGA1 Demographic and Clinical Data

Title Documentation:PANGA1 Demographic and Clinical Data
Author(s) Sweetness H Dube
Date 18/04/2019
Country South Africa
Language English
Filename PANGA1 Demographic and Clinical Data.pdf

DDI:PANGA1.Data.2019.v1

Title DDI:PANGA1.Data.2019.v1
Author(s) Sweetness H Dube
Date 18/04/2019
Country South Africa
Language English
Filename AHRI.PANGA1.Data.2019.v1.pdf
