

PlasmoVis



GENOMIC VARIANT BROWSER

PlasmoVis is a user-friendly web-based visualisation tool to assist with the analysis and visualisation of sequencing data, specifically to interrogate the genomic variation of *Plasmodium malariae* parasites.

Table of Contents

- [About](#)
- [Prerequisites](#)
- [Setup](#)
 - [Download PlasmoVis](#)
 - [Install PlasmoVis Dependencies](#)
 - [Run PlasmoVis](#)
 - [Close PlasmoVis](#)
- [PlasmoVis](#)
 - [Home Page](#)
 - [Navigation Bar](#)
 - [Genomic Variant Browser](#)
 - [Overview](#)
 - [Tacks](#)
 - [Zooming Functionality](#)
 - [Scrolling](#)
 - [Input Specific Chromosome Positions](#)
 - [Cursor Guides](#)
 - [Track Settings](#)
 - [Track Disposition](#)
 - [Gene Search](#)
 - [Venn Diagram](#)
 - [Data](#)
 - [Sample Charts](#)
 - [Sample Search](#)
 - [About](#)
- [Author](#)

Prerequisites

PlasmoVis is compatible for being run either on Mac or Windows Operating Systems. Please ensure you have **Node.js** installed on your computer before proceeding.

- [Node.js](#)

Setup

1. Download PlasmoVis

PlasmoVis can be downloaded in two ways:

- Click on the green **Code** button on the upper-right corner above and select **Download ZIP** (**Figure 1**). Remember to unzip the file before proceeding.

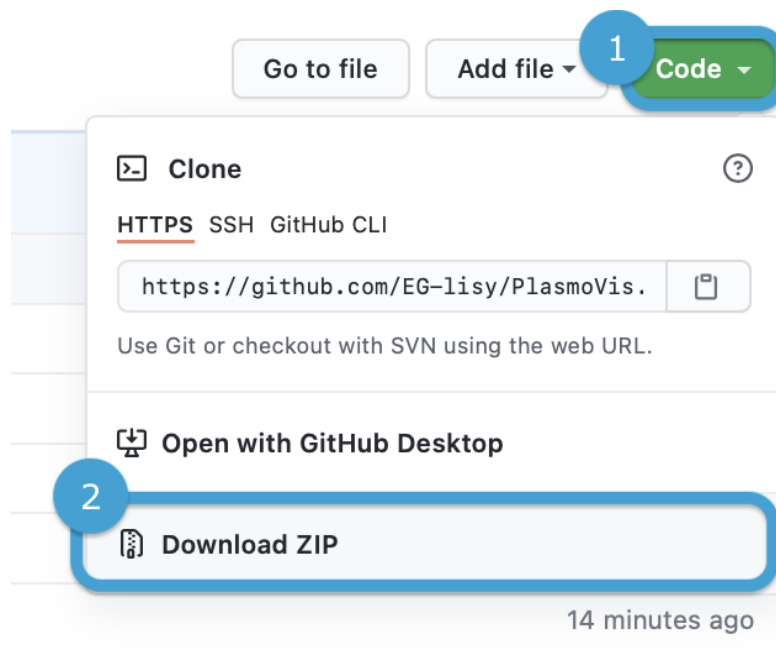


Figure 1. Download PlasmoVis - GitHub screenshot.

- Alternatively, in case you have [Git](#) installed on your computer, run the following commands on your terminal:

```
# set your working directory (i.e. folder in which you would like to save
PlasmoVis)
cd <yourdir/yourfolder>
# clone the repository
git clone https://github.com/EG-lisy/PlasmoVis.git
```

2. Install PlasmoVis Dependencies

Proceed installing PlasmoVis dependencies (node modules).

1. From the terminal, set your working directory inside **PlasmoVis/PlasmoVis**. Please note that the project subfolder has the same name of the main one.

```
# set your working directory (i.e. folder in which you would like to save
PlasmoVis)
cd <yourworkingdirectory>/PlasmoVis/PlasmoVis
# install dependencies
npm install
```

Once the installation is completed, a **node_modules** folder containing all the required dependencies will be created inside **PlasmoVis/PlasmoVis**.

3. Run PlasmoVis

1. Without changing the working directory, run the following command from the terminal:

```
node app.js
```

Note for Developers

Nodemon has been installed to allow changes being automatically updated on the server.
If you wish to edit the code and run PlasmoVis in the developer mode, run **npm start** instead.

2. If all the steps have been followed correctly, the following welcoming message will show up on the console:

```
MacBook-Pro-di-Elisabetta:PlasmoVis lisy$ node app.js

Server is listening on port: 3000
----- PlasmoVis -----
-
Welcome to PlasmoVis!
1. Open your browser of choice
2. Visit: http://localhost:3000
```

3. You will be now able to visit PlasmoVis on <http://localhost:3000>.

Note: to stop PlasmoVis from running, type **Ctrl+C** on your terminal.

PlasmoVis

This section aims to assist with the navigation of PlasmoVis web-pages.

Home Page

From <http://localhost:3000> you will end up on the landing page of PlasmoVis (**Figure 2**).

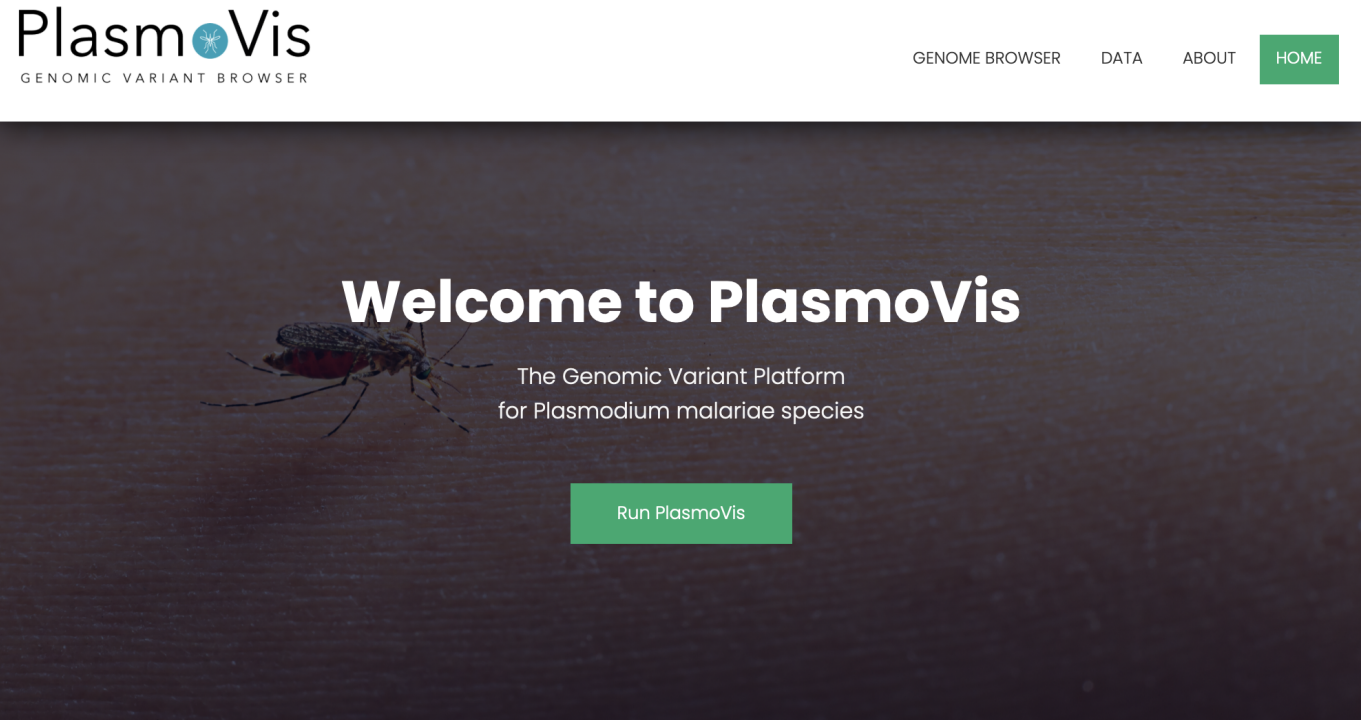


Figure 2. PlasmoVis Landing Page.

Bottom-Section Navigation

Below the welcoming message, self-explanatory clickable links can be found (**Figure 3**).

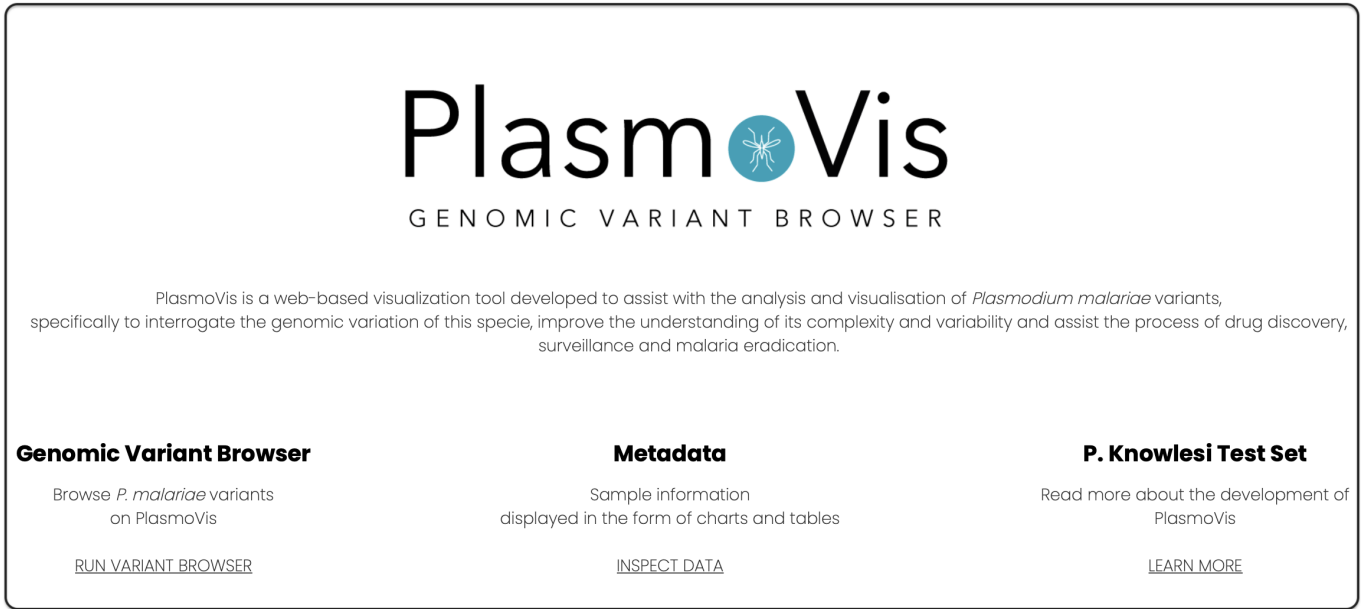


Figure 3. PlasmoVis bottom-section navigation links.

P. malariae variants can be inspected on the [Genomic Variant Browser](#) page, either by clicking on the green **Run PlasmoVis** button (refer to **Figure 2**) or by using the navigation bar (see below).

Navigation bar

The navigation bar allows switching in between pages (**Figure 4**).



Figure 4.PlasmoVis Navigation Bar

Where:

1. **PlasmoVis logo** takes back to the [home page](#)
2. **GENOME BROWSER** takes to the [Genomic Variant Browser](#) page
3. **DATA** takes to the [Data](#) page
4. **ABOUT** includes additional information on *P. knowlesi* data used whilst developing PlasmoVis
5. **HOME** takes back to [home page](#)

Genomic Variant Browser

This page allows to inspect *P. malariae* variants over an [IGV.js](#) framework(**Figure 5**).

Overview

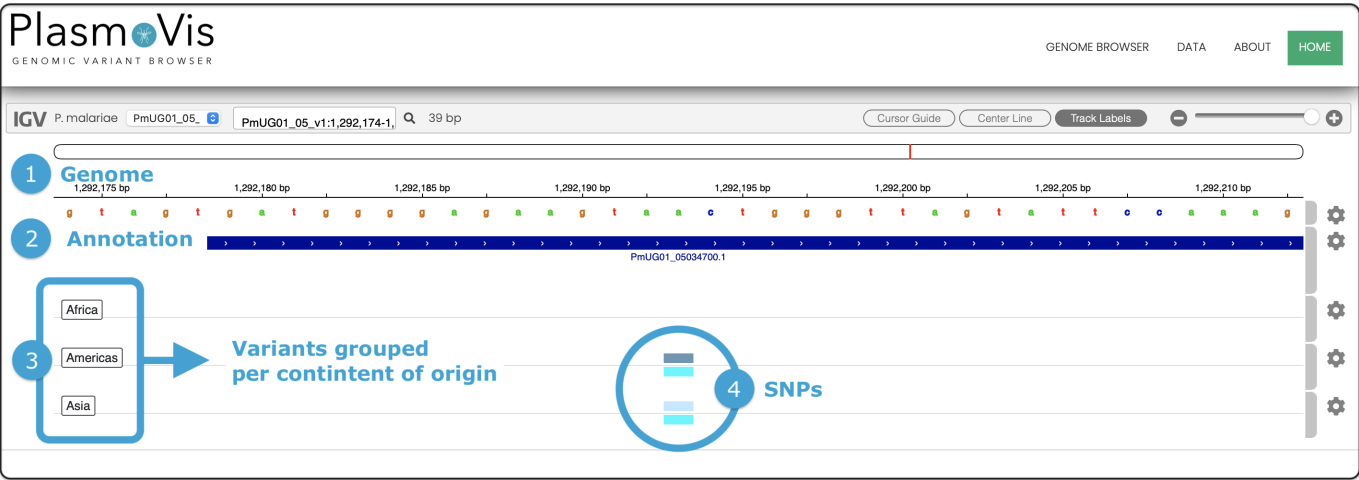


Figure 5. Screenshot of the Genomic Variant Browser

Where:

1. *Plasmodium Malariae* Genome ([PmUG01](#))
2. Annotation (gene IDs)
3. Intersected samples variants based on continent of origin (Africa, Americas and Asia)
4. SNPs (Single Nucleotide Polymorphisms)

Tracks

Genome, annotation and variants tracks are all interactive.

By clicking on a specific gene track, an info box will pop up displaying the gene ID/parent ID, which can be copied to the clipboard to identify the name of the gene using the Gene Search table found at the bottom of the genomic variant browser (see [Gene Search](#) section).

SNPs tracks display two subtracks. The upper subtracks are coloured in different shades of blue based on the continent of origin; by clicking on it, an info box will pop up displaying the information stored inside the VCF file. Likewise, the lower subtrack will pop up an info box displaying genotype and additional allelic information.

See **Figure 6** below for a better understanding.

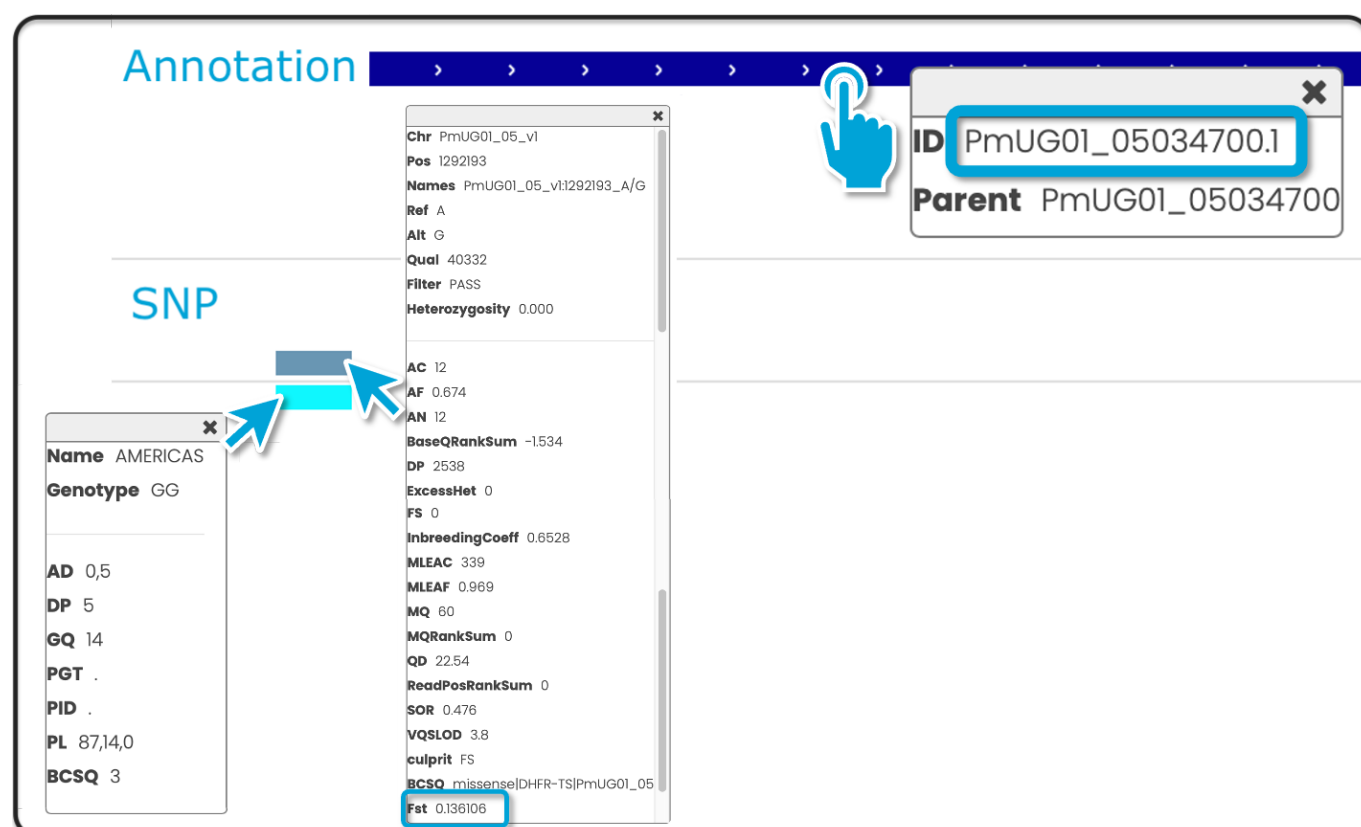


Figure 6. Tracks Info Boxes

Upper subtrack:

CHR chromosome name

Pos SNP position

Names SNP name added during the annotation pipeline used to calculate fixation indices (R script)

Ref reference allele

ALT alternative allele

Qual a phred-scaled quality score assigned by the variant caller

Filter PASS if specific position has passed all given filters when generating the vcf file

AC allele count in genotypes, for each ALT allele, in the same order as listed

AF allele Frequency, for each ALT allele, in the same order as listed

AN total number of alleles in called genotypes

BaseQRankSum z-score from Wilcoxon rank sum test of Alt Vs. Ref base qualities

DP approximate read depth; some reads may have been filtered

ExcessHet phred-scaled p-value for exact test of excess heterozygosity

FS phred-scaled p-value using Fisher's exact test to detect strand bias

InbreedingCoeff inbreeding coefficient as estimated from the genotype likelihoods per-sample when compared against the Hardy-Weinberg expectation

MLEAC maximum likelihood expectation (MLE) for the allele counts (not necessarily the same as the AC), for each ALT allele, in the same order as listed

MLEAF maximum likelihood expectation (MLE) for the allele frequency (not necessarily the same as the AF), for each ALT allele, in the same order as listed

MQ RMS (root mean square) Mapping Quality

MQRankSum z-score From Wilcoxon rank sum test of Alt vs. Ref read mapping qualities

QD variant Confidence/Quality by Depth

ReadPosRankSum z-score from Wilcoxon rank sum test of Alt vs. Ref read position bias

SOR symmetric Odds Ratio of 2x2 contingency table to detect strand bias

VQSLOD log odds of being a true variant versus being false under the trained gaussian mixture model

culprit the annotation which was the worst performing in the Gaussian mixture model, explains the reason why the variant was filtered out (e.g. FisherStrand (FS), QualByDepth (QD), StrandOddsRatio (SOR), RMSMappingQuality (MQ), MappingQualityRankSumTest (MQRankSum), ReadPosRankSumTest (ReadPosRankSum)...)

BCSQ haplotype-aware consequence annotation from BCFtools/csq, see

<http://samtools.github.io/bcftools/howtos/csq-calling.html> for details. Format:

Consequence|gene|transcript|biotype|strand|amino_acid_change|dna_change

Fst Fixation index (range 0-1)

Lower subtrack:

Name sample name (continent)

Genotype genotype info. Please note that Plasmodium species are haploid. This diploid output is due to the GATK pipeline

DP approximate read depth (reads with MQ=255 or with bad mates are filtered)

GQ genotype quality

PGT physical phasing haplotype information, describing how the alternate alleles are phased in relation to one another; will always be heterozygous and is not intended to describe called alleles

PID physical phasing ID information, where each unique ID within a given sample (but not across samples) connects records within a phasing group

PL phred-scaled likelihoods for genotypes as defined in the VCF specification

BCSQ Haplotype-aware consequence value

Zooming Functionality

You can zoom in and out of the genomic variant browser either by using the zooming bar found on the upper-right corner of the genome browser (**Figure 7**)

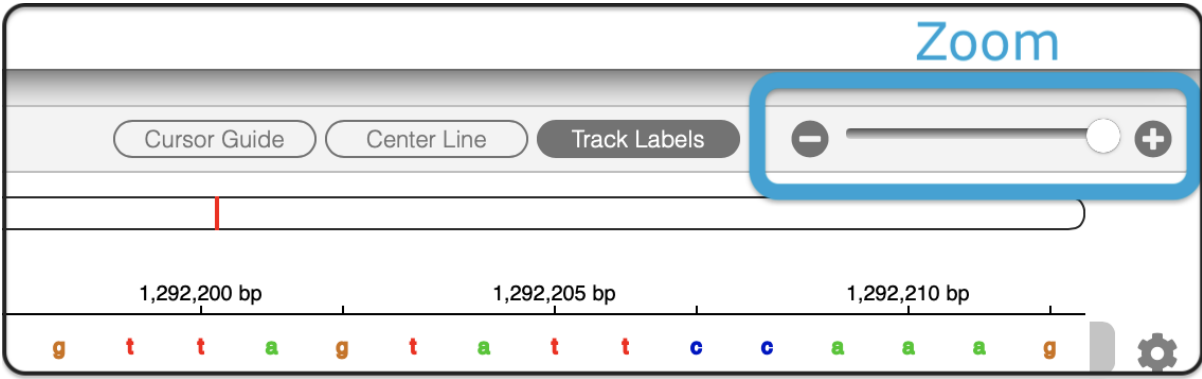


Figure 7. Zooming Bar

or by clicking, dragging and dropping over the genome range of interest (Figure 8).

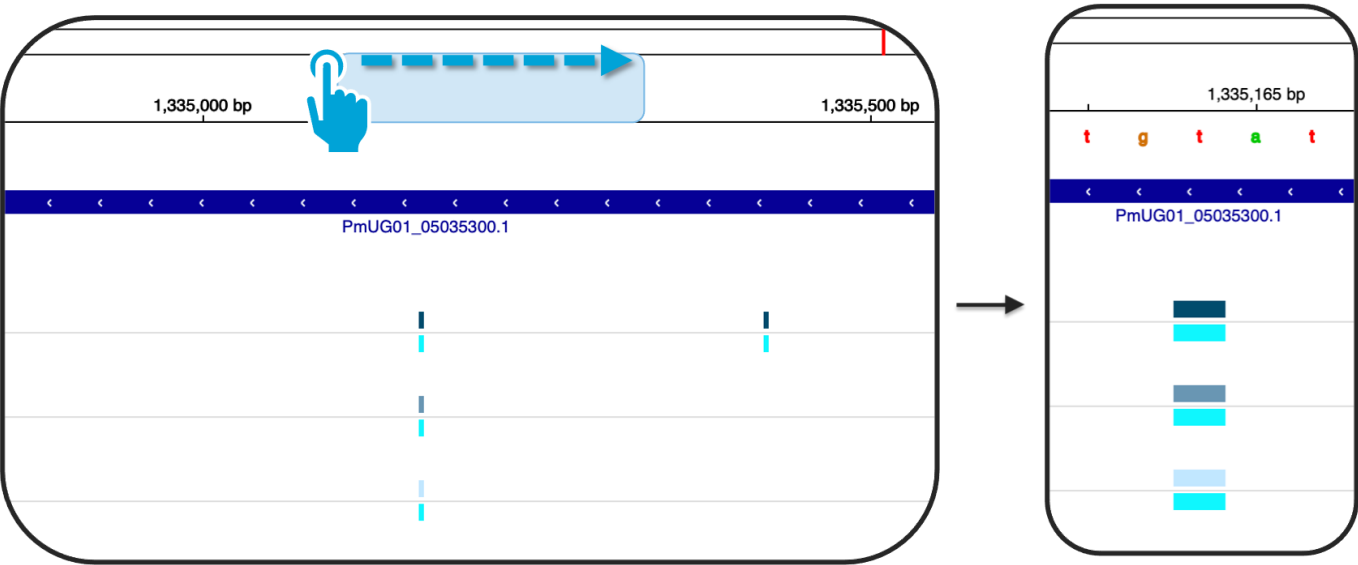


Figure 8. Drag & drop zooming functionality

Scrolling

To manually move along the genome, drag and drop over the annotation/variant tracks section (refer to Figure 5, bulletpoints 2 and 3).
Alternatively, to input a specific location refer to the [section below](#).

Input Specific Chromosome Positions

Chromosomes can be selected using the drop-down menu found on the upper-left corner of the genomic variant browser (Figure 9).

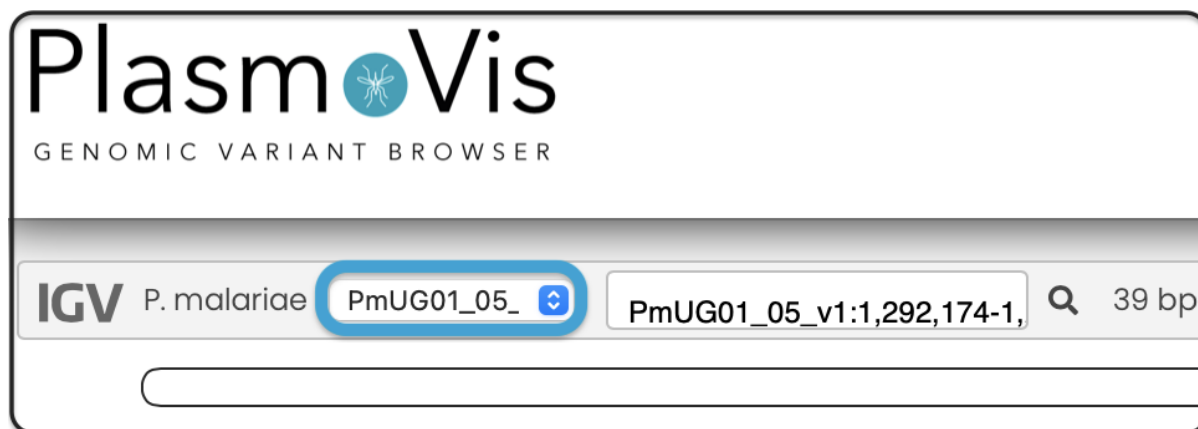


Figure 9. Chromosome Selection

To avoid manually [scrolling](#) to a position of interest, genomic coordinates can be directly pasted in the search box found on the upper-left corner of the genomic variant browser, opposite the chromosome selection (**Figure 10**).



Figure 10. Genomic Coordinates Search Box

Taking as an example `PmUG01_05_v1:1,335,145-1,335,183`:

- `PmUG01_05_v1` refers to the chromosome of interest
- `:` precedes the chromosome coordinates
- `1,335,145-1,335,183` refers to the chromosome coordinates (in this case, from `1,335,145` to `1,335,183`)

Note that this will also work when specifying just the start position of interest.

Cursor Guides

Visualisation guides are also included in the IGV framework (**Figure 11**).

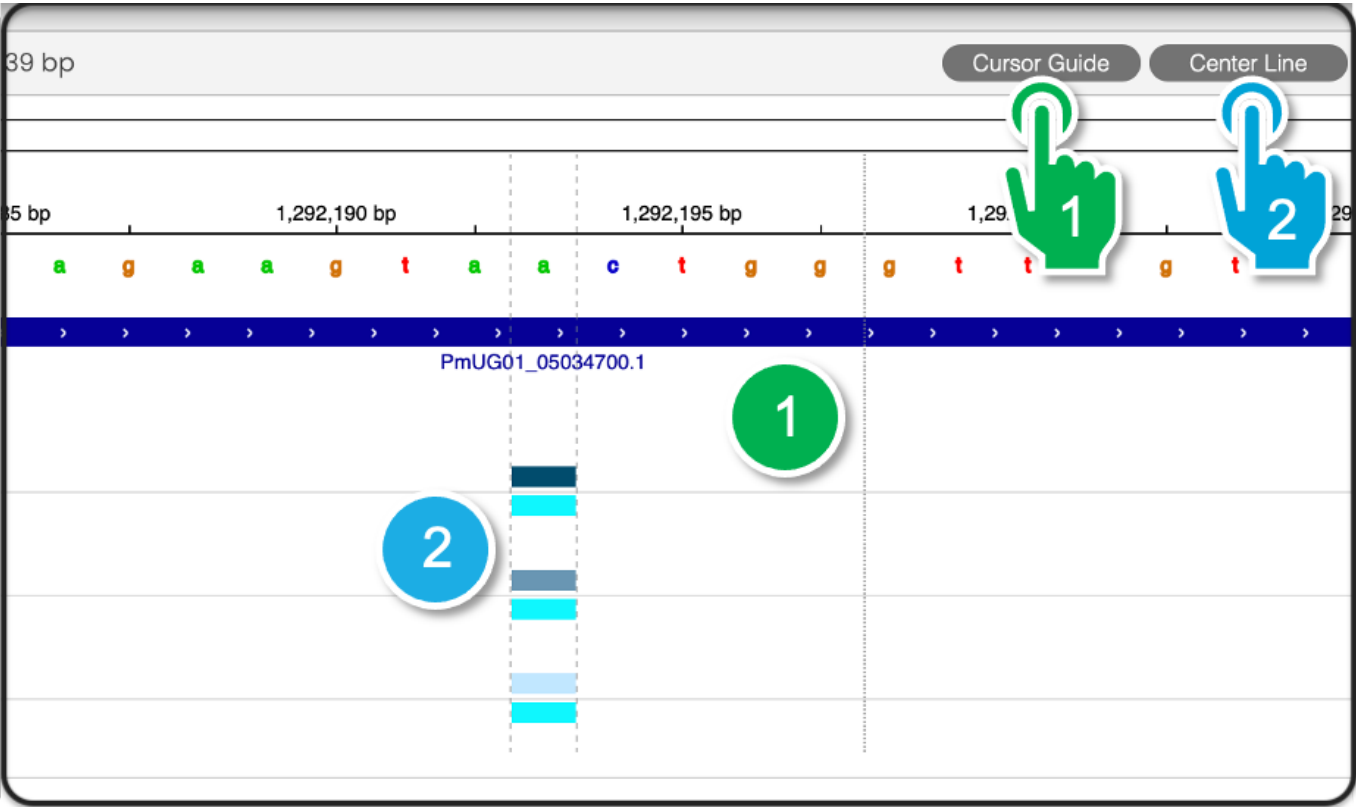


Figure 11. Visualisation Guides

Where:

- 1. **Cursor Guide** will display a guide which will follow the movements of your cursor
- 2. **Center Line** will display two guides in the middle of the Genomic Variant Browser

Track Settings

Track labels can be hidden/shown using the **Track Labels** button found on the upper section next to the cursor guides buttons (**Figure 12**).



Figure 12. Track Labels button

Tracks can be also customised using the grey gears found on the right-hand side of the genomic variant browser (**Figure 13**).



Figure 13. Track Settings

Where:

1. Allows setting different track names
2. Allows setting a different track height
3. Allows setting a different track colour
4. Allows setting the Collapsed view
5. Allows setting the Squished view
6. Allows setting the Expanded view (set by default)
7. Allows setting the visibility window (window height)
8. Removes the track

Note: all the above options are reversible, except point **8**. If you remove a track by mistake you will need to refresh the page.

Track Disposition

The disposition of the tracks can be modified by dragging and dropping the dark-grey track bars on the right-hand-side (**Figure 14**).

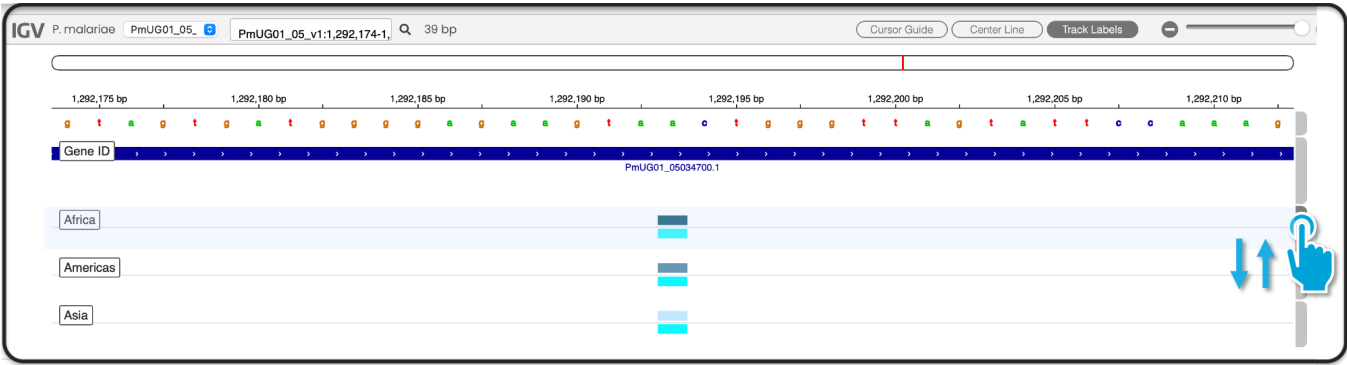


Figure 14. Track Dispositions

Gene Search

The **GENOME BROWSER** page includes a **Gene Search** section (**Figure 15**) from which both gene names and gene IDs can be searched.

Gene Name Search

Reported gene names can be inspected directly on [Plasmodb](#)
Please note that not all gene IDs have a corresponding gene name

Show

10

 entries

Search:

Gene ID	Chromosome	Start	End	Gene Name
PmUG01_00010200	PmUG01_00_v1	10901	11796	
PmUG01_00010300	PmUG01_00_v1	17599	19530	
PmUG01_00010400	PmUG01_00_v1	28779	29762	
PmUG01_00010500	PmUG01_00_v1	33277	33872	
PmUG01_00010700	PmUG01_00_v1	45868	46753	
PmUG01_00010800	PmUG01_00_v1	51200	52191	
PmUG01_00010900	PmUG01_00_v1	62038	63052	
PmUG01_00011000	PmUG01_00_v1	66335	67225	
PmUG01_00011200	PmUG01_00_v1	75118	77693	
PmUG01_00011300	PmUG01_00_v1	83884	84763	

Showing 1 to 10 of 6,078 entries

Previous

1

2

3

4

5

...

608

Next

Figure 15. Gene Search Table

12 / 16

Note: columns can be sorted in ascending/descending order by clicking on the header title of interest(**Gene ID**, **Chromosome**, **Start**, **End**, **Gene Name**).
Up to 100 entries can be displayed at the same time.

To obtain more information about a specific gene of interest, a direct link to the official database of *Plasmodium* parasites (**PlasmoDB**) is also included.

Venn Diagram

An interactive venn diagram is included at the bottom of the **GENOME BROWSER** page, displaying the total number of unique/shared SNPs between groups (**Figure 16.**).

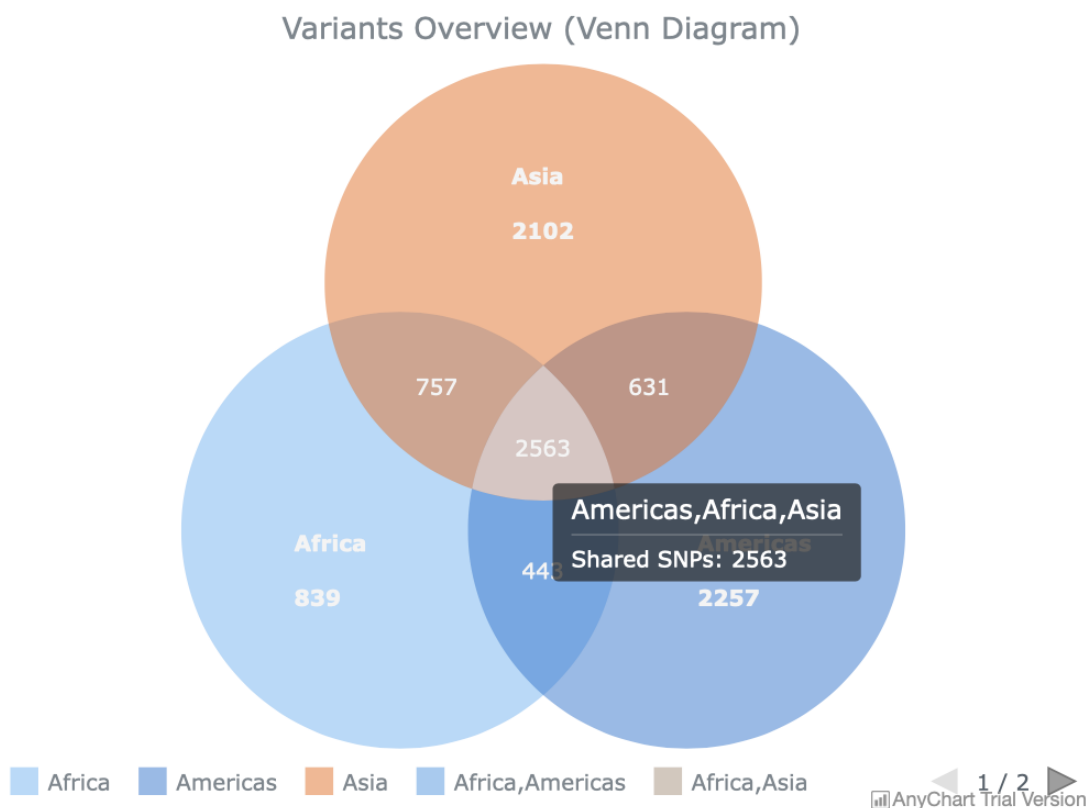


Figure 16. Interactive Venn Diagram

Data

This page allows to visually inspect sample information in interactive charts and tables. An interactive choropleth map is also available, displaying sample counts based on their country of origin (**Figure 17.**).

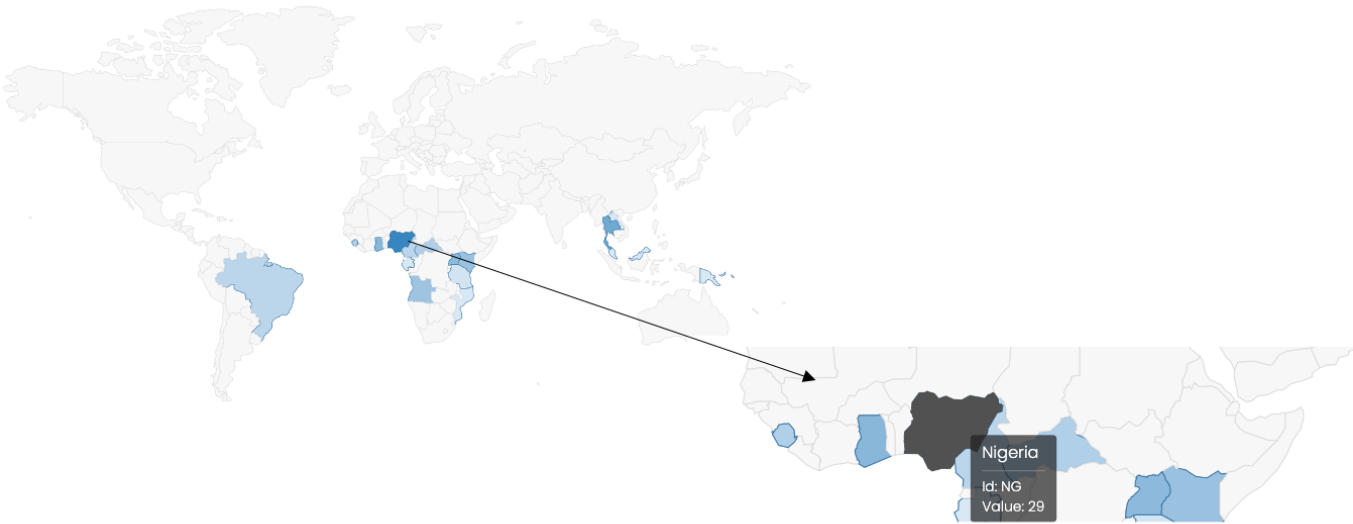


Figure 17. Cloropleth Map based on sample counts

Sample Charts

Interactive sample-charts displaying sample counts based on country and continent of origin are also included (Figure 18).

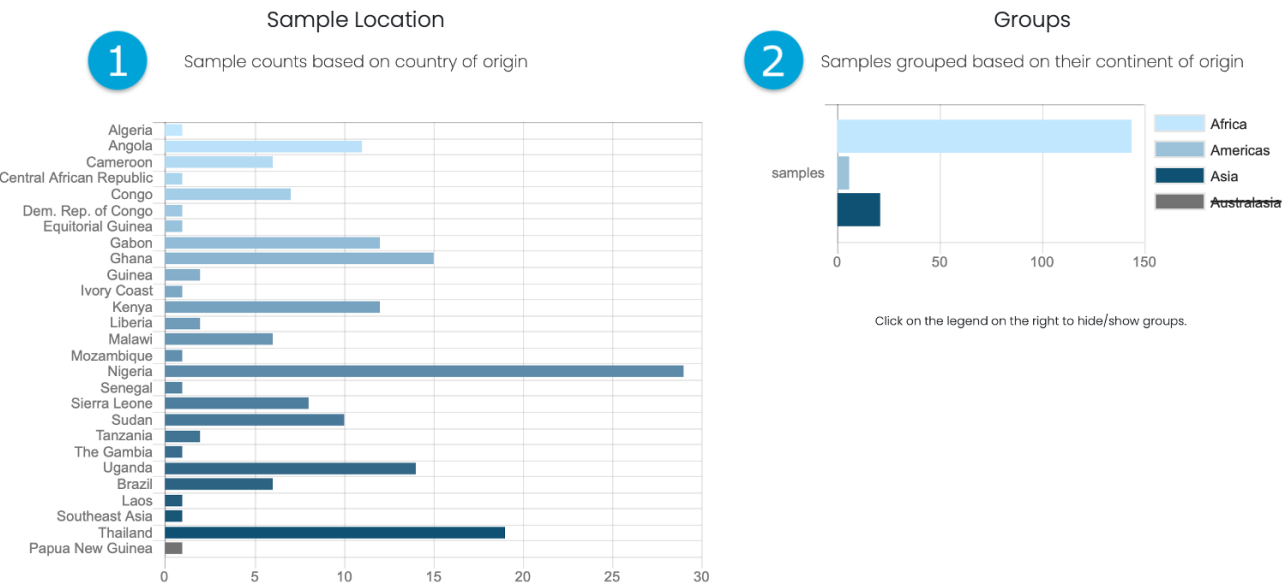


Figure 17. Interactive Sample Charts

Note: sample counts will display by hoovering over the bar charts.

Sample Search

A Sample Search Table is included at the bottom of this page, from which sample ID/location information can be retrived (Figure 18).

Sample Search

Samples can be searched by ID, location or year using the search bar below

Show entries

Search:

Sample	Country	Continent	Year
PM0015_Ishtm	Angola	Africa	NA
PM_AGO_001	Angola	Africa	NA
PM_AGO_002	Angola	Africa	NA
PM_AGO_003	Angola	Africa	NA
PM_AGO_004_com	Angola	Africa	2014
PM_AGO_006	Angola	Africa	2001
PM_AGO_007	Angola	Africa	2005
PM_AGO_009	Angola	Africa	2019
PM_AGO_010	Angola	Africa	2019
PM_AGO_012	Angola	Africa	2017

Showing 1 to 10 of 172 entries

[Previous](#) [1](#) [2](#) [3](#) [4](#) [5](#) [...](#) [18](#) [Next](#)

Figure 18. Interactive Sample Charts

About

This page includes additional information about the development of PlasmoVis. Direct links to the *P. knowlesi* dataset initially tested on PlasmoVis are also included (**Figure 19**).

PlasmoVis

GENOMIC VARIANT BROWSER

PlasmoVis is a web-based visualization tool developed to assist with the analysis and visualisation of *Plasmodium malariae* variants. PlasmoVis specifically aims to interrogate the genomic variation of *P. malariae* parasites, improve the understanding of the complexity and variability of this species and assist the process of drug discovery, surveillance and malaria eradication.

PlasmoVis was initially built working with a subset of *P. knowlesi* samples, which can be viewed using the links below.

P. knowlesi (v1)

Comparison of a subset of 9 samples.
Note that this version runs slow.

[RUN GENOME BROWSER](#)

P. knowlesi (v2)

Comparison of intersected *P. knowlesi* samples.
This version was adapted for *P. malariae* dataset.

[RUN GENOME BROWSER](#)

P. knowlesi Data

P. knowlesi sample information
displayed in the form of charts and tables.

[INSPECT DATA](#)

Figure 19. About section

Note: *P. malariae* samples were intersected in a different way compared to *P. knowlesi*. Be aware that the visualisation of *P. knowlesi* samples might appear different.

Sequencing Files

THIS SECTION MIGHT NOT BE NEEDED

Download the sequencing files running the following command

Author

Developed by Elisabetta Galatola

Applied Bioinformatics MSc
Cranfield University – Cranfield, Bedford, UK
Academic Year: 2020–2021