mapDATAge instructions

Table of Contents

About mapDATAge 1

Installing R and R dependencies 1

Running mapDATAge 2

Description of Example Files 2

General Instructions 3

- Input Files 3
- Panels 6
- PANEL1: DATA 7
- PANEL2: SMAP 7
- PANEL3: AMAP 9
- PANEL4: TRAJECTORY 12
- PANEL5: ANCESTRY 14
- PANEL6: PCA 16
- PANEL7: MULTIPLESNPS 17
- PANEL8: HAPLO 18
- PANEL9: ONECLICK 19

References 22

About mapDATAge

The mapDATAge package is designed to explore the presence of geographic and temporal patterns in ancient DNA data. It takes simple tabulated text files as input and contains different panels to draw a number of maps and plots that are common to most ancient DNA studies, including:
(1) the spatial and temporal distributions of a given set of samples (SMAP) or alleles (AMAP);
(2) temporal allelic trajectories (TRAJECTORY);
(3) maps of individual ancestry profiles (ANCESTRY);
(4) PCA (or MDS) and related maps of spatio-temporal distributions (PCA);
(5) spatio-temporal distributions of alleles at one or multiple loci (MULTIPLESNPS);
(6) maps of (sub)haplogroups (HAPLO), and;
(7) finally, options are provided to automatically generate figures, following the selection of a preselected range of spatial and temporal parameters, which can be useful to contrast data from different loci and/or species (ONECLICK).

mapDATAge is implemented and maintained in ShinyR by Dr Xuexue Liu at the Centre for Anthropobiology and Genomics of Toulouse (CAGT) (https://cagt.cnrs.fr/).
The underlying code is available at [https://github.com/xuefenfei712/mapDATAge].
Please cite the following publication [Xuexue Liu and Ludovic Orlando (2022). mapDATAge: a ShinyR package to chart ancient DNA data through space and time. Bioinformatics XX:XX-XX]
Please contact Dr Xuexue Liu (xuexue.liu@univ-tlse3.fr), if you have any questions, or suggestions for improvement.

Installing R and R dependencies

The mapDATAge package was developed using the R programming language, which thus must be properly downloaded and installed from the Comprehensive R Archive Network (CRAN) (www.r-project.org), before running mapDATAge. For a better user experience, we recommend downloading and installing RStudio desktop, available at www.rstudio.com. The mapDATAge package depends on a number of R libraries and requires the R version 4.1.2 or above. Once the R environment is installed, the various dependencies can be installed using the simple following commands:

```
packages=c("plotly","shiny","shinyFiles","leaflet","shinythemes","RColorBrewe
r","esquisse","scales","ggplot2","plotly","markdown","leaflet.minicharts","ht
mltools","leaflegend","sf","sp","stringi","leaflet.extras","dplyr","rcolors",
"DT","forcats","foreach","htmlwidgets")

install.packages(packages, repo="http://cran.rstudio.org", dependencies=TRUE)

install.packages("remotes")

remotes::install_github("thomasp85/shinyFiles", upgrade = "never")
```

Running mapDATAge

We provide a user interface layout file (ui.R) at github, as well as all of the related code (/code/ subfolder), and all markdown files (/ms/ subfolder). The mapDATAge package can be directly obtained from github by clicking on the green "*Code*" download button and choosing the *download ZIP* option. The resulting .zip file will contain the entire repository content and can be unzipped on your local computer. Alternatively, users can automatically obtain mapDATAge from the github repository by running:

```
git clone https://github.com/xuefenfei712/mapDATAge.git
```

on a terminal. Once installed, the mapDATAge package can be run by defining a variable path indicated the full path to the mapDATAge folder (e.g. 'path=C://MapR//mapDATAge', which contains the ui.R and Server.R) and then typing 'shiny::runApp(appDir = path)'. This will open a graphical interface, which provides all the different menus provided as part of the mapDATAge package. Local installation is preferred. However, if you want to run mapDATAge from a URL or Github, please use the command below either in R or Rstudio.

Through URL

```
shiny::runUrl("https://github.com/xuefenfei712/mapDATAge/archive/refs/heads/main.zip")
```

Through Github

```
shiny::runGitHub("mapDATAge","xuefenfei712", ref="main")
```

Please note that these two commands will download the code of mapDATAge to a temporary folder on your computer and launch the mapDATAge in your default web browser. Once the web browser is closed, the downloaded code is deleted.

Description of Example Files

To illustrate the range of options provided by mapDATAge, we have prepared four example files providing previously-published geolocated and time-stamped ancient DNA data. The first data set (mapDATAge-rs4988235_MCM6.txt) was obtained from the Allen Ancient DNA Resources for human data, (https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data). We restricted the available data set to the individual genotype provided at rs4988235, a locus involved in lactose tolerance in Europeans (Anguita-Ruiz, et al., 2020), as well as mitochondrial and Y chromosome haplotypes. The numeric allele was coded from the genotype data, e.g. a sample showing a 'C/T' genotype at a given locus was recoded with 1 for C and T, and 0 for A and G; similarly, 'C/C' genotypes were recoded with 2 for C, and 0 for A, G, and T. We filtered out samples with missing alleles. The code "ped2Table.R" is provided to generate the input file. Overall, we considered a total 2,120 ancient and modern Europeans dated to the last 10,000 years, disregarding individuals with missing genotypes for efficient processing. The purpose of this first example file was to illustrate the flexibility of six menus provided by mapDATAge package (namely, SMAP, AMAP, TRAJECTORY, MULTIPLESNPS, HAPLO and ONECLICK).

The second data set consists of three individual files (Horse-Anc.txt, PCA.eigenval.txt, and PCA.eigenvec.txt), which include the individual genetic ancestry profiles for 271 ancient horses characterized by Librado and colleages (Librado, et al., 2021). It illustrates how the mapDATAge ANCESTRY panel can be used to replicated one of their original figures in just a few clicks (Figure 2, https://www.nature.com/articles/s41586-021-04018-9/figures/2). This dataset also includes the results from their PCA analysis so as to illustrate the features of the PCA panel. Together, these different files can be used to illustrate the features of the SMAP, ANCESTRY and/or PCA panels.

We also provide a third data set, which consists of four additional example files (mapDATAge-chr23.22391254_DMRT3.txt, mapDATAge-chr8.20644555_TBX3.txt, mapDATAge-chr9.73114851_GSDMC.txt and HorseColor-combination.txt). These files provide allele counts at 9 loci causative for locomotory (*DMRT3*, (Andersson, et al., 2012)), stature (*TBX3*, (Liu, et al., 2021)), locomotory ((Librado, et al., 2021)) and coat coloration phenotypes (*TRPM1* for leopard spotting (Bellone, et al., 2013)*; STX17* for grey (Sundström, et al., 2012)*; PMEL* for silver (Andersson, et al., 2013)*; KIT* of tobiano (Brooks, et al., 2002)*; MITF* of splash white (Hauswirth, et al., 2012), and; *SLC45A2* responsible for cream coat color (Mariat, et al., 2003)). Coordinates are based on the EquCab3 reference genome (Kalbfleisch, et al., 2018) and the genomic data include all ancient horse genomes previously published by our group ((Schubert, et al., 2014), (Librado, et al., 2015), (Librado, et al., 2017), (Fages, et al., 2019), (Gaunitz, et al., 2018), and (Librado, et al., 2021)). Combined, these data sets help illustrate the features present in five panels of the mapDATAge package (SMAP, AMAP, TRAJECTORY, MULTIPLESNPS and ONECLICK).

General Instructions

Input Files

Input files are tab delimited .txt files or comma-separated .csv files that include the following, compulsory column headers: ‘Sample’, ‘Latitude’, ‘Longitude’, ‘Site’, ‘Species’, ‘Age’, ‘Sex’, and allele counts or numeric alleles for both allelic states (see Table S1A). Allele counts and/or numeric alleles are provided in columns with the following two headers ‘SNP_SNPNAME_N1’ and ‘SNP_SNPNAME_N2’, where N1 and N2 are C, A, G, T or D (where ‘D’ represents then the derived allele, as this can prove easier to distinguish when using the ‘MULTIPLESNPS’ panel). As many such columns as necessary can be added to allow users visualize data from multiple loci. Importantly, headers are not case sensitive; internally, mapDATAge will change all header labels to uppercase, making it easier for users to prepare their input files.

Other columns can be optionally added, including ancestral components, haplogroups and PCA (Principle Component Analysis, or Multi-Dimensional Scaling) coordinates (see Table S1B, frames highlighted in green). Columns providing ancestry components are labelled Anc1, Anc2, etc (or similarly, PC1, PC2, etc for those providing PCA/MDS coordinates). Mitochondrial, Y-chromosome haplogroups as well as other categorical data are provided under column headers ‘CAT_mtDNA’ and/or ‘CAT_Y’ (other categorical data are accepted, and can be added CAT_NAME, editing NAME according to the data type intended).

(A)

Sample	Site	Latitude	Longitude
18ELTu18_Spa_m588	ElTurunuelo_Spain	38.95	-6.06
AC7970_Tur_m290	Acemhoyuk_Turkey	38.41	33.84
AC8811_Tur_m2125	Acemhoyuk_Turkey	38.41	33.84
AC9016_Tur_m1900	Acemhoyuk_Turkey	38.41	33.84
AM115_Ira_472	Shahr-i-Qumis_Iran	35.96	54.04

SNP_chr9.73114851_GSDMC_A	SNP_chr9.73114851_GSDMC_C	Species	Sex
0	4	horse	male
0	0	horse	male
0	3	horse	female
0	2	horse	male
0	6	horse	male

SNP_rs4988235_C	SNP_rs4988235_T	Age
2	0	2538
1	0	2240
0	2	4075
2	0	3850
0	1	1478

(B)

PC1	PC2	PC3	Ance1	Ance2	Ance3	CAT_mtDNA	CAT_Y	CAT_A
0.0507	-0.0088	0.0086	0.0000	0.0000	0.0000	K	unknown	I
0.0532	-0.0056	0.0099	0.0000	0.0060	0.0000	U	unknown	I
0.0478	-0.0040	0.0096	0.0000	0.0000	0.1861	J	I	unknown
0.0454	-0.0050	0.0077	0.0000	0.0000	0.0492	K	I	unknown
0.0552	-0.0025	0.0099	0.0005	0.4538	0.0000	H	unknown	H

Table S1. Input file format overview.

A) Compulsory columns, alleles for chr9.73114851_GSDMC are listed with read counts, and rs4988235 of numeric alleles from genotype data. B) Optional columns.

Important note: missing values, special characters are not accepted by mapDATAge (using special characters such as "`", "&", "$", etc, will produce warnings and stop the programme).

For visualizing PCA results, the vector of eigenvalues must be provided so as to estimate the fraction to the total variance explained by the associated PCs (Table S2). This two-columns file does not need any header, but should be provided in .txt or .csv formats. There are many tools available for carrying out PCA, including smartPCA (Zhang, 2009), LASER (Taliun, et al., 2017) and Plink ((Purcell, et al., 2007) --pca 3 flag). It is noteworthy that any analysis

decomposing individual genomes in a multi-coordinate system can be visualized under the 'PCA' panel (e.g. Multi-Dimensional Scaling, (Wang, et al., 2010)). Please note that the mapDATAge package is not limited to the first three PCs (or 3 coordinates), but that any additional number of dimensions can be provided as long as columns with correct headers and data are provided (i.e. PC4, PC5, and more).

PC1	9.30
PC2	8.79
PC3	6.05

Table S2. Input file of eigenvalues.

In addition to providing data, and analytical results (such as sex, species, and alleles), users can provide a parameter file so as to pre-select some visualization options. This file will force visualization to the information provided by default, but all panel options in the graphical interface will remain available and editable. The Parameter file corresponds to a two-columns file, tab-separated, with 'Para' and 'Value' as main headers; rows as listed following Table S3, and multiple values can be listed, as long as separated with underscore (e.g. male_female). The 'Allele' row can be left empty if all alleles are to be considered. The last row ('Type') indicates whether read counts (*'ReadsCounts')* or genotype data (*'Genotype'*) are provided. No other parameters are accepted. Finally, a list of input files can be processed from the same path to provide the allelic counts for each locus of interest, one file per locus, with identical sample names (the order does not matter). Those files must be named mapDATAge-XXX.txt, where XXX is the name of the snp, e.g. chr9.73114851_GSDMC and chr23.22391254_DMRT3. They will be processed from the mapDATAge graphical interface once the path is indicated by entering in a path directory on '*Please enter your project root*', and then clicking on the *PATH TO YOUR PROJECT FOLDER* button.

Para	Value
Sex	male_female
Species	horse
Start	0
End	6000
WinSize	1000
StepSize	500
Allele	
Longitudest	-143
Longitudeend	143
Latitudest	35
Latitudeend	73
Type	ReadsCounts

Table S3. Parameter file.

Important note: It is recommended to enter the root of your working disk; for example, in Windows systems, you can enter "C:/" or "D:\", and "/Users" on Mac OS systems. If your Mac

OS R studio/R session shows warnings (but works normally), it likely indicates that your shinyFile package is not updated; please use the below command to update it:

```
install.packages("remotes")
remotes::install_github("thomasp85/shinyFiles", upgrade = "never")
```

Panels

The mapDATAge package includes a total of nine panels/menus/. Most panels show a sidebar subpanel (see the red rectangle labeled **1** on Fig. S1), providing parameters for visualization, including colors, the time and/or geographic range to considered and more. The resulting maps and graphs are generated within the central main subpanel (see the red rectangle labeled **2** on Fig. S1). This area is common to all panels, excepting the DATA and ONECLICK panels, which instead provide a visualization of the flat input file and additional features to display and reorganize the data tables (sort, search, paginated, and more).

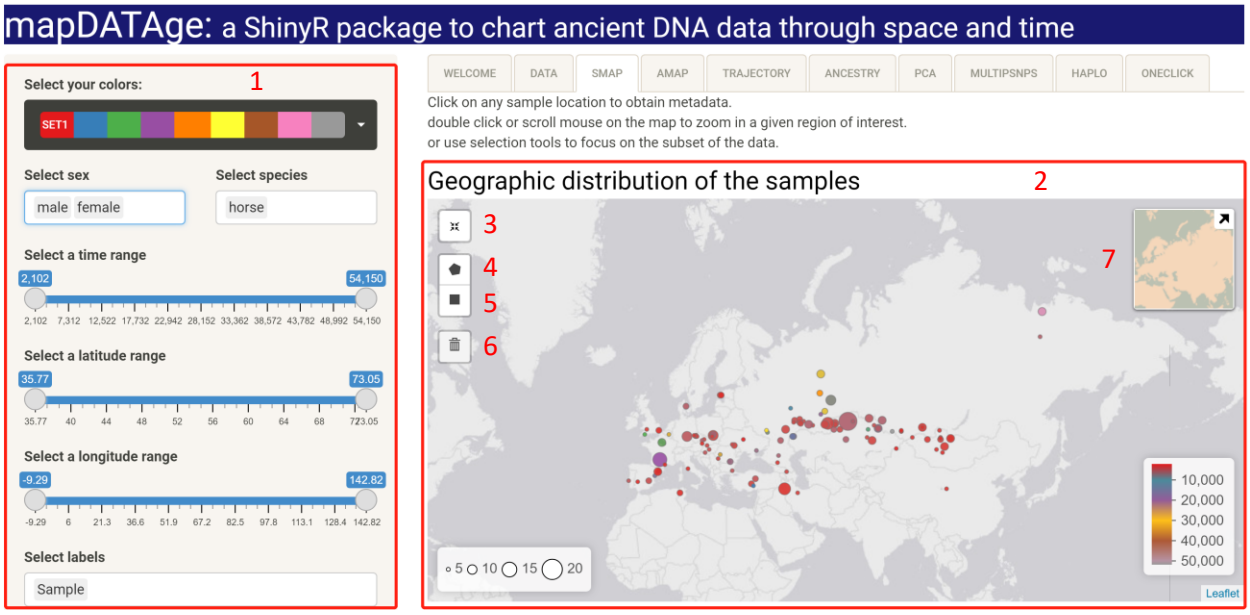


Figure S1. General visual structure of mapDATAge panels.
The display of the SMAP panel is shown as an example.

Users can move their mouse pointer over any data point shown on the map to visualize the metadata associated. Double clicking on the map allow users to zoom in and out. It is also possible to select manually a region of interest by brushing the geographic area of interest directly on the map itself once the polygon symbol (see label **4** on Fig. S1), or the square symbol (see label **5** on Fig. S1), have been selected. During this process, the dots belonging to the region selected are highlighted in red to facilitate selection. Current areas can be saved or deleted by clicking on the trash symbol (see label **6** on Fig. S1) or the reset button (see label **3** on Fig. S1), and new regions can be considered by repeating this process as many times as necessary. The mini-map showed in label **7** indicates the geographic location of the display map.

PANEL1: DATA

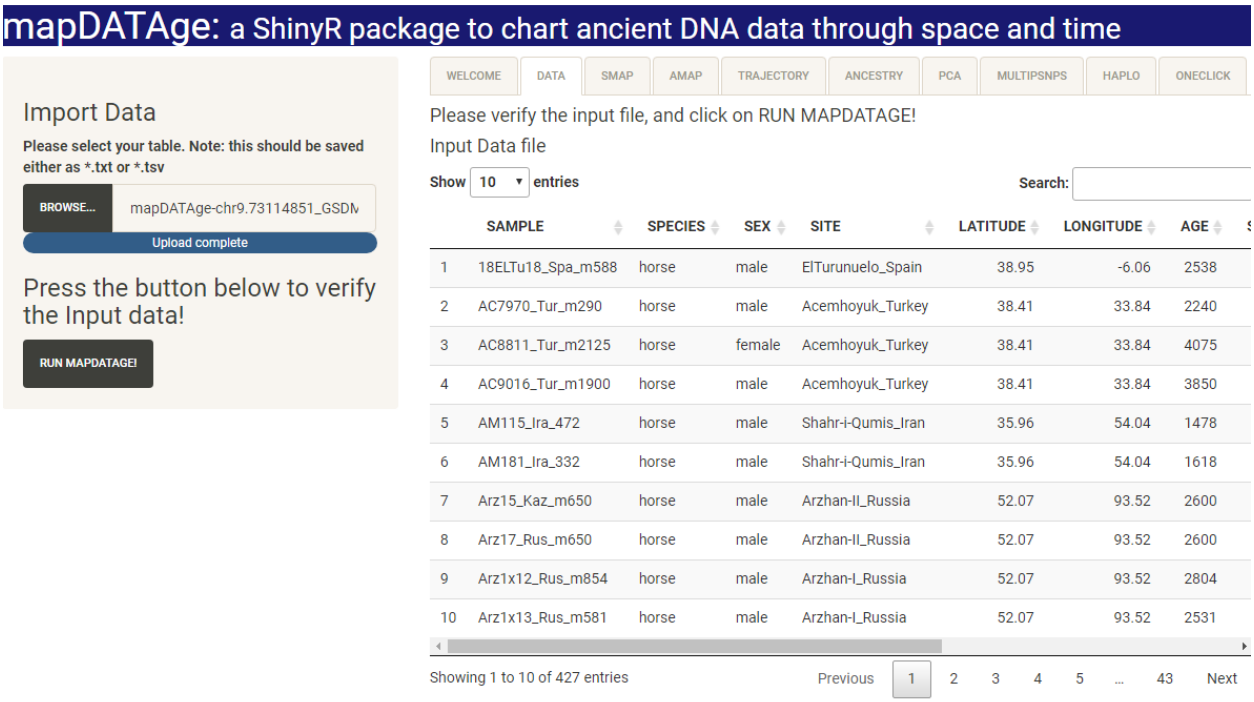


Figure S2. 'DATA' panel. (horse data set mapDATAge-chr9.73114851_GSDMC.txt).

Once browsed, input files are automatically loaded by clicking on the '*RUN MAPDATAGE*' button (Fig. S2). In case of formatting issues, errors will be flagged; the message '*No errors detected!*' will appear in case the data could be loaded properly. Data can be further sorted and searched. We recommend to refresh/stop and start the programme over again whenever new data sets must be loaded to avoid running into RAM memory issues. The resulting maps and graphs are accessible from the different panels following 'DATA' on the main menu bar, at the top (i.e. 'SMAP', 'AMAP', etc).

PANEL2: SMAP

Sample distributions are plotted using the R package leaflet, which can create interactive and user-friendly maps through the JavaScript library. This is illustrated in Fig. S3, through one of the horse data sets provided as examples (mapDATAge-chr9.73114851_GSDMC.txt). On the left side of the panel, a series of widgets are available so as to modify visualization options:

1) Color schemes from the 'RColorBrewer' package, label **1**;

2) Sex and Species filters (listing by default those indicated in the Parameter file; all instances detected from the input files otherwise), label **2**;

3) Time range (by default from 0 – i.e. now, to the age of the oldest sample present in the input file; this can be adjusted by sliding the button indicating minimum and maximum boundaries), label **3**;

4) Geographic range (by default those space limits provided in the input file; this can be adjusted by sliding the button indicating minimum and maximum latitude/longitude boundaries), label **4**;

5) Label filters for selecting those metadata appearing while overlaying with the mouse those sites/samples of interest, label **5**.

The main panel provides the polygon (see Fig. S3, label **7**) helps, square (see Fig. S3, label **8**), and save or trash (see Fig. S3, label **9**) options that allow users to select the region of interest by clicking directly on the map. The temporal distribution of those samples selected (i.e. present in the selected time range and geographic range) is provided below the map (see Fig. S3, label **11**). Finally, a reset button (see Fig. S3, label **6**), helps return to the original map after zooming in or out while a mini-map (see Fig. S3, label **10**) helps keeping track of the region under investigation, respective to the whole map. This minimap can be hidden (or restored) by clicking on the arrow in the upper right corner.

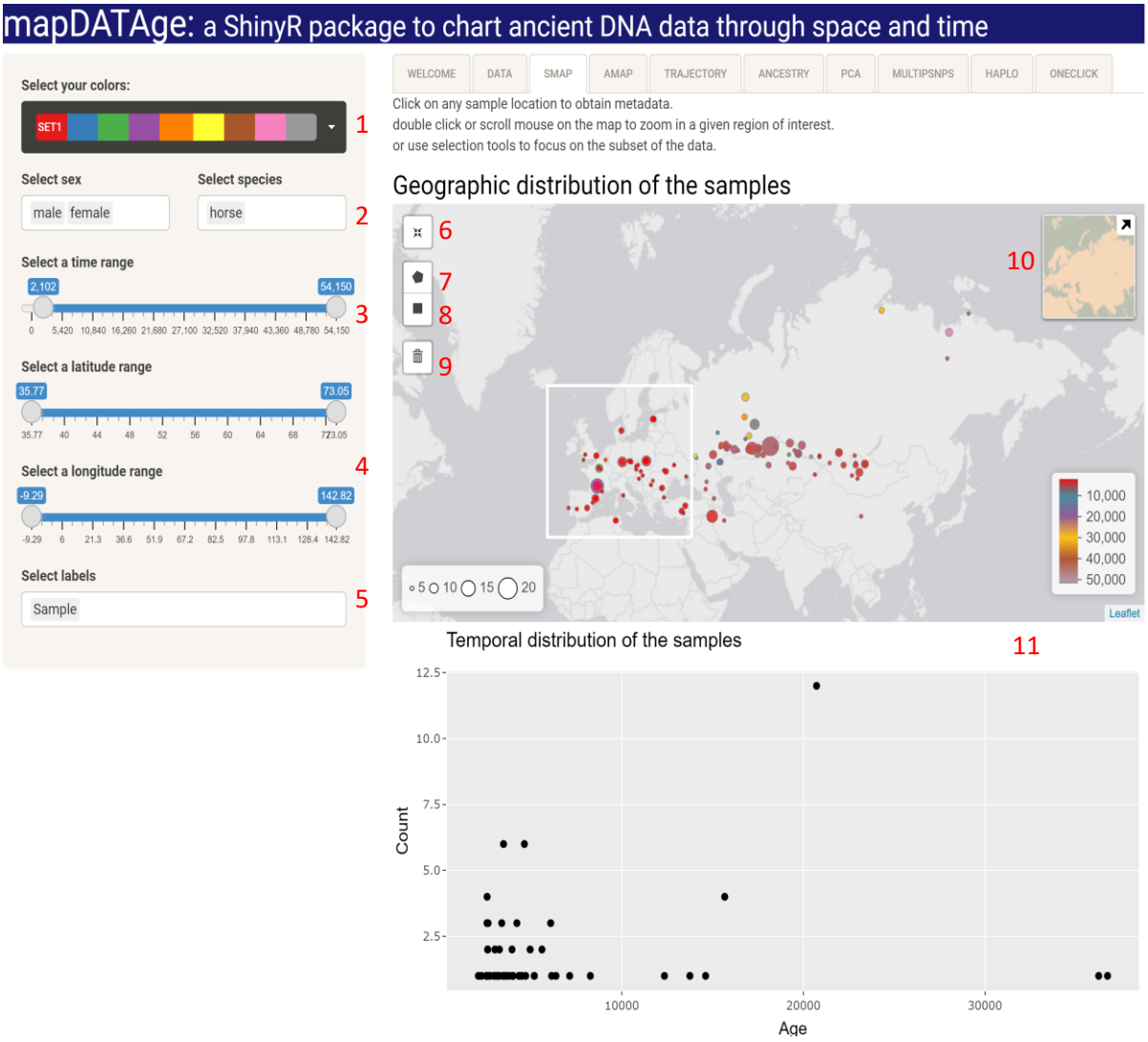


Figure S3. 'SMAP' panel (horse data set mapDATAge-chr9.73114851_GSDMC.txt).

PANEL3: AMAP

Allele distributions, ancestry components and haplogroups are mapped by using the 'addminicharts' package (https://cran.r-project.org/web/packages/leaflet.minicharts/leaflet.minicharts.pdf). The 'AMAP' panel provides the geographic and temporal distributions of the two allelic states at a given locus. Each archaeological site is represented as an individual pie chart, merging together all samples selected in the corresponding time range. This is illustrated in Fig. S4, which map allele frequency for rs4988235 (C, T and "NA", "NA" represents for samples with missing allele, e.g. C/- and T/-) in present-day and ancient Europeans dated to the last 5,000 years. The size of each pie chart is proportional to the number of samples, with corresponding keys provided at the bottom left. Options for visualization are customized through using the toolbar on the left side of the panel (see Fig. S4, indicators labelled 1-11):

1) Color schemes from the 'RColorBrewer' package, label **1**;

2) SNP filters, if the input file includes more than one SNPs, label **2**;
3) 'Select Allele type', providing 'Genotype' and 'ReadCounts' options according to the type of data provided as input, those fields providing options for users are indicated in red bold characters; 'Chart type', allowing users to represent frequencies through 'pie chart', 'bar plot', 'polar-area', or 'polar-radius', label **3**;
4) 'Sex' and 'Species' filters (listing by default those indicated in the Parameter file; all instances detected from the input files otherwise), label **4**;
5) 'Time range' (by default from 0 i.e. now, to the age of the oldest sample present in the input file; this can be adjusted by sliding the button indicating minimum and maximum boundaries), label **5**;
6) 'Geographic range' (by default those space limits provided in the input file; this can be adjusted by sliding the button indicating minimum and maximum latitude/longitude boundaries), label **6**.
7) "Select labels", filters based on Labels for selecting those metadata appearing while overlaying mouse with sites/samples of interest, label **7**.

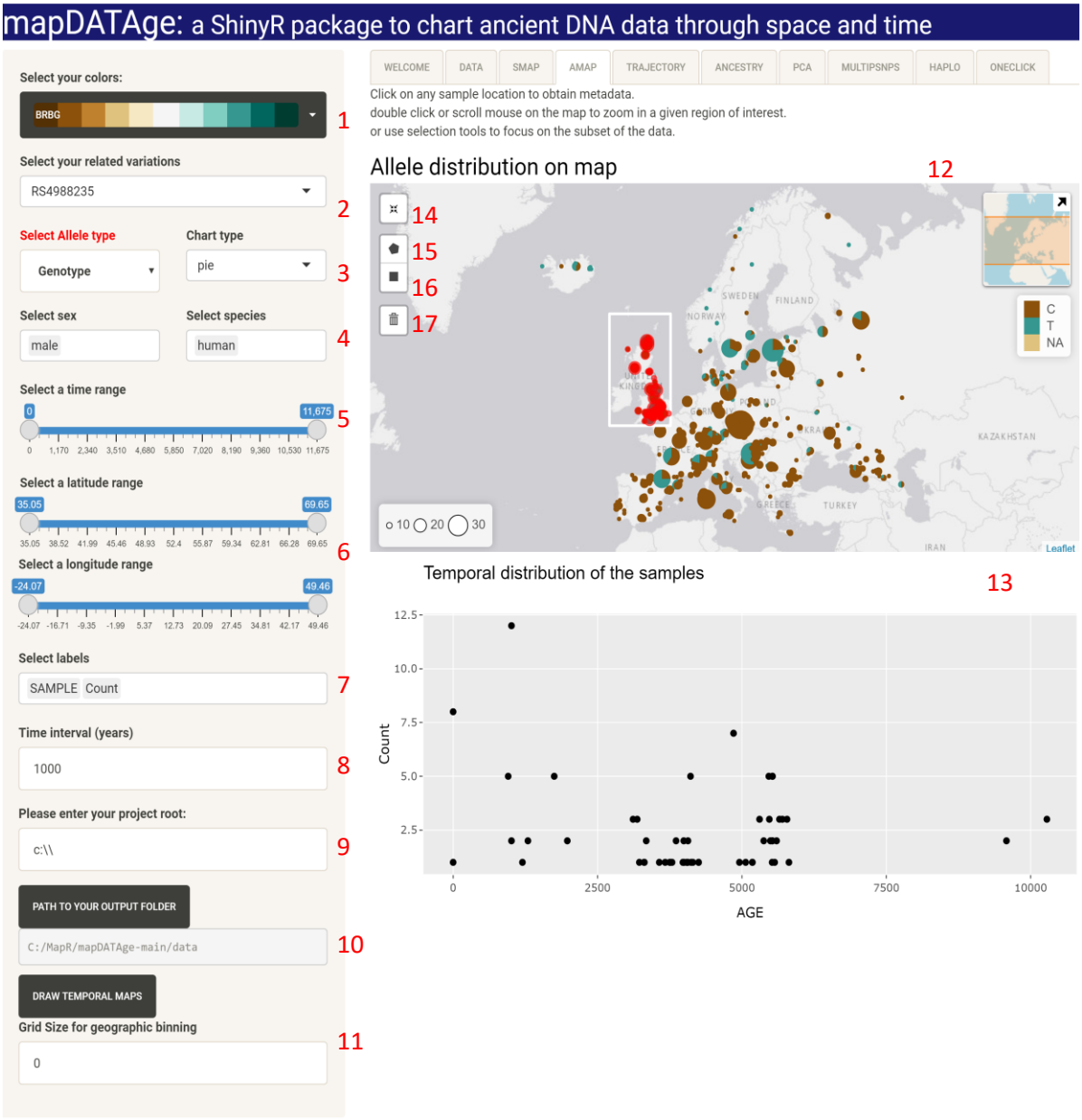


Figure S4. 'AMAP' panel. Users can select any region for their analysis. Selected regions are highlighted in red (indicator labelled **12**). (human data set mapDATAge-rs4988235_MCM6.txt). NA = samples not fully covered at this site.

Users can also slice their data within time bins of a given time range (e.g. each 1,000 years between 1,000 and 5,000 years ago; see Fig. S4, indicator labelled **8**). The output path can be selected by providing a project root at Fig. S4, indicator labelled **9**, and clicking the menu '*PATH TO YOUR OUTPUT FOLDER'* at label **10**. Once the scale of each time bin ('Time interval (years)') has been selected, and destination folder indicated, clicking on '*DRAW TEMPORAL MAPS*' will automatically generate .html output files delivering the corresponding maps. For example, if the overall time range is selected to 1,000-5,000 (at indicator labelled 5 on Fig. S4), and the time interval is 1,000 (at indicator 8), four output files will be generated, providing allelic distributions

with the [1000-2000[time bin ('1000-2000BP-piemap.html'), the [2000-3000[time bin ('2000-3000BP-piemap.html'), the [3000-4000[time bin ('3000-4000BP-piemap.html'), and the [4000-5000[time bin ('4000-5000BP-piemap.html').

Users can also group their data within geographic bins of a selected range (e.g. all sites present within 1,000 kilometers are grouped, and frequency estimates are calculated as one single location). This is controlled through the '*Grid Size for Geographic Binning'* parameter, which projects a longitudinal and latitudinal grid according to the value indicated (Fig. S4, indicator **11**).

A polygon or a square on the map will highlight the dots selected (see Fig. S4, label **12**), and show their temporal distribution at label **13**. For labels **14-17,** please refer to labels **6-9** in Fig. S3.

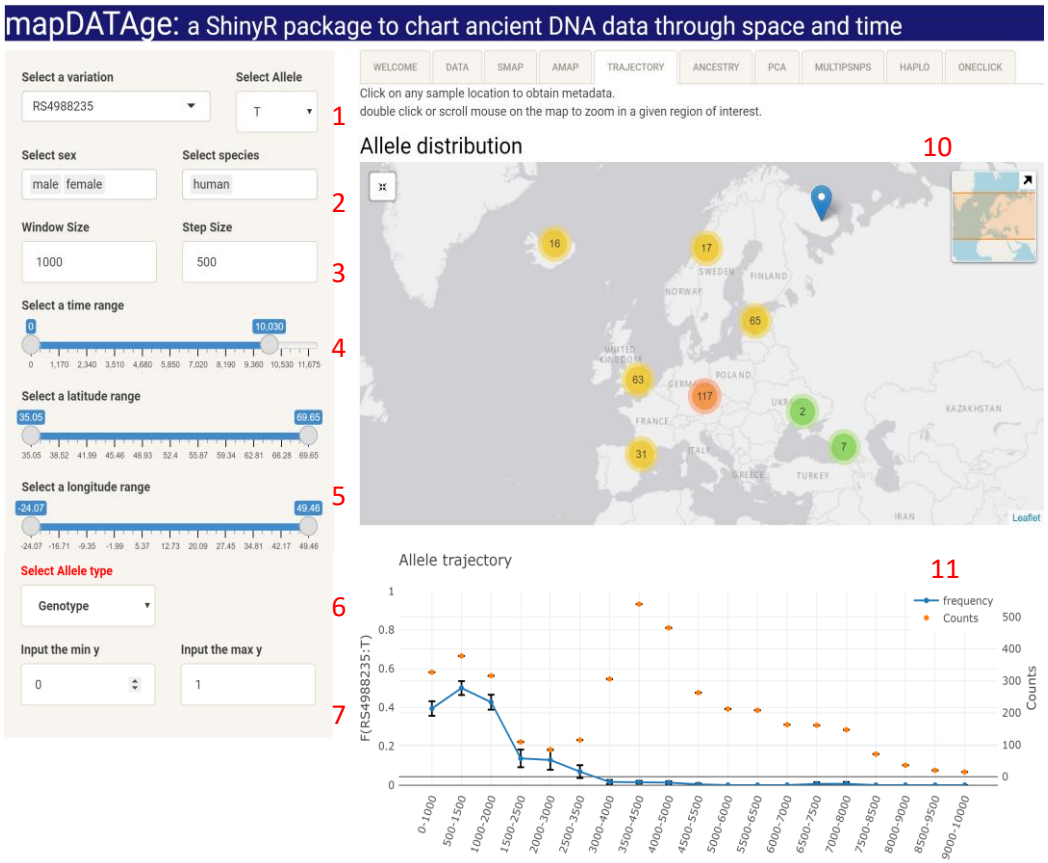
Important note: In case the grid map does not refresh, parameters must be changed back and forth to reinitialize the corresponding map.

PANEL4: TRAJECTORY

The mapDATAge package can automatically draw allele temporal trajectories at any given locus present in the input file. This can be illustrated by the temporal trajectory of the T allele at rs4988235 in Europeans (Fig. S5A) and C allele at chr9.73114851_GSDMC in horses (Fig. S5B). The left side of the panel provides the same control parameters as those provided in the 'AMAP' panel. The interval covered by each individual time bin ('Window Size') and the step size (i.e. the time shift from one-time bin to the next; 'Step Size') can be adjusted (Fig. S5A and B, indicator labelled 3). Allele frequencies and confidence intervals are calculated assuming binomial sampling from the individual genotypes provided ('Genotype' is selected on the *'Select Allele type'* menu (Fig. S5A, indicator **6**), or through random sampling one allele if the 'ReadCounts' option is selected (Fig. S5B, indicator **6**), repeating drawings according to the number of times indicated through indicator 6. Those fields providing options for users are indicated in red bold characters.

The geographic distribution of a given allele is plotted on the map (Fig. S5A and B, indicator 10), with clustered circle markers labelled according to the number of sample carriers present in a given region. Clicking on such markers will unfold all underlying samples and associated metadata. The blue, green, yellow and orange colors reflect the number of individuals, i.e. 1,]1-10],]10-100] and >100, respectively.

(A)


mapDATAge: a ShinyR package to chart ancient DNA data through space and time
Select a variation
RS4988235
Select Allele
T
Select sex
male female
Select species
human
Window Size
1000
Step Size
500
Select a time range
0
10,030
Select a latitude range
35.05
69.65
Select a longitude range
-24.07
49.46
Select Allele type
Genotype
Input the min y
0
Input the max y
1
WELCOME
DATA
SMAP
AMAP
TRAJECTORY
ANCESTRY
PCA
MULTIPSNPS
HAPLO
ONECLICK
Click on any sample location to obtain metadata.
double click or scroll mouse on the map to zoom in a given region of interest.
Allele distribution
Allele trajectory
frequency
Counts
F(RS4988235:T)
Leaflet
1
2
3
4
5
6
7
10
11

(B)

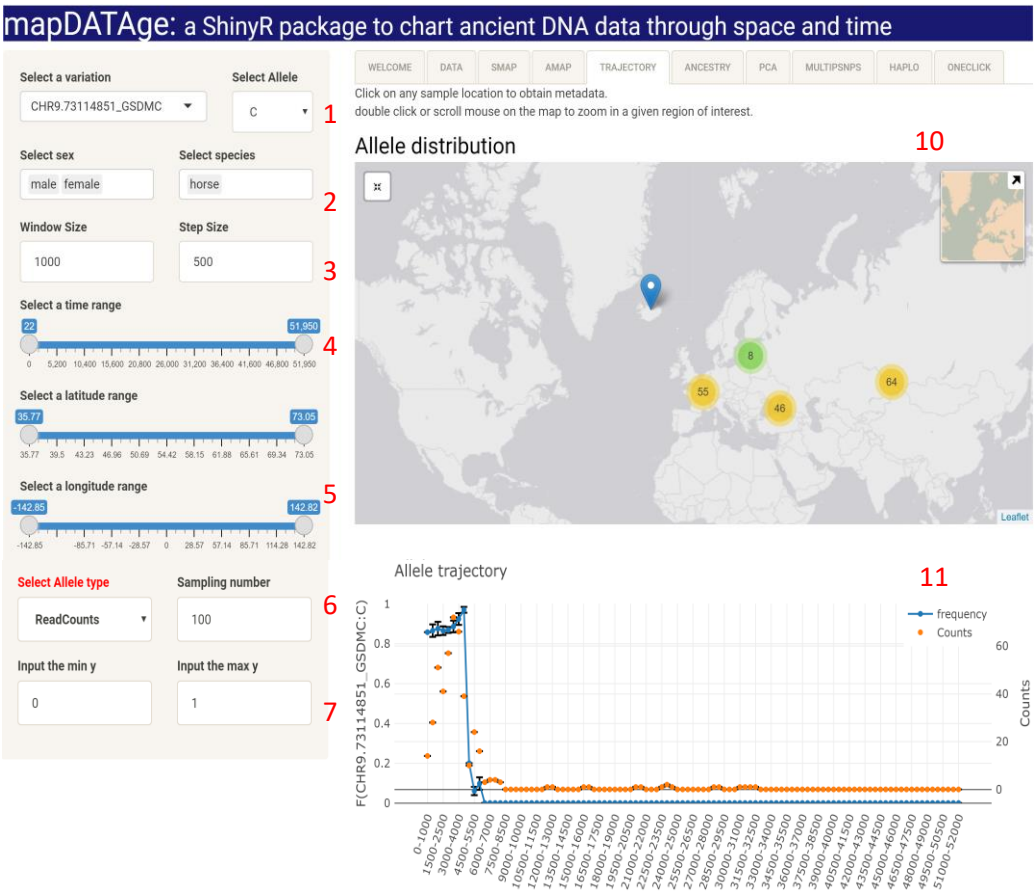

mapDATAge: a ShinyR package to chart ancient DNA data through space and time
Select a variation
CHR9.73114851_GSDMC
Select Allele
C
Select sex
male female
Select species
horse
Window Size
1000
Step Size
500
Select a time range
22
51,950
Select a latitude range
35.77
73.05
Select a longitude range
-142.85
142.82
Select Allele type
ReadCounts
Sampling number
100
Input the min y
0
Input the max y
1
WELCOME
DATA
SMAP
AMAP
TRAJECTORY
ANCESTRY
PCA
MULTIPSNPS
HAPLO
ONECLICK
Click on any sample location to obtain metadata.
double click or scroll mouse on the map to zoom in a given region of interest.
Allele distribution
Allele trajectory
frequency
Counts
F(CHR9.73114851_GSDMC:C)
Leaflet
1
2
3
4
5
6
7
10
11

Figure S5. ‘TRAJECTORY’ panel. The scale of the Y-axis can be edited through the ‘Input the min y’ and ‘Input the max y’ parameters (center vs bottom). A) Human data set mapDATAge-rs4988235_MCM6.txt). B) Horse data set mapDATAge-chr9.73114851_GSDMC.txt).

PANEL5: ANCESTRY

This panel projects individual ancestry profiles on a map for those individuals present in one or multiple temporal intervals. This feature allows users to immediately explore possible changes in the geographic distribution of ancestry profiles during specific events (e.g. global environmental, societal and/or cultural changes) (Fig. S6). The overall temporal range to be considered is controlled through the same buttons as those controlling the minimal and maximal time boundaries in the previous panels, while the time point used to split the data in two temporal windows is controlled through the ‘Time Split’ parameter. Mis-specifications (e.g. no temporal data are provided within the time intervals selected) return an error message.

The features provided by the ‘ANCESTRY’ panel can be illustrated using the horse individual ancestry profiles from Librado and colleagues (2021), in which horses belonging to the DOM2 genetic lineage spread outside their domestication homeland some ~4,000-4,200 years ago (Fig. S6). As a result, the ancestry component labelled ANC6 became dominant across Eurasia. Individual can be filtered according to a number of parameters, including sex, species, time range, geographic range. This is apparent with the example file Horse-Anc.txt, when indicating 4,200 as a ‘Time Split’ parameter and selecting 0-10,000 as an overall time range.

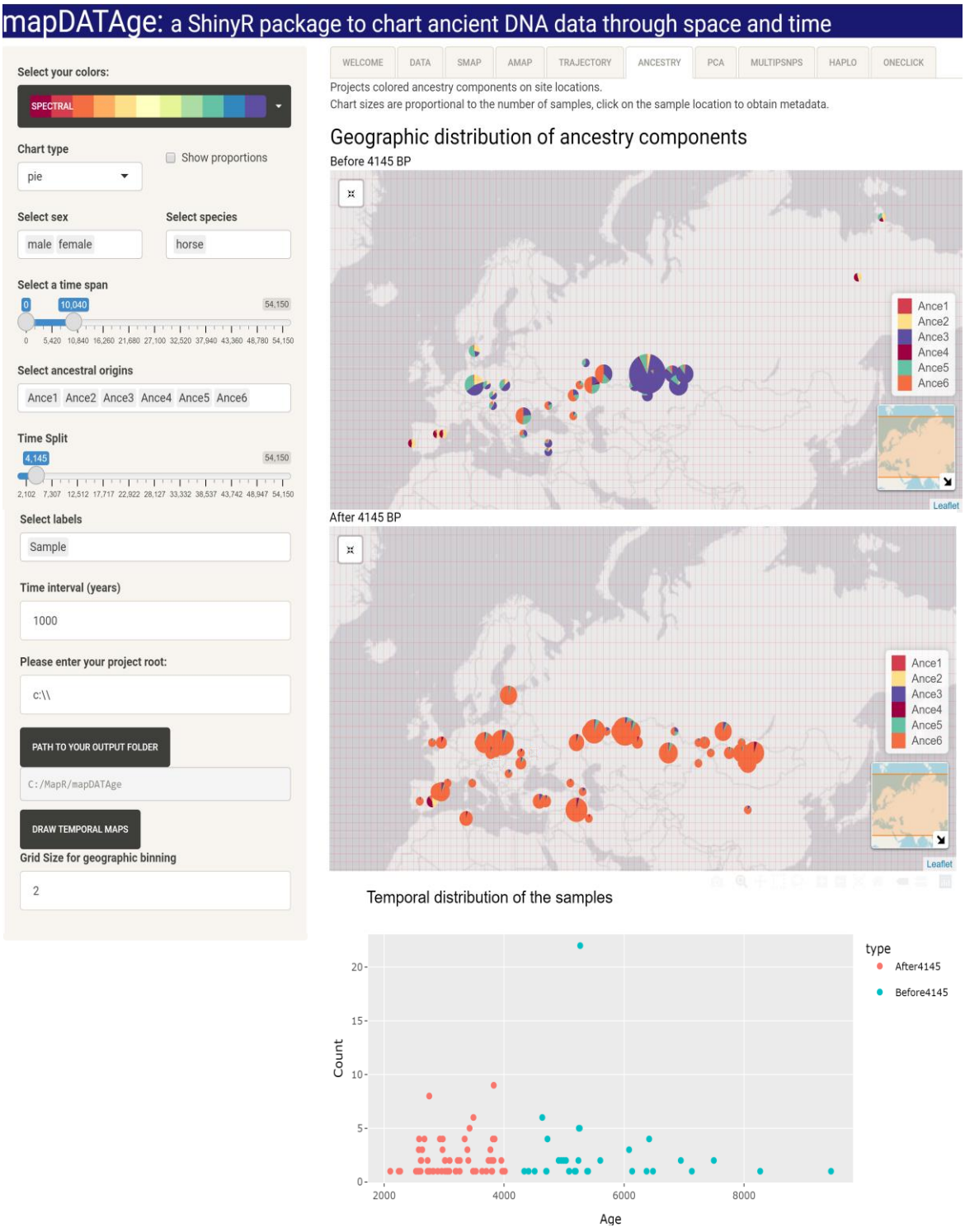


Figure S6. ‘ANCESTRY’ Panel. (horse data set Horse-Anc.txt).

The ‘ANCESTRY’ panel also provides the same options as the ‘AMAP’ panel for merging data within geographic bins of a selected size and/or for slicing the data in more than two times intervals. Once the corresponding parameters are indicated, clicking on ‘*DRAWTEMPORAL MAPS*’ will generate the corresponding maps, sliced for the selected time intervals. Output files will be written in the destination folders indicated on the left panel, following the same logic as that described in the ‘AMAP’ panel section. For example, if the time range is 2,000-6,000 years BP, and the time interval is 1,000, a total of four maps will be generated. These are shown on Fig. S7, and available from the output files named as ‘2000-3000BP-ancemap.html’, ‘3000-4000BP-ancemap.html’, ‘4000-5000BP-ancemap.html’, and; ‘5000-6000BP-ancemap.html’.

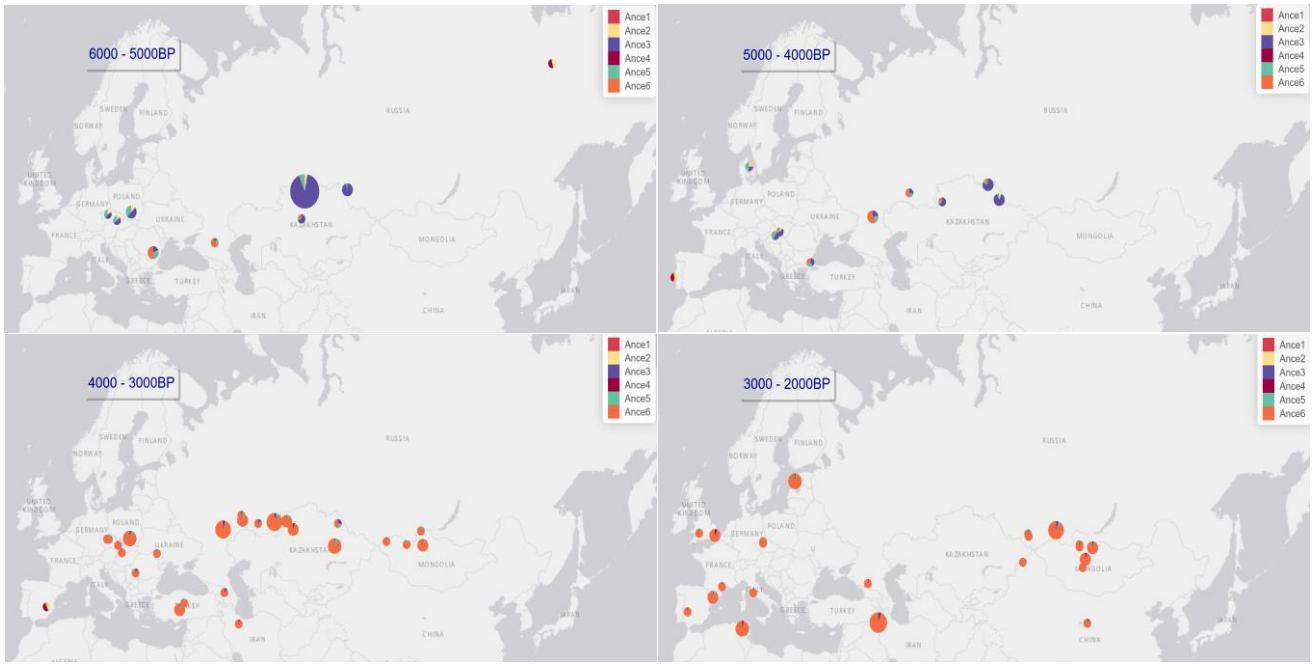


Figure S7. ‘ANCESTRY’. Panel for time split of 1,000 years (horse data set Horse-Anc.txt).

PANEL6: PCA

This panel allow users to visualize the genetic structure of their data set, using PCA, MDS, or any multivariate analysis decomposing individual genetic data in Euclidean coordinate systems. Several options are provided for filtering the data according to sex, species, time, geographic location and more. Individual metadata are available when pointing the mouse over individual dots. Fig. S8 illustrates the PCA obtained from (Librado and Orlando 2022), which made use of the same set of horses as those analysed in the ‘ANCESTRY’ panel. The geographic and temporal location of the samples to be visualized can be selected using the control panel on the left side. Please note that it does not rerun the underlying analysis, but only restricts visualization to those samples of interest, in order to enhance clarity and facilitate data exploration.

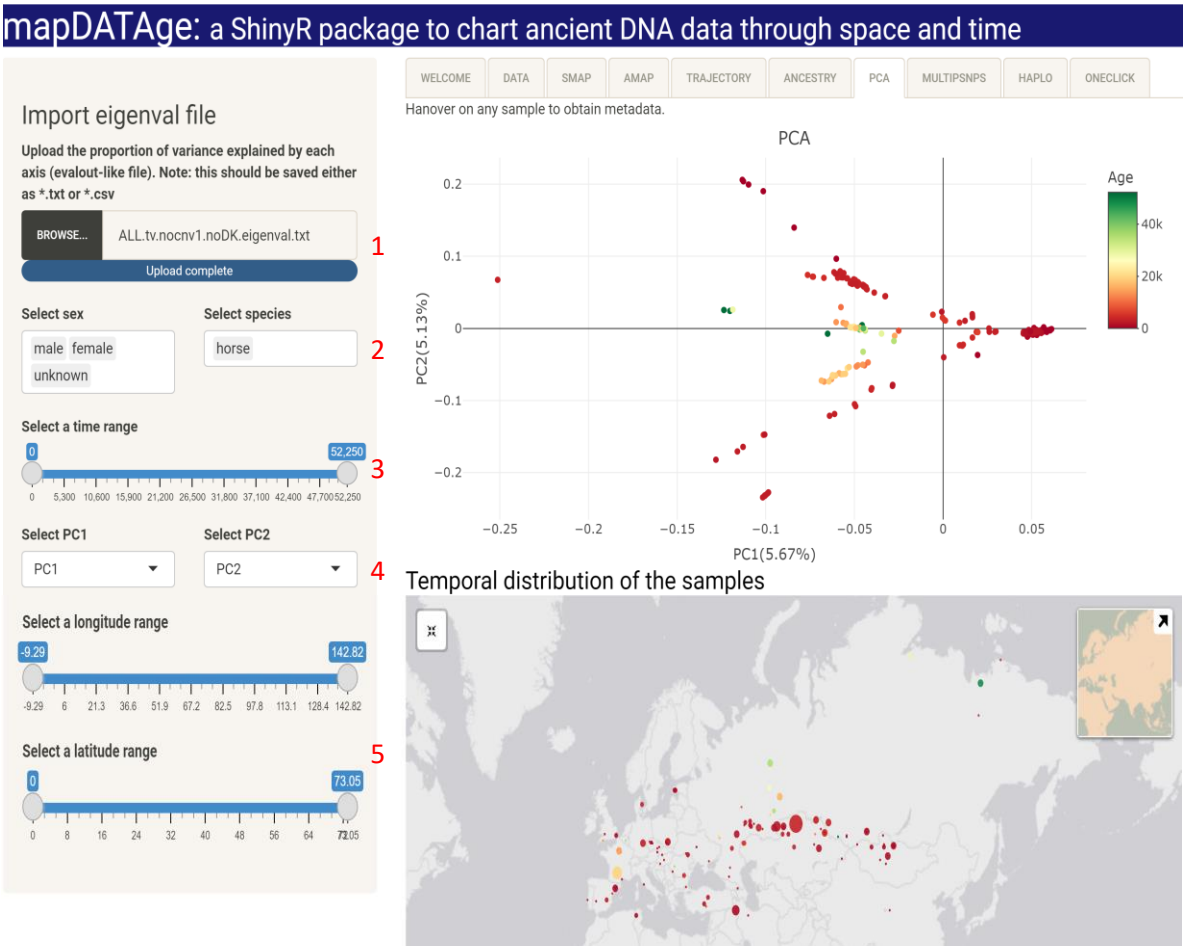


Figure S8. 'PCA' panel (horse data set PCA.eigenvec.txt).

PANEL7: MULTIPLESNPS

This panel helps users map the frequency of specific alleles at one or multiple loci. Bar plot will report the corresponding frequency of the alleles selected within all the samples present at each individual site for the time period considered. Only samples for which data are available at all the loci selected are considered (i.e. they show non-zero coverage at each and every locus). The example HorseColor-combination.txt file provides read counts for 427 ancient horses at six loci causing coat coloration phenotypes (*TRPM1*, responsible for leopard spotting (Bellone, et al., 2013); *STX17*, responsible for grey coats (Sundström, et al., 2012); *PMEL*, responsible for silver coats (Andersson, et al., 2013); *KIT*, responsible for tobiano (Brooks, et al., 2002); *MITF*, responsible for splash white (Hauswirth, et al., 2012), and; *SLC45A2*, responsible for cream coats (Mariat, et al., 2003)). Fig. S9 shows the estimated allele frequencies at each selected locus. Barplots are reported for each location, merging together all data points originating from the same site and comprised in the selected time range. It demonstrates striking differences across sites.

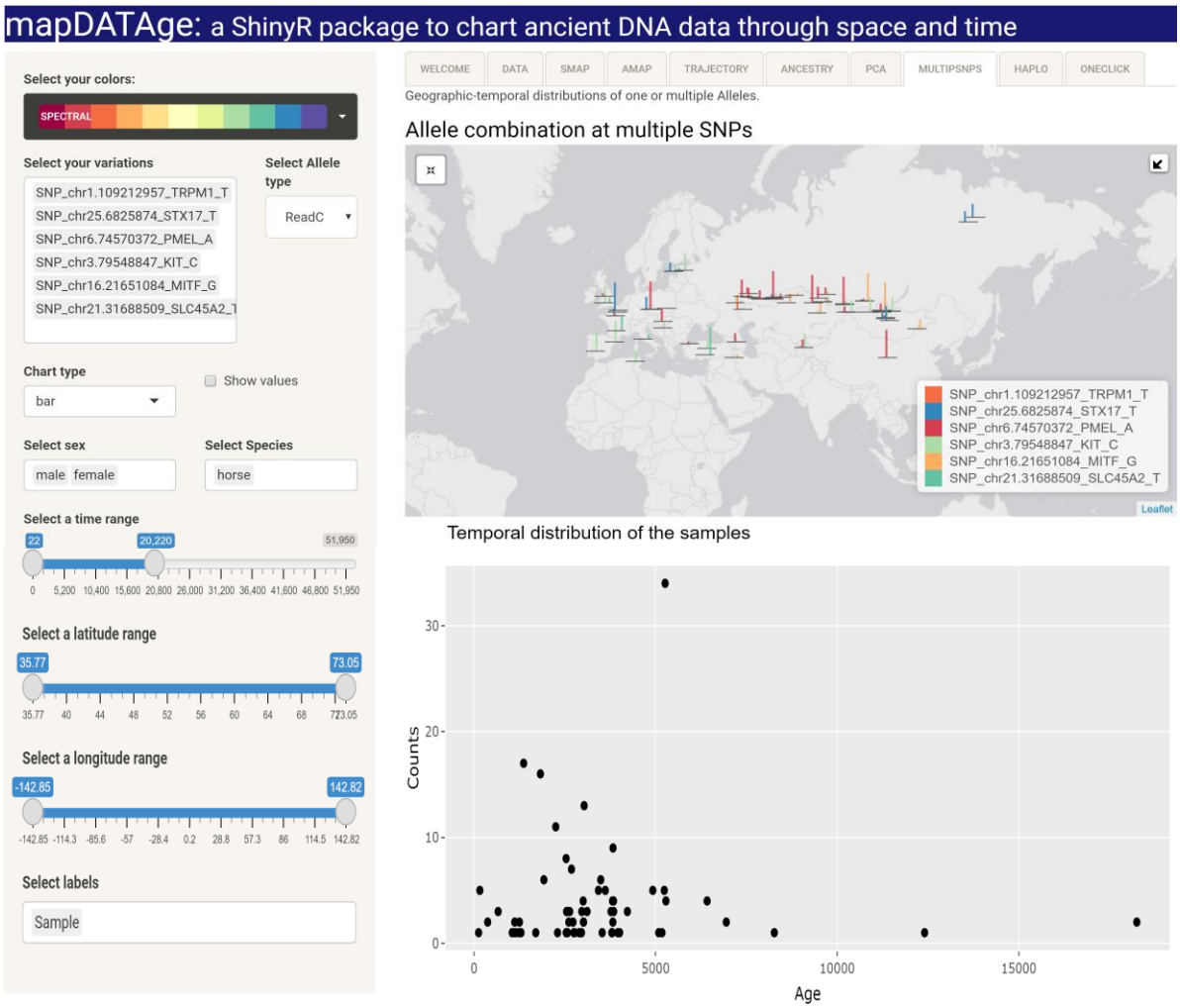


Figure S9. ‘MultipleSNPs’ panel (horse data sets HorseColor-combination.txt).

PANEL8: HAPLO

This panel helps users visualize the geographic distribution of mitochondrial and/or Y-chromosomal haplogroups within a pre-selected time range. Please note that it could also be used to map any qualitative, categorical data (as long as columns with the header CAT_XX, where XX is the name of the cluster, are provided; e.g. cultural data and/or affiliations). Typical maps generated through the ‘HAPLO’ panel are illustrated in Fig. S10, which was generated using the human data set provided as an example (mapDATAge-rs4988235_MCM6.txt). The data can be filtered and sites present within geographic grids can be merged, following the same procedures as those described in the ‘AMAP’ and ‘ANCIENT’ panels. For Y-chromosomal haplogroups, only males are considered.

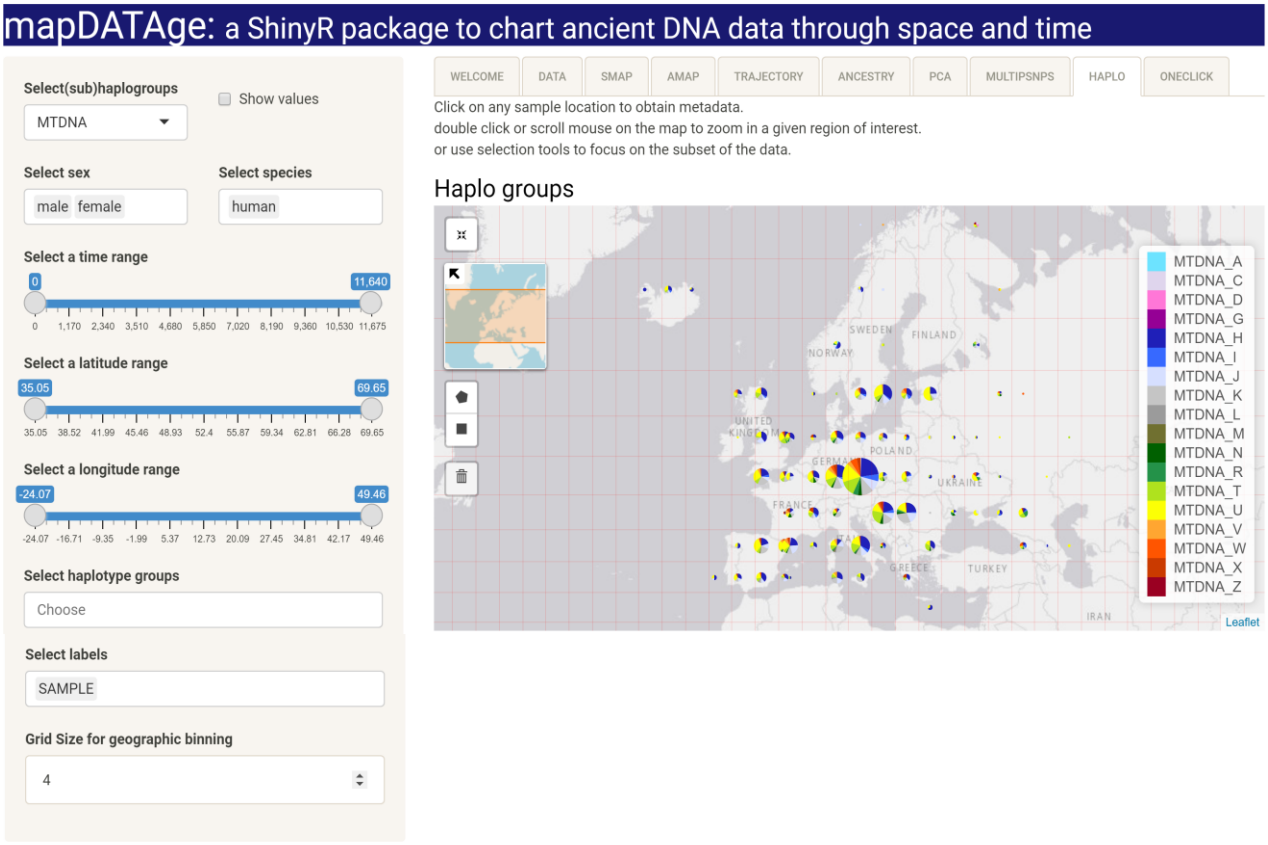


Figure S10. 'HAPLO' Panel. This chart summarizes the diversity of mitochondrial haplogroups present in human populations from Europe over the last 12,000 years.

PANEL9: ONECLICK

The 'ONECLICK' panel allow users to provide a pre-defined set of parameters (i.e. those controlling visualization on the left side of each individual panel) to be applied to one data file (Fig. S11). This will automatically generate all possible corresponding figures, without any need for filling up parameters and/or clicking on any of panels. This panel is for instance very useful when aimed at generating individual temporal trajectories for alleles at many loci, considering the same geographic and temporal ranges. The companion maps providing the geographic distribution of the samples considered for the analysis will also be generated and saved in a user-defined location as html (temporal trajectory, maps) files (Fig. S11).

Important Note: If the allele was not provided at the row 7 of the parameter file, mapDATAge will generate allele trajectories for both alleles provided in the input file, showed as Fig. S11.

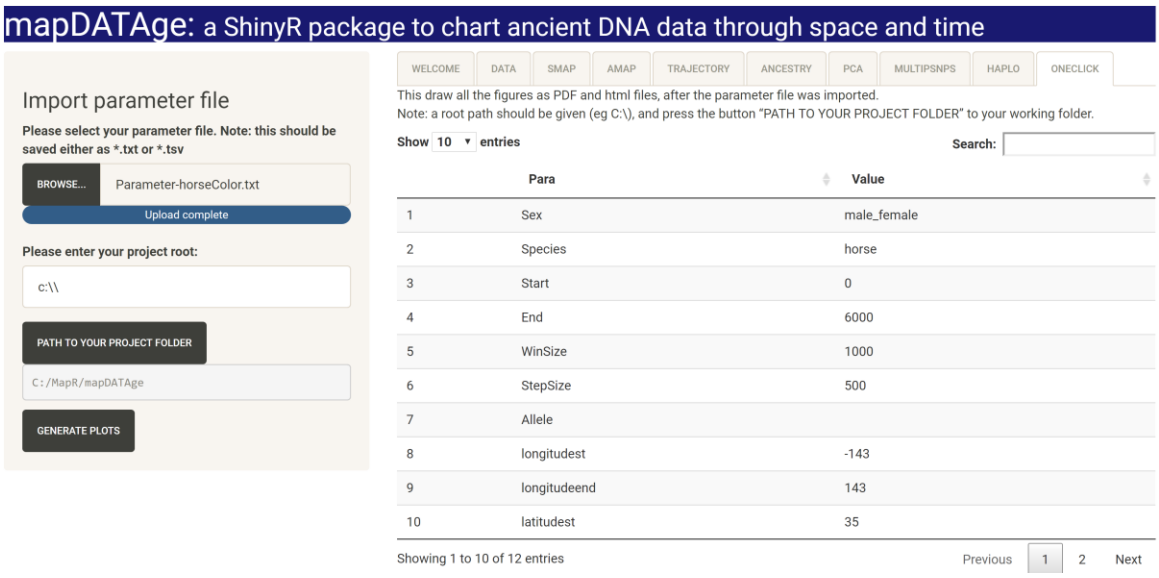


Figure S11 'ONECLICK' panel.

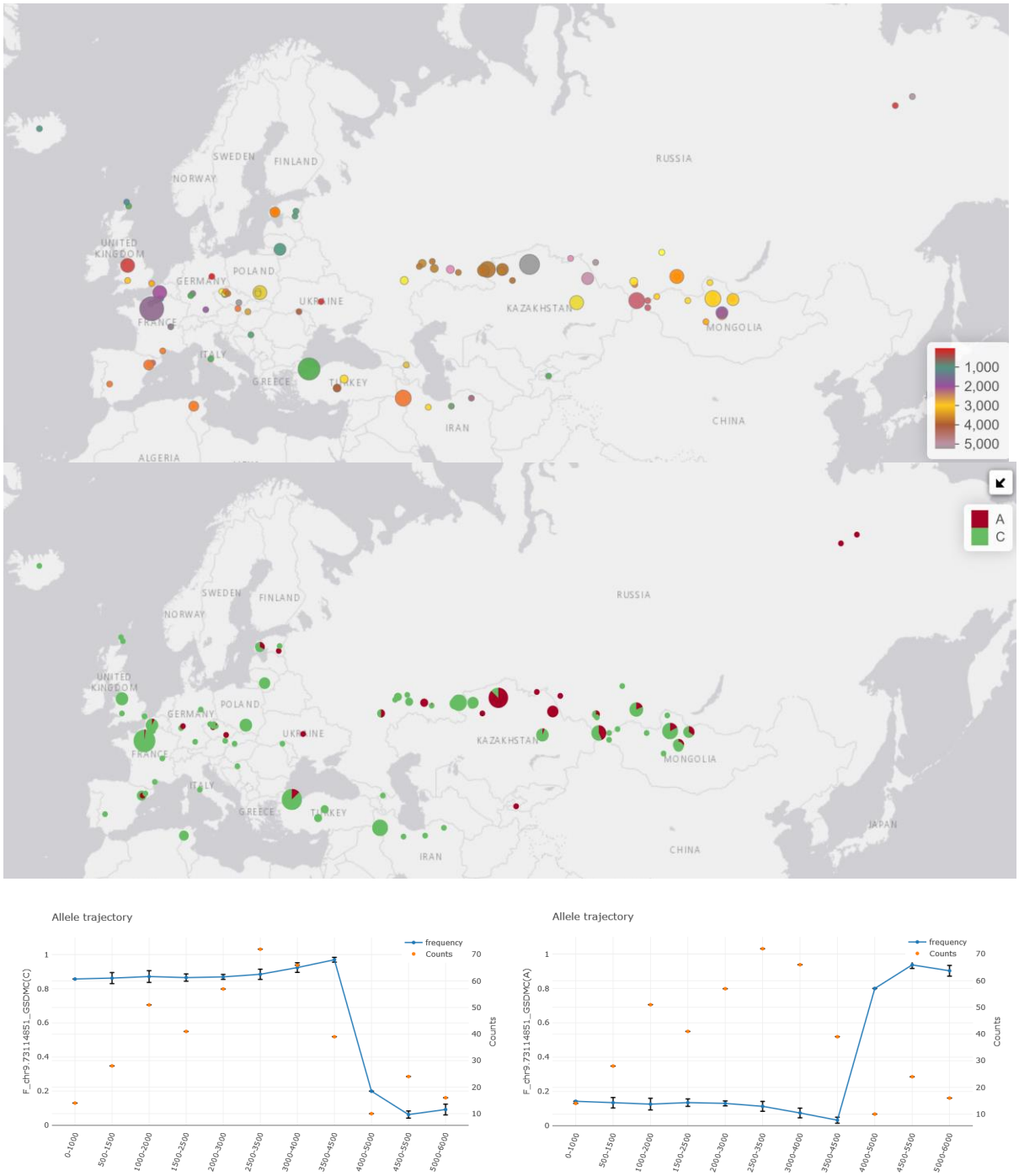


Figure S12. Example of 'ONECLICK' output (horse data set mapDATAge-chr9.73114851_GSDMC.txt).

Different maps are returned, providing the temporal and geographic distributions of samples (top), average frequency estimates for both alleles (center, here the A and C alleles at position 73,114,581 of the EquCab3 chromosome 9), as well as the temporal trajectories of both alleles (bottom; the C allele was previously shown to have almost reached fixation as the DOM2 genetic lineage of modern horses spread outside their initial homeland, some 4,200 years ago; the orange

dots indicate the number of samples at each time bin (with the 'Counts' axis providing the corresponding scale) (Librado, et al. (2021)).

References

Andersson, L.S*., et al.* Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice. *Nature* 2012;488(7413):642-646.

Andersson, L.S*., et al.* Equine multiple congenital ocular anomalies and silver coat colour result from the pleiotropic effects of mutant PMEL. *PloS one* 2013;8(9):e75639.

Anguita-Ruiz, A., Aguilera, C.M. and Gil, Á. Genetics of Lactose Intolerance: An Updated Review and Online Interactive World Maps of Phenotype and Genotype Frequencies. *Nutrients* 2020;12(9):2689.

Bellone, R.R*., et al.* Evidence for a Retroviral Insertion in TRPM1 as the Cause of Congenital Stationary Night Blindness and Leopard Complex Spotting in the Horse. *PloS one* 2013;8(10):e78280.

Brooks, S.A., Terry, R.B. and Bailey, E. A PCR-RFLP for KIT associated with tobiano spotting pattern in horses. *Animal Genetics* 2002;33(4):301-303.

Fages, A*., et al.* Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series. *Cell* 2019;177(6):1419-1435 e1431.

Gaunitz, C*., et al.* Ancient genomes revisit the ancestry of domestic and Przewalski's horses. *Science* 2018;360(6384):111-114.

Hauswirth, R*., et al.* Mutations in MITF and PAX3 Cause "Splashed White" and Other White Spotting Phenotypes in Horses. *PLOS Genetics* 2012;8(4):e1002653.

Kalbfleisch, T.S*., et al.* Improved reference genome for the domestic horse increases assembly contiguity and composition. *Communications biology* 2018;1:197.

Librado, P*., et al.* Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proceedings of the National Academy of Sciences of the United States of America* 2015;112(50):E6889-6897.

Librado, P*., et al.* Ancient genomic changes associated with domestication of the horse. *Science* 2017;356(6336):442-445.

Librado, P*., et al.* The origins and spread of domestic horses from the Western Eurasian steppes. *Nature* 2021;598(7882):634-640.

Liu, X*., et al.* A single-nucleotide mutation within the TBX3 enhancer increased body size in Chinese horses. *Current biology: CB* 2021;32.

Mariat, D., Taourit, S. and Guérin, G. A mutation in the MATP gene causes the cream coat colour in the horse. *Genetics Selection Evolution* 2003;35(1):119-133.

Purcell, S*., et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81(3):559-575.

Schubert, M*., et al.* Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature Protocols* 2014;9(5):1056-1082.

Sundström, E*., et al.* Copy number expansion of the STX17 duplication in melanoma tissue from Grey horses. *BMC genomics* 2012;13(1):365.

Taliun, D*., et al.* LASER server: ancestry tracing with genotypes or sequence reads. *Bioinformatics* 2017;33(13):2056-2058.

Wang, C*., et al.* Comparing spatial maps of human population-genetic variation using Procrustes analysis. *Statistical applications in genetics and molecular biology* 2010;9(1).

Zhang, Y. Smart pca. In, *Twenty-First International Joint Conference on Artificial Intelligence*. Citeseer; 2009.