

# Package ‘PopGenHelpR’

August 16, 2023

**Title** Streamline Population Genomic and Genetic Analyses

**Version** 1.2.1

**Description** Estimate commonly used population genomic statistics and generate publication quality figures. The current version of 'PopGenHelpR' uses vcf and csv files to generate output, however, future implementations will expand the input file type options.

**URL** <https://kfarleigh.github.io/PopGenHelpR/>

**BugReports** <https://github.com/kfarleigh/PopGenHelpR/issues>

**License** GPL (>= 3)

**Encoding** UTF-8

**RoxygenNote** 7.2.3

**Suggests** knitr, rmarkdown, testthat (>= 3.0.0)

**Config/testthat/edition** 3

**Imports** adegenet, dartR, dplyr, ggplot2, gstat, hierfstat, magrittr, methods, poppr, raster, reshape2, rlang, scatterpie, StAMPP, stats, sp, spData, spdep, utils, vcfR

**Depends** R (>= 2.10)

**LazyData** true

**VignetteBuilder** knitr

**NeedsCompilation** no

**Author** Keaka Farleigh [aut, cph, cre]

(<<https://orcid.org/0000-0002-9195-121X>>),

Mason Murphy [aut, cph, ctb] (<<https://orcid.org/0000-0002-5317-1386>>),

Christopher Blair [aut, cph, ctb]

(<<https://orcid.org/0000-0002-4782-6788>>),

Tereza Jezkova [aut, cph, ctb]

(<<https://orcid.org/0000-0003-4114-4564>>)

**Maintainer** Keaka Farleigh <[keakafarleigh@gmail.com](mailto:keakafarleigh@gmail.com)>

**Repository** CRAN

**Date/Publication** 2023-08-16 09:32:37 UTC

## R topics documented:

Dif_stats . . . . .	2
Dif_Stats_Map . . . . .	3
Div_stats . . . . .	4
Div_Stats_Map . . . . .	5
Fst_dat . . . . .	6
Het_dat . . . . .	7
HornedLizard_Pop . . . . .	8
HornedLizard_VCF . . . . .	9
Pairwise_heatmap . . . . .	9
Plot_ancestry . . . . .	10
Q_dat . . . . .	11

<b>Index</b>	<b>13</b>
--------------	-----------

---

Dif_stats	<i>A function to calculate differentiation statistics and perform significance testing with a vcf file.</i>
-----------	---

---

### Description

A function to calculate differentiation statistics and perform significance testing with a vcf file.

### Usage

```
Dif_stats(
  VCF,
  pops,
  ploidy,
  statistic = "both",
  boots,
  write = FALSE,
  prefix = NULL
)
```

### Arguments

VCF	Character string indicating the name of the vcf file to be used in analysis.
pops	Character string indicating the name of the population assignment file. This file should have four columns and be in the same order as your vcf file. The first column named Sample indicates the sample name. The second column named Population indicates the population assignment of each individual. The third column named Long indicates the longitude of the sample. The fourth column named Lat indicates the latitude of the sample.
ploidy	Numeric. The ploidy of the data.
statistic	Character string. Options are both, FST, and NeisD.

boots	Numeric. The number of bootstraps to use to evaluate statistical significance. Only relevant for FST estimation.
write	Boolean. Whether or not to write the output to a file in the current working directory.
prefix	Character string that will be appended to file output.

**Value**

A list containing data frames for the requested statistic.

**Examples**

```
data("HornedLizard_Pop")
data("HornedLizard_VCF")
Test <- Dif_stats(VCF = HornedLizard_VCF, pops = HornedLizard_Pop,
ploidy = 2, statistic = "both", boots = 10, write = FALSE)
```

---

Dif_Stats_Map	<i>A function to map differentiation statistics.</i>
---------------	--

---

**Description**

A function to map differentiation statistics.

**Usage**

```
Dif_Stats_Map(
  dat,
  pops,
  neighbors,
  col,
  breaks = NULL,
  Lat_buffer,
  Long_buffer
)
```

**Arguments**

dat	Data frame or character string that supplies the input data. If it is a character string, the file should be a csv. If it is a csv, the 1st row should contain the individual/population names. The columns should also be named in this fashion.
pops	Data frame or character string that supplies the input data. If it is a character string, the file should be a csv. The columns should be named Sample, containing the sample IDs; Population indicating the population assignment of the individual; Long, indicating the longitude of the sample; Lat, indicating the latitude of the sample.

neighbors	Numeric. The number of neighbors to plot connections with.
col	Character vector indicating the colors you wish to use for plotting.
breaks	Numeric. The breaks used to generate the color ramp when plotting. Users should supply 3 values if custom breaks are desired.
Lat_buffer	Numeric. A buffer to customize visualization.
Long_buffer	Numeric. A buffer to customize visualization.

**Value**

A list containing the map and the matrix used to plot the map.

**Examples**

```
data(Fst_dat)
Fst <- Fst_dat[[1]]
Loc <- Fst_dat[[2]]
Test <- Dif_Stats_Map(dat = Fst, pops = Loc,
neighbors = 2,
col = c('#fd8d3c', '#fc4e2a', '#e31a1c', '#bd0026', '#800026'), Lat_buffer = 1, Long_buffer = 1)
```

---

Div_stats	<i>A function to estimate heterozygosity and the number of private alleles from a vcf file.</i>
-----------	---

---

**Description**

A function to estimate heterozygosity and the number of private alleles from a vcf file.

**Usage**

```
Div_stats(VCF, pops, ploidy, write = FALSE, prefix)
```

**Arguments**

VCF	Character string indicating the name of the vcf file to be used in analysis.
pops	Character string indicating the name of the population assignment file. This file should have four columns and be in the same order as your vcf file. The first column named Sample indicates the sample name. The second column named Population indicates the population assignment of each individual. The third column named Longitude indicates the longitude of the sample. The fourth column named Latitude indicates the latitude of the sample.
ploidy	Numeric. The ploidy of the data.
write	Boolean. Whether or not to write the output to a file in the current working directory.
prefix	Character string that will be appended to file output.

**Value**

A list containing the estimated diversity statistics, model output from linear regression of these statistics against latitude, and model plots.

**Examples**

```
data("HornedLizard_Pop")
data("HornedLizard_VCF")
Test <- Div_stats(VCF = HornedLizard_VCF, pops = HornedLizard_Pop,
ploidy = 2, write = FALSE)
```

---

Div_Stats_Map	<i>A function to map diversity statistics.</i>
---------------	--

---

**Description**

A function to map diversity statistics.

**Usage**

```
Div_Stats_Map(
  dat,
  plot.type = "all",
  statistic,
  breaks = NULL,
  col,
  Lat_buffer = 1,
  Long_buffer = 1,
  write = FALSE,
  prefix = NULL
)
```

**Arguments**

dat	Data frame or character string that supplies the input data. If it is a character string, the file should be a csv. The first column should be the statistic to be plotted and named the same as the statistic argument. The second column is Population indicating which population each row belongs to. The third column is the standard deviation, the fourth column is Long indicating the longitude, and the fifth column is Lat, indicating the latitude.
plot.type	Character string. Options are all, point, or interpolated. All is recommended and will generate a map with points colored according to heterozygosity as well as a raster of interpolated heterozygosity values.
statistic	Character string. The statistic to be plotted.
breaks	Numeric. The breaks used to generate the color ramp when plotting. Users should supply 3 values if custom breaks are desired.

col	Character vector indicating the colors you wish to use for plotting, three colors are allowed (low, mid, high). The first color will be the low color, the second the middle, the third the high.
Lat_buffer	Numeric. A buffer to customize visualization.
Long_buffer	Numeric. A buffer to customize visualization.
write	Boolean. Whether or not to write the output to a file in the current working directory.
prefix	Character string that will be appended to file output.

### Value

A list containing maps and the data frames used to generate them.

### Examples

```
data(Het_dat)
Test_het <- Div_Stats_Map(dat = Het_dat, plot.type = 'all',
  statistic = "Heterozygosity",
  Lat_buffer = 1, Long_buffer = 1, write = FALSE, prefix = 'Test_het')
```

---

Fst_dat	<i>A genetic differentiation matrix and locality information for each population. This data was generated by subsetting data of Farleigh et al., 2021.</i>
---------	--

---

### Description

A symmetric matrix with estimated genetic differentiation (Fst) between 3 populations.

### Usage

```
data(Fst_dat)
```

### Format

A list with two elements:

**Fst\_dat** Data frame with three rows and three columns

**Loc\_dat** Data frame containing the locality information for each population

...

### Source

Farleigh, K., Vladimirova, S. A., Blair, C., Bracken, J. T., Koochekian, N., Schield, D. R., ... & Jezkova, T. (2021). The effects of climate and demographic history in shaping genomic variation across populations of the Desert Horned Lizard (*Phrynosoma platyrhinos*). *Molecular Ecology*, 30(18), 4481-4496.

## Examples

```
data(Fst_dat)
Fst <- Fst_dat[[1]]
Loc <- Fst_dat[[2]]

Test <- Dif_Stats_Map(dat = Fst, pops = Loc,
neighbors = 2,
col = c('#fd8d3c', '#fc4e2a', '#e31a1c', '#bd0026', '#800026'), Lat_buffer = 1, Long_buffer = 1)

Fstat_plot <- Pairwise_heatmap(dat = Fst, statistic = 'FST')
```

---

Het_dat	<i>A data frame of hypothetical heterozygosity data produced by Div_Stats.</i>
---------	--

---

## Description

Data frame containing 5 columns and 3 rows

## Usage

```
data(Het_dat)
```

## Format

A data frame with 5 columns and 3 rows:

**Heterozygosity** Estimated heterozygosity

**Pop** Population assignment

**Standard.Deviation** standard deviation

**Longitude** Longitude

**Latitude** Latitude

...

## Source

Coordinates and population names taken from Farleigh, K., Vladimirova, S. A., Blair, C., Bracken, J. T., Koochekian, N., Schield, D. R., ... & Jezkova, T. (2021). The effects of climate and demographic history in shaping genomic variation across populations of the Desert Horned Lizard (*Phrynosoma platyrhinos*). *Molecular Ecology*, 30(18), 4481-4496.

**Examples**

```
data(Het_dat)
Test_het <- Div_Stats_Map(dat = Het_dat, plot.type = 'all',
  statistic = "Heterozygosity",
  Lat_buffer = 1, Long_buffer = 1, write = FALSE, prefix = 'Test_het')
```

---

HornedLizard_Pop	<i>A population assignment data frame to be used in Div_stats and Dif_stats.</i>
------------------	--

---

**Description**

Data frame containing 4 columns and 72 rows

**Usage**

```
data(HornedLizard_Pop)
```

**Format**

A data frame with 4 columns and 72 rows:

**Sample** Sample Name

**Population** Population assignment according to sNMF results (see citation)

**Longitude** Longitude

**Latitude** Latitude

...

**Source**

Coordinates and population names taken from Farleigh, K., Vladimirova, S. A., Blair, C., Bracken, J. T., Koochekian, N., Schield, D. R., ... & Jezkova, T. (2021). The effects of climate and demographic history in shaping genomic variation across populations of the Desert Horned Lizard (*Phrynosoma platyrhinos*). *Molecular Ecology*, 30(18), 4481-4496.

**Examples**

```
data("HornedLizard_Pop")
data("HornedLizard_VCF")
Test <- Div_stats(VCF = HornedLizard_VCF, pops = HornedLizard_Pop,
  ploidy = 2, write = FALSE)
```



---

HornedLizard\_VCF      *A vcfR object to be used in Div\_stats and Dif\_stats.*

---

**Description**

Data frame containing 4 columns and 72 rows

**Usage**

```
data(HornedLizard_Pop)
```

**Format**

A vcfR object

**vcfR object** A vcfR object containing genotype and sample information for 72 individuals.

...

**Source**

Farleigh, K., Vladimirova, S. A., Blair, C., Bracken, J. T., Koochekian, N., Schield, D. R., ... & Jezkova, T. (2021). The effects of climate and demographic history in shaping genomic variation across populations of the Desert Horned Lizard (*Phrynosoma platyrhinos*). *Molecular Ecology*, 30(18), 4481-4496.

**Examples**

```
data("HornedLizard_Pop")
data("HornedLizard_VCF")
Test <- Div_stats(VCF = HornedLizard_VCF, pops = HornedLizard_Pop,
ploidy = 2, write = FALSE)
```

---

Pairwise\_heatmap      *A function to plot a heatmap from a symmetric matrix.*

---

**Description**

A function to plot a heatmap from a symmetric matrix.

**Usage**

```
Pairwise_heatmap(dat, statistic, col = NULL)
```

**Arguments**

dat	Data frame or character string that supplies the input data. If it is a character string, the file should be a csv. If it is a csv, the 1st row should contain the individual/population names. The columns should also be named in this fashion.
statistic	Character indicating the statistic represented in the matrix, this will be used to label the plot.
col	Character vector indicating the colors to be used in plotting. The vector should contain two colors, the first will be the low value, the second will be the high value.

**Value**

A heatmap plot

**Examples**

```
#' data(Fst_dat)
Fst <- Fst_dat[[1]]
Fstat_plot <- Pairwise_heatmap(dat = Fst, statistic = 'FST')
```

---

Plot\_ancestry

---

*Plot an ancestry matrix and map of ancestry pie charts.*


---

**Description**

Plot an ancestry matrix and map of ancestry pie charts.

**Usage**

```
Plot_ancestry(
  anc.mat,
  pops,
  K,
  plot.type = "all",
  col,
  Lat_buffer,
  Long_buffer
)
```

**Arguments**

anc.mat	Data frame or character string that supplies the input data. If it is a character string, the file should be a csv. The first column should be the names of each sample/population, followed by the estimated contribution of each cluster to that individual/pop.
---------	--

pops	Data frame or character string that supplies the input data. If it is a character string, the file should be a csv. The columns should be named Sample, containing the sample IDs; Population indicating the population assignment of the individual, population and sample names must be the same type (i.e., both numeric or both characters); Long, indicating the longitude of the sample; Lat, indicating the latitude of the sample.
K	Numeric. The number of genetic clusters in your data set, please contact the package authors if you need help doing this.
plot.type	Character string. Options are all, individual, and population. All is default and recommended, this will plot a barchart and piechart map for both the individuals and populations.
col	Character vector indicating the colors you wish to use for plotting.
Lat_buffer	Numeric. A buffer to customize visualization.
Long_buffer	Numeric. A buffer to customize visualization.

**Value**

A list containing your plots and the data frames used to generate the plots.

**Examples**

```
data(Q_dat)
Qmat <- Q_dat[[1]]
rownames(Qmat) <- Qmat[,1]
Loc <- Q_dat[[2]]
Test_all <- Plot_ancestry(anc.mat = Qmat, pops = Loc, K = 5,
plot.type = 'all', col <- c('red', 'maroon', 'navy', 'cyan', 'blue'),
Lat_buffer = 1, Long_buffer = 1)
```

---

Q_dat	<i>A list representing a q-matrix and the locality information associated with the qmatrix</i>
-------	--

---

**Description**

List with two elements

**Usage**

```
data(Q_dat)
```

**Format**

A list with two elements:

**Qmat** A q-matrix with 6 columns and 30 rows, the first column lists the sample name and the remaining 5 represent the contribution a genetic cluster to that individuals ancestry

**Loc\_dat** The locality information for each individual in the q-matrix

...

**Source**

Data was generated by package authors.

**Examples**

```
data(Q_dat)
Qmat <- Q_dat[[1]]
rownames(Qmat) <- Qmat[,1]
Loc <- Q_dat[[2]]
Test_all <- Plot_ancestry(anc.mat = Qmat, pops = Loc, K = 5,
plot.type = 'all', col <- c('red', 'maroon', 'navy', 'cyan', 'blue'),
Lat_buffer = 1, Long_buffer = 1)
```

# Index

## \* datasets

Fst\_dat, [6](#)

Het\_dat, [7](#)

HornedLizard\_Pop, [8](#)

HornedLizard\_VCF, [9](#)

Q\_dat, [11](#)

Dif\_stats, [2](#)

Dif\_Stats\_Map, [3](#)

Div\_stats, [4](#)

Div\_Stats\_Map, [5](#)

Fst\_dat, [6](#)

Het\_dat, [7](#)

HornedLizard\_Pop, [8](#)

HornedLizard\_VCF, [9](#)

Pairwise\_heatmap, [9](#)

Plot\_ancestry, [10](#)

Q\_dat, [11](#)