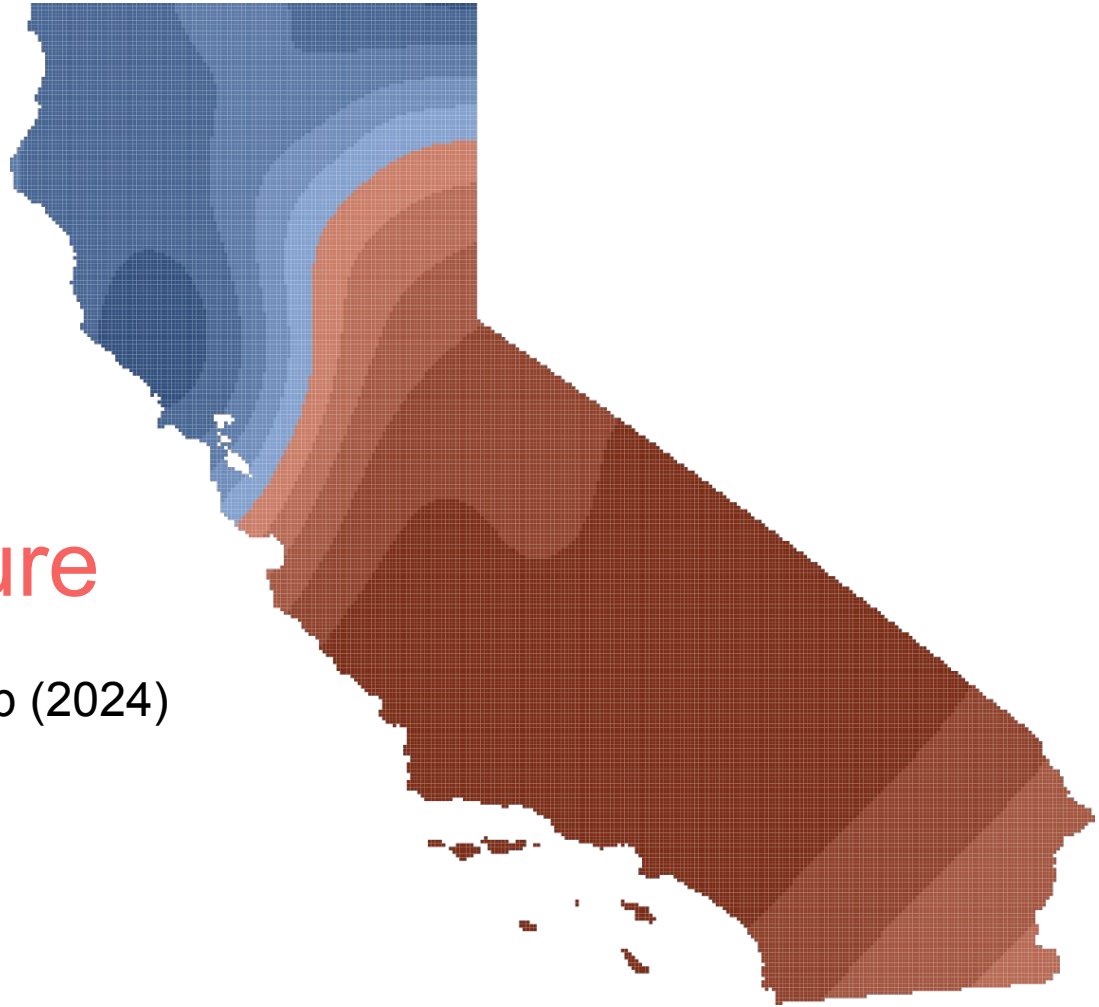
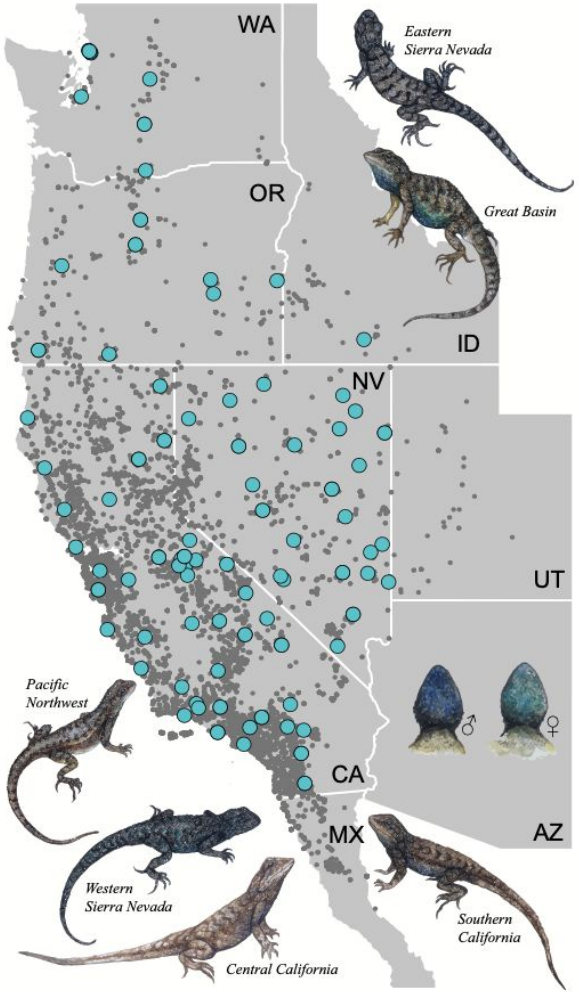
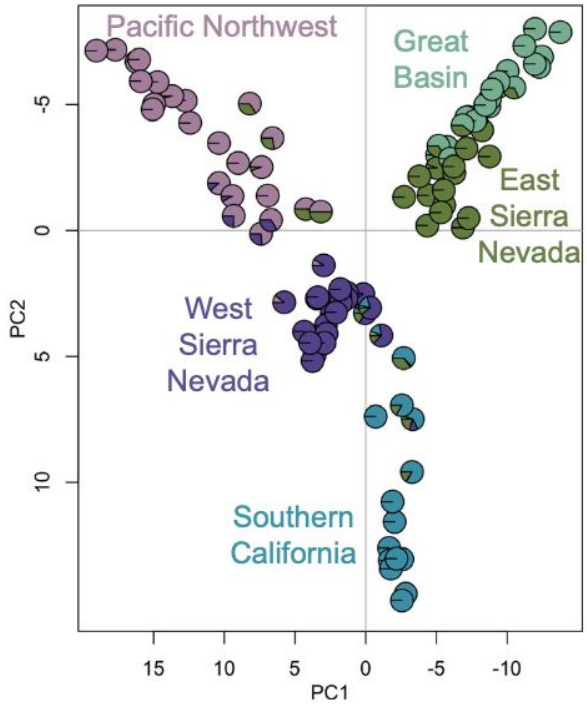
# Population structure

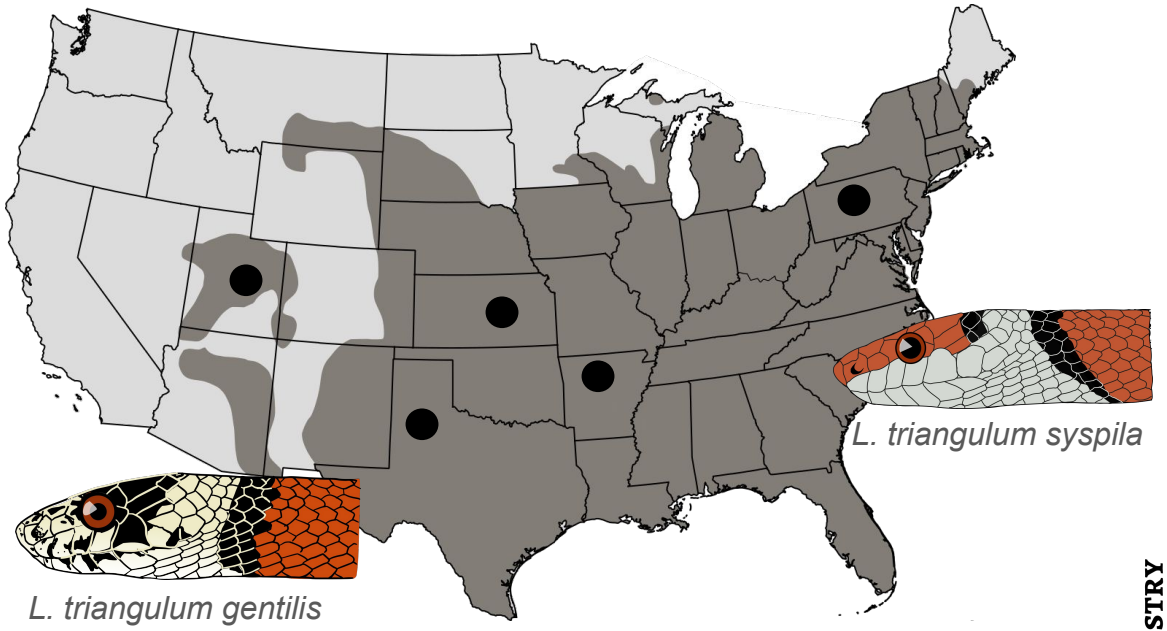
Anne Chambers & Anusha Bishop (2024)



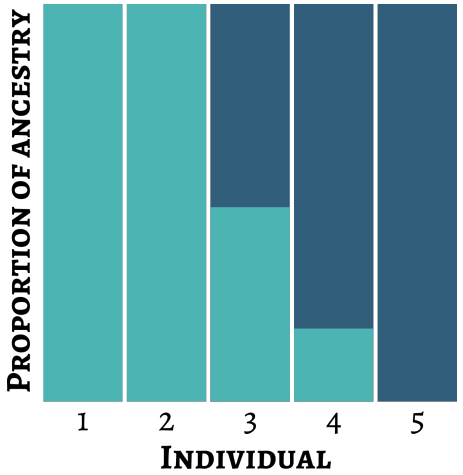


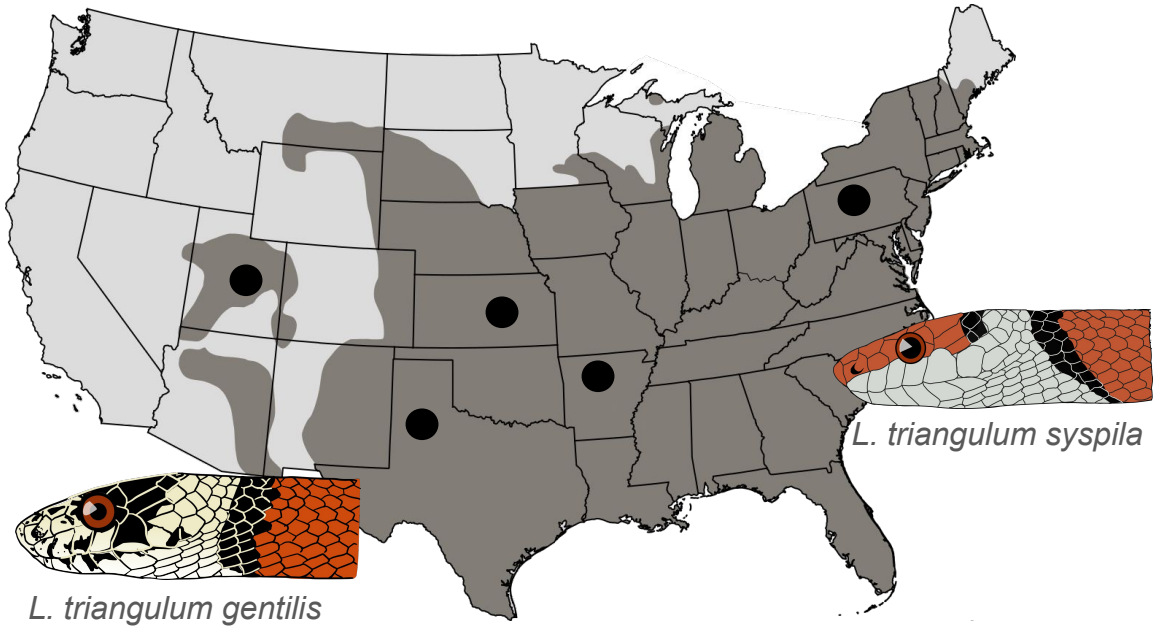




Bouzid et al. (2022)

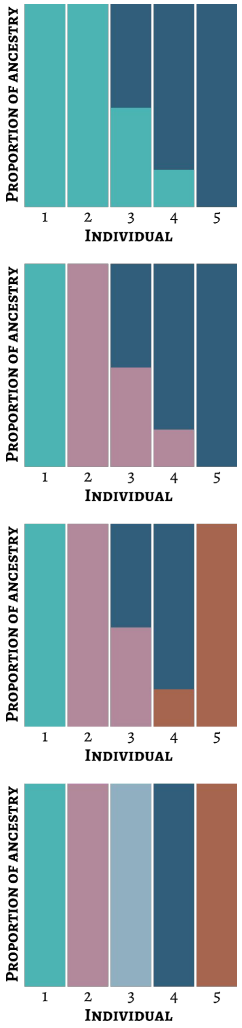



Two clusters
(*K* = 2)



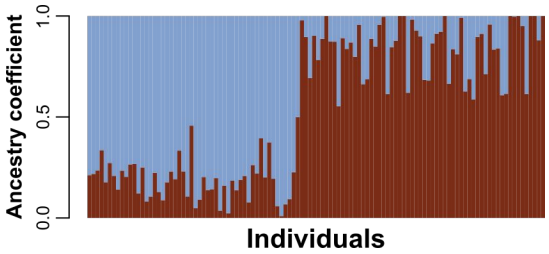



Range of clusters ($K$ = 2–5)

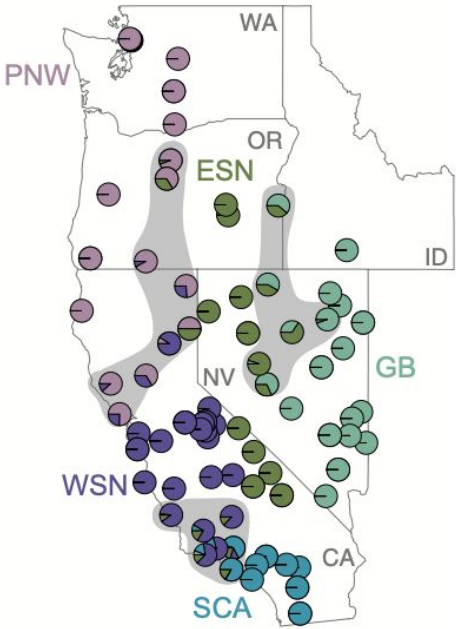


# Visualizing population structure results

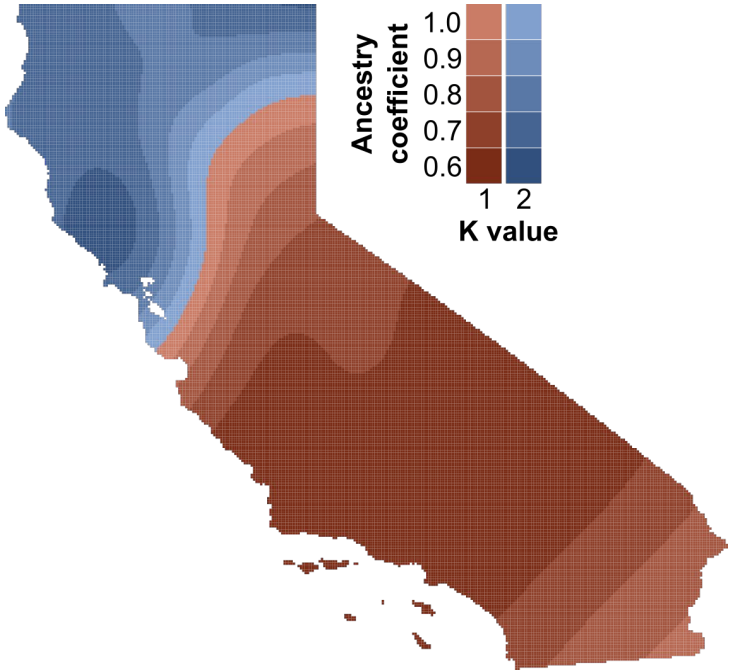
## *Structure plot*




## *Pie charts*




## *Map of kriged ancestry coefficients*




Bouzid et al. (2022)

## Methods to infer population structure

STRUCTURE (Pritchard et al. 2000, *Genetics*)
fastSTRUCTURE, CLUMPP, FRAPPE, GENELAND
**ADMIXTURE, TESS, Ohana**

All of these methods estimate allele frequencies, are sensitive to sampling (individuals and loci), and based on genetic assumptions of HWE and LE between loci

**Ensure your sampling is thorough and even, sites are unlinked, test a range of cluster numbers, and have a good knowledge of your study system!**

# EXERCISE

https://thewanglab.github.io/algatr/articles/TESS_vignette.html

# Process input data files

```
# Load example dataset
load_algatr_example()

# Convert vcf to genotype matrix
liz_dosage <- vcf_to_dosage(liz_vcf)
```

VCF (object or file path)

See what **`liz_dosage`** looks like using **`head(liz_dosage)`**

# Process input data files

```
# First, create a grid for kriging
krig_raster <- raster::aggregate(CA_env[[1]],
                                 fact = 6)
```

Raster layer

**`fact`** = factor to aggregate cells

See what happens when you change **`fact`** and replot

```
terra::plot(krig_raster,
            col = mako(100),
            axes = FALSE)
```

Raster layer

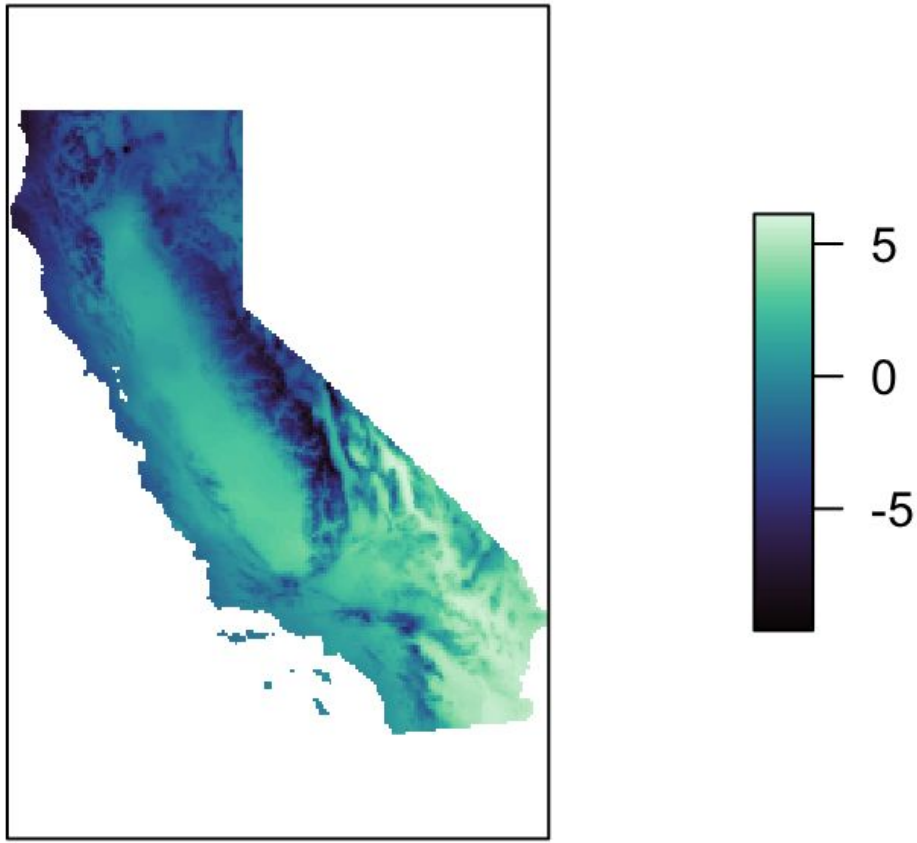
**`col`** = color palette

**`axes`** = whether to display axes

# Process input data files

```
terra::plot(krig_raster,
            col = mako(100),
            axes = FALSE)
```

See how this plot compares to one made using **`CA_env[[1]]`**



# Run TESS with a set *K*

```
# Run TESS with a set K
tess3_K3 <- tess3r::tess3(x = liz_dosage,
                          coord = as.matrix(liz_coords),
                          K = 3,
                          ploidy = 2)
```

**x** = dosage matrix

**coord** = sampling coordinates

**K** = *K* value to run TESS

**ploidy** = ploidy of genetic data

**tess3** = function to run TESS (within the tess3r package)

Take a look at the **tess3_K3** object; what's contained in it?

# Run TESS with *K* selection

```
tess3_result <- tess_ktest(liz_dosage,
                           liz_coords,
                           Kvals = 1:10,
                           ploidy = 2,
                           K_selection = "auto")
```

Dosage matrix

Sampling coordinates

**`Kvals`** = range of *K* values to test

**`ploidy`** = ploidy of genetic data

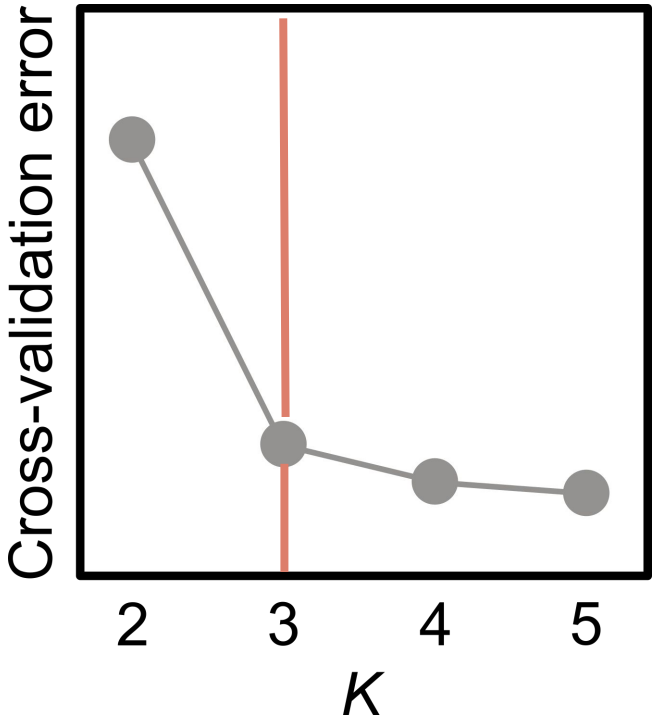
**`K_selection`** = type of *K* selection

*You can provide a single value to the **`Kvals`** argument and it won't perform K-selection but will run TESS*

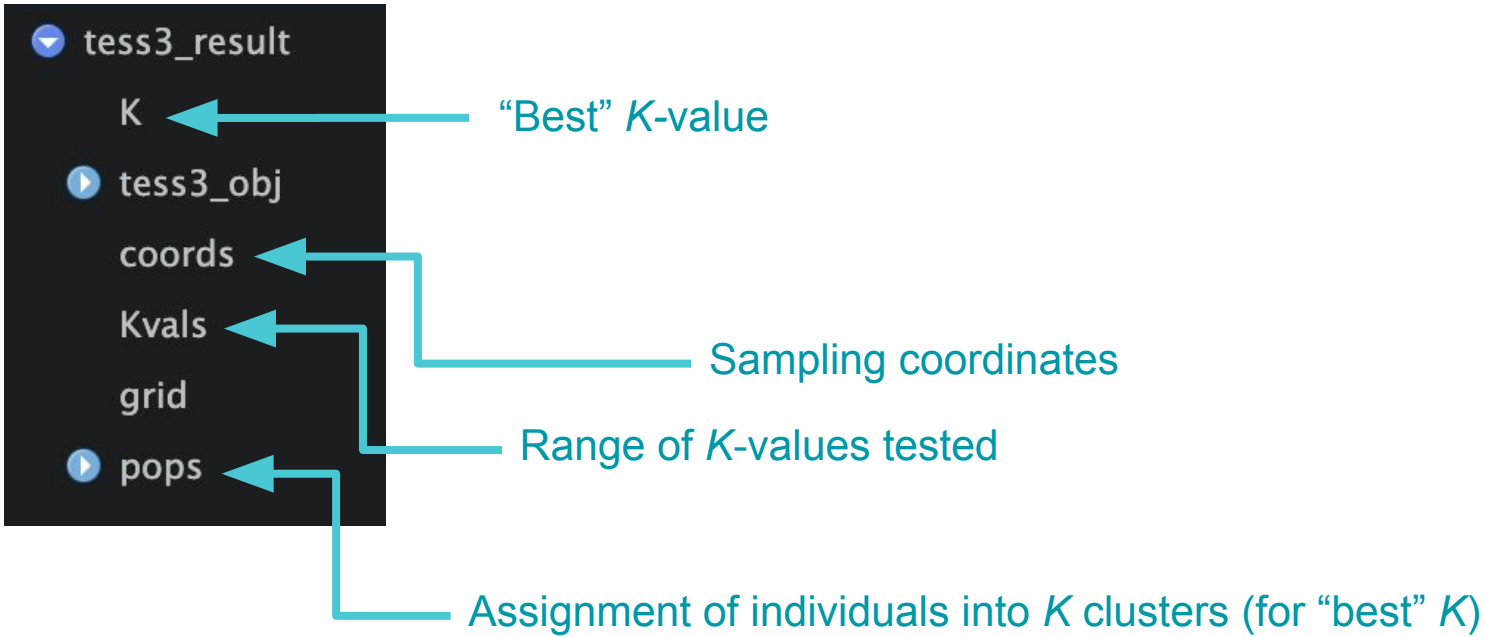
# Run TESS with *K* selection

```
tess3_result <- tess_ktest(liz_dosage,
                           liz_coords,
                           Kvals = 1:10,
                           ploidy = 2,
                           K_selection = "auto")
```

See what happens when you set
**K_selection = “manual”**



# Interpret TESS results




What is in the **`tess3_obj`** element of this list?

# Interpret TESS results


```
# Get TESS object and best K from results
tess3_obj <- tess3_result$tess3_obj
bestK <- tess3_result[["K"]]
```

Extract TESS object and best *K* value from TESS results

```
# Get Qmatrix with ancestry coefficients
qmat <- tess3r::qmatrix(tess3 = tess3_obj,
                        K = bestK)
```

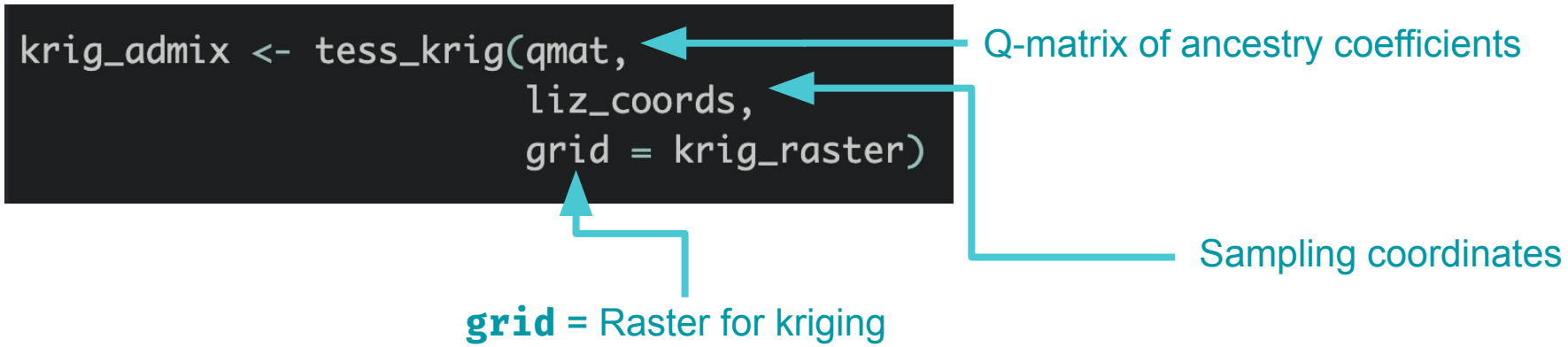
**`tess3`** = TESS object

**`K`** = *K* value to extract ancestry coefficients for

Take a look at the Q-matrix by doing **`head(qmat)`**

# Krige ancestry coefficients using your raster

```
krig_admix <- tess_krig(qmat,
                        liz_coords,
                        grid = krig_raster)
```

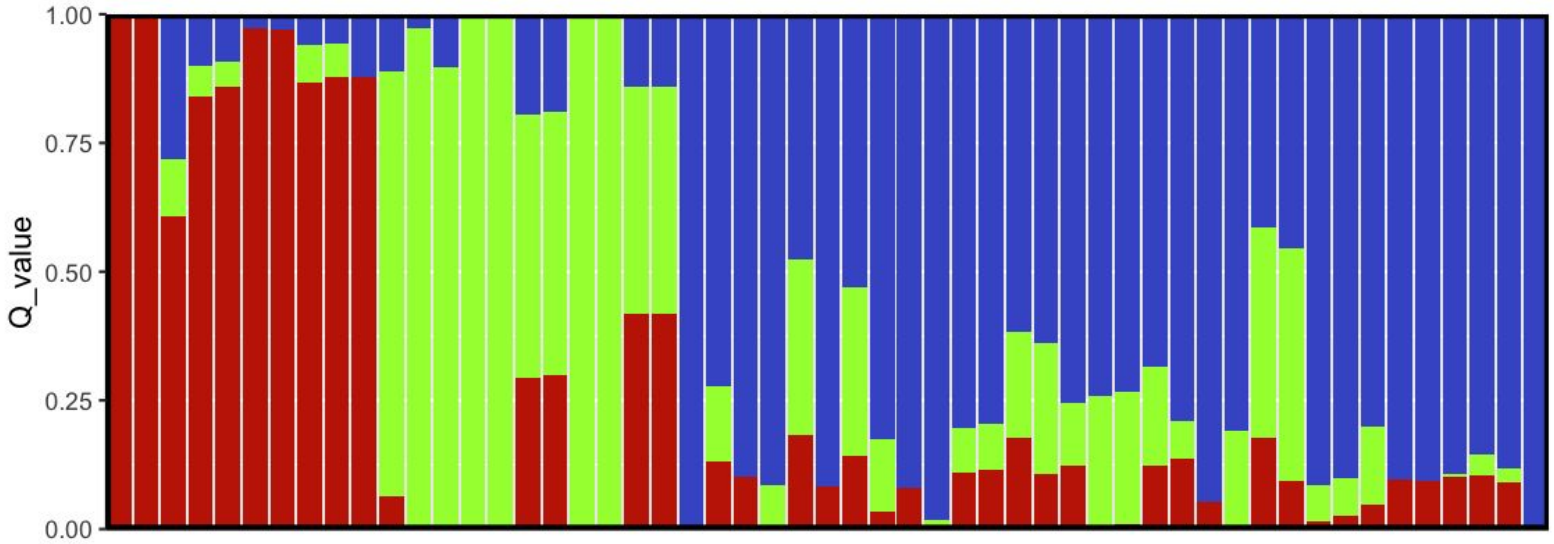



See what happens when you plot **krig_admix** (using **terra::plot**) compared to your earlier plot of **krig_raster**

# Plot TESS results

```
tess_ggbarplot(qmat,
               legend = FALSE,
               sort_by_Q = TRUE)
```

See what happens when you set **`sort_by_Q()`** to FALSE



# Plot TESS results

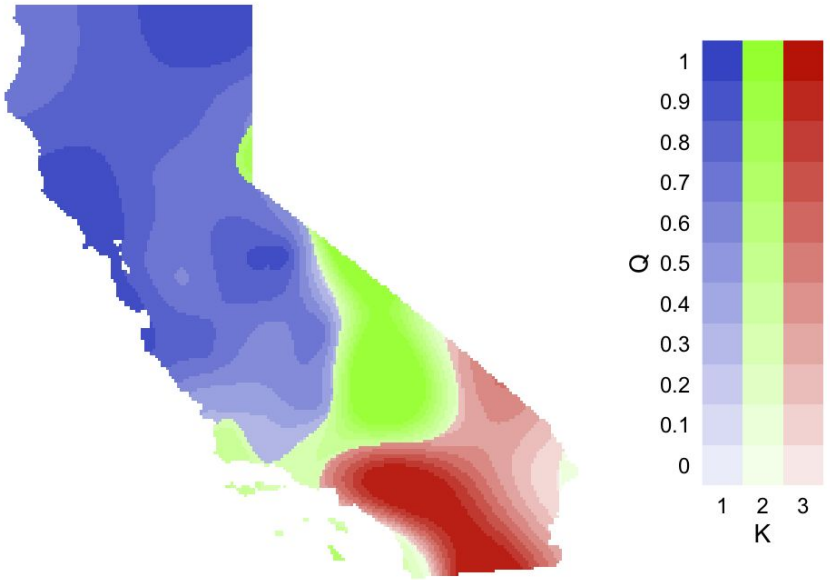
```
tess_ggplot(krig_admix,
            plot_method = "maxQ")
```

Raster with kriged ancestry coefficients

**plot_method** = which ancestry coefficients to colorize map using

# Plot TESS results

```
tess_ggplot(krig_admix,
            plot_method = "maxQ")
```

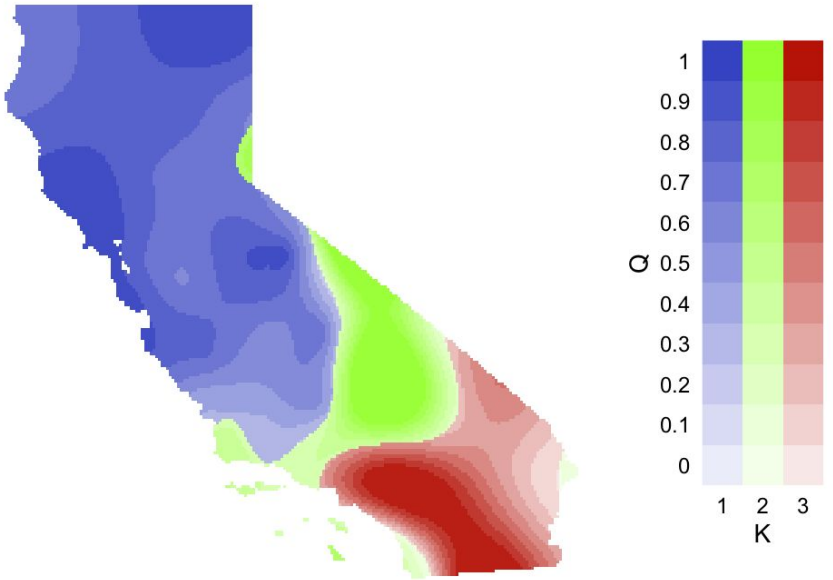



See what happens when you change the **`plot_method`** argument in the **`tess_ggplot()`** function

See what happens when you specify **`liz_coords`** using the **`coords`** argument in the **`tess_ggplot()`** function

# Plot TESS results

```
tess_ggplot(krig_admix,
            plot_method = "maxQ")
```




```
Warning messages:
1: In crs_check(coords, grid) :
  No CRS found for the provided coordinates. Make sure the coordinates and the raster have the
same projection (see function details or vignette)
2: [mask] CRS do not match
```

https://thewanglab.github.io/algatr/articles/TESS_vignette.html

# Exercise

1. Load the example dataset
2. Process genetic data:
   a. Convert vcf to dosage using `vcf_to_dosage()`
3. Process environmental data:
   a. Create a raster for mapping (kriging) using `raster::aggregate()`
4. Select the “best” number of *K* clusters and run TESS using `tess_ktest()`
5. Visualize results with a barplot using `tess_ggbarplot()`
6. Visualize results with a kriged map using `tess_krig()` and `tess_ggplot()`