

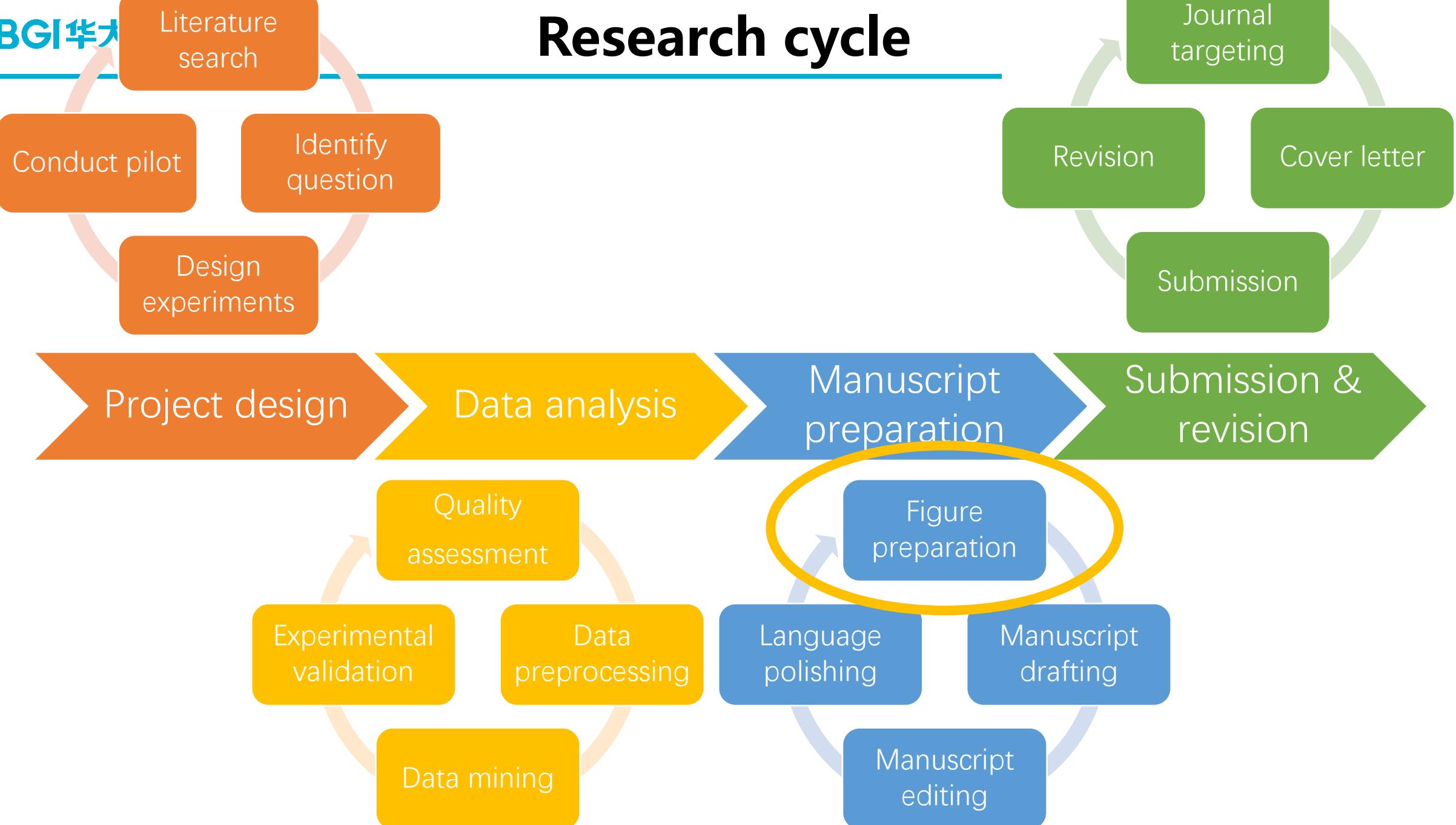
Scientific figure

魏桐

11/2/2022

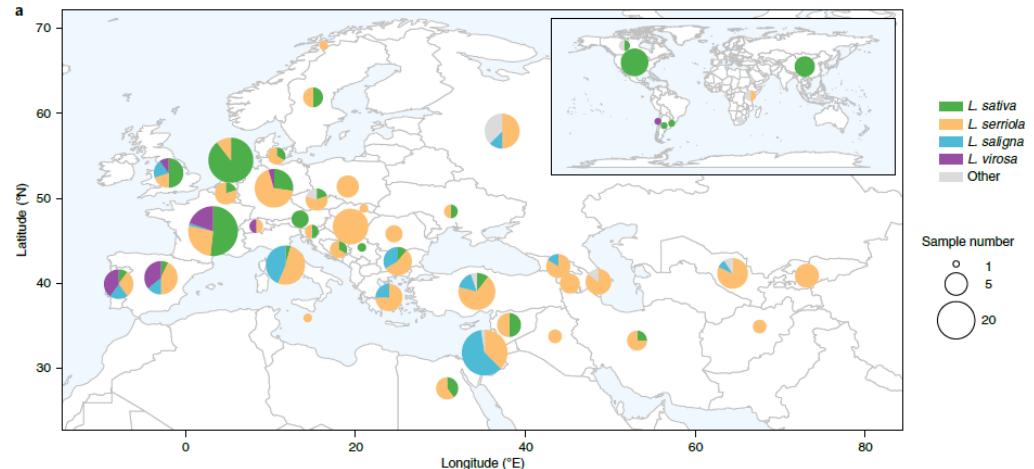
- What is scientific illustration/science figure?
- Plot types
- An exemplary case

Research cycle



What is scientific figure

Provide **an efficient way to communicate** your research



VS

Plant materials and sequencing. The collection of *Lactuca* SSD lines (<http://www.wur.eu/cgnsc002>) used in this study comprises a core subset of the regular collection of the Centre for Genetic Resources, the Netherlands (CGN) and includes all crop types of cultivated lettuce and main wild relatives used in plant breeding³². The total study set of 445 SSD lines included 131 cultivated lettuce (*L. sativa*) accessions collected worldwide, 201 *L. serriola* accessions, 57 *L. saligna* accessions, 37 *L. virosa* accessions and 19 lines from another eight *Lactuca* species (Supplementary Table 1).

Part I

Plot types

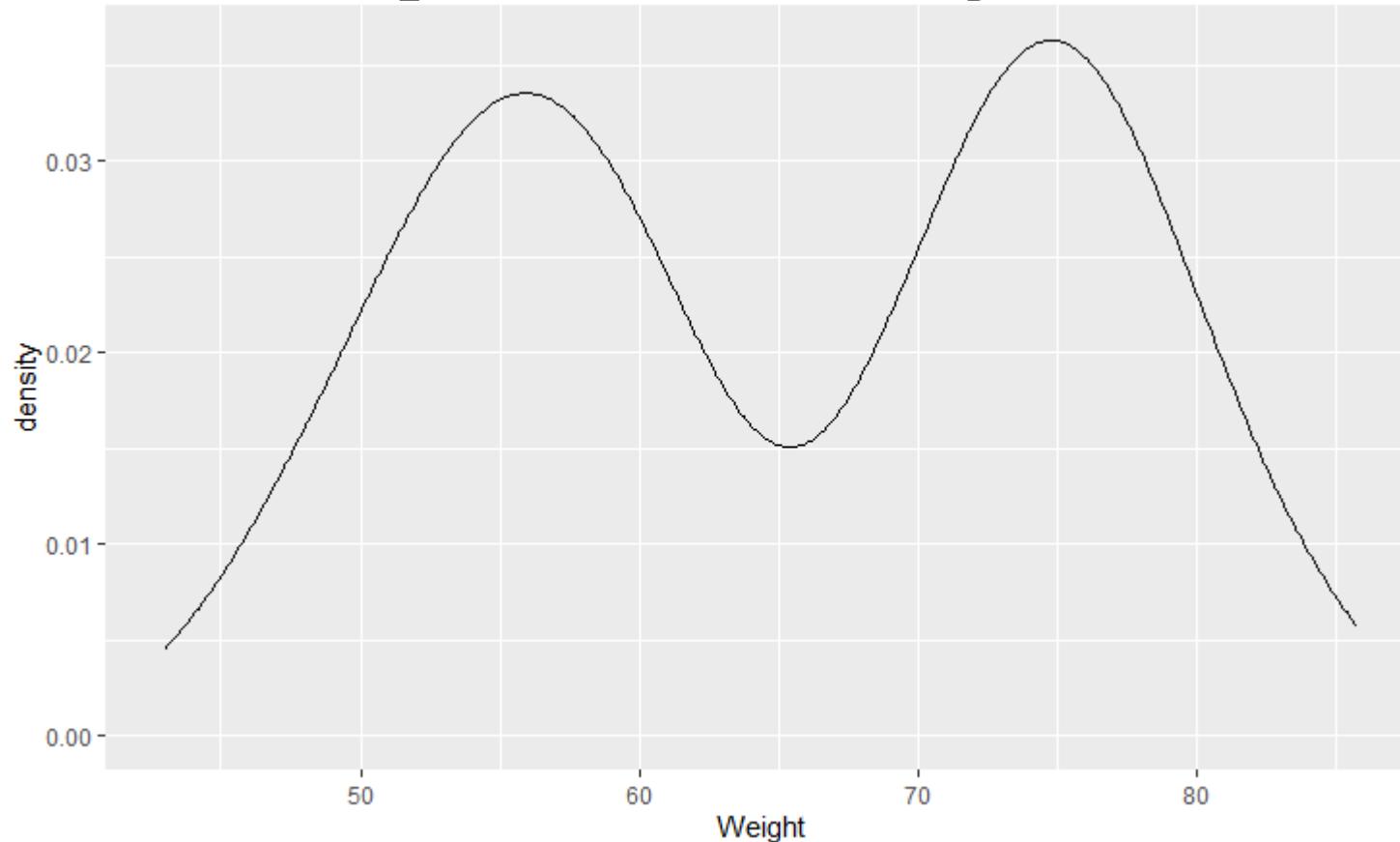
- Data types: numerical, categorical & ordinal
- 1-D* numerical data: density lines (summary stats)
- 1-D categorical: barplot of counts
- 2-D numerical: scatterplot
- 2-D numerical vs categorical: boxplot/violin
- 2-D categorical: heatmap
- 3-D numerical/categorical: scatterplot w/ other aesthetics
- >3-D numerical/categorical: facet

A simulated data frame

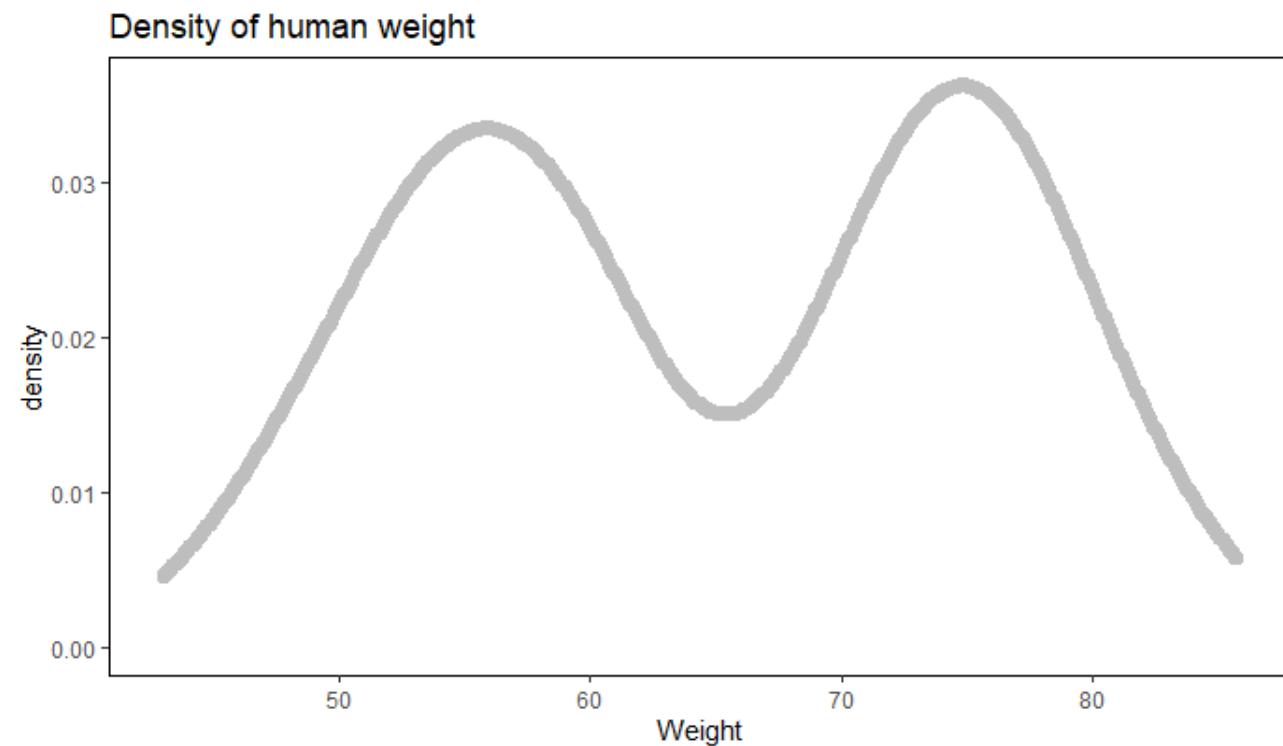
```
dat <- data.frame(Height = c(rnorm(100, 175, 10),  
                           rnorm(100, 175, 10)),  
                  Weight = c(rnorm(100, 75, 5),  
                             rnorm(100, 55, 5)),  
                  Sex = rep(c("Male", "Female"),  
                            each = 100),  
                  Age = rep(c("Young", "Old"),  
                            each = 50, times = 2))
```

```
> head(dat)  
#> #>   Height  Weight Sex   Age  
#> 1 180.4326 75.56678 Male Young  
#> 2 166.4858 73.68677 Male Young  
#> 3 179.8365 75.75200 Male Young  
#> 4 178.4043 77.01459 Male Young  
#> 5 176.3072 81.62829 Male Young  
#> 6 185.1516 73.08120 Male Young
```

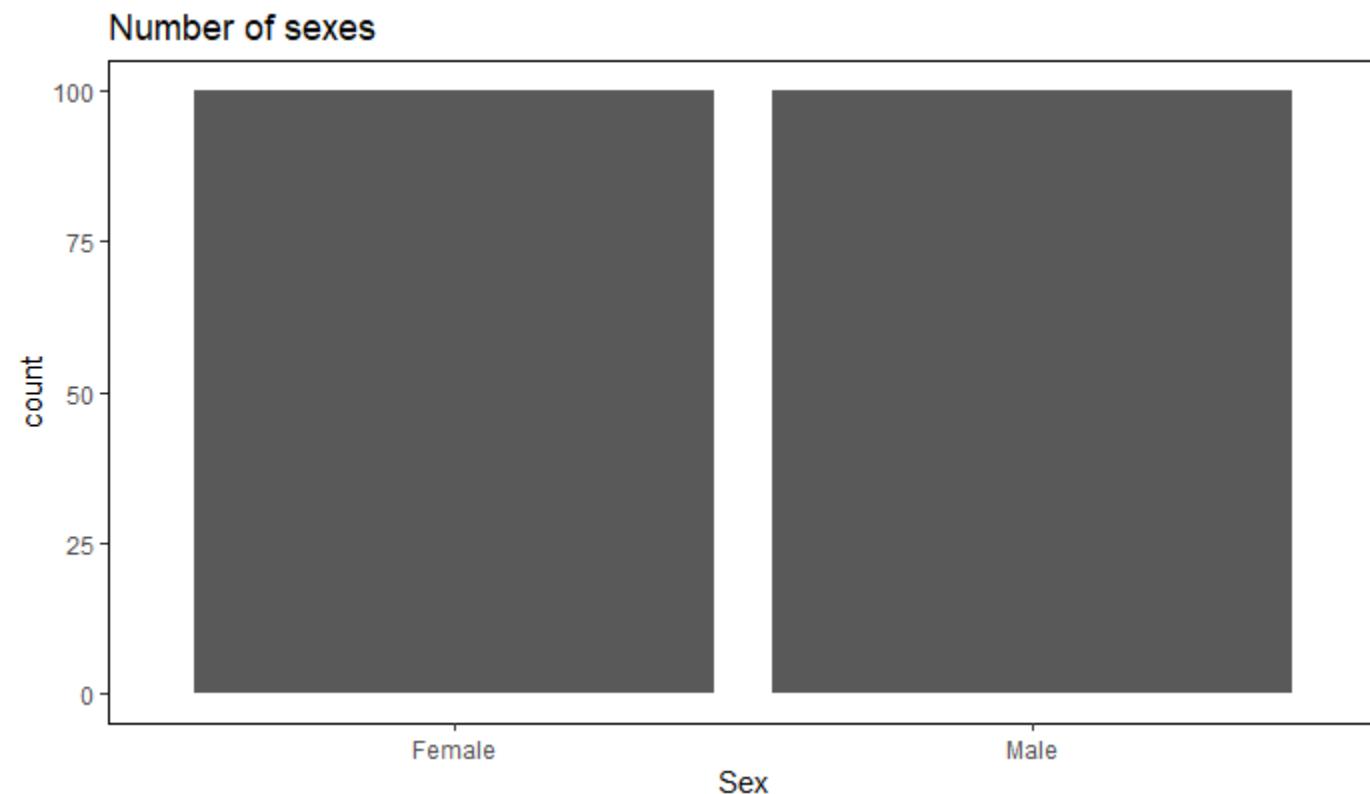
```
library(ggplot2)
g <- ggplot(dat)
g + geom density(aes(x = Weight)) # density plot
```



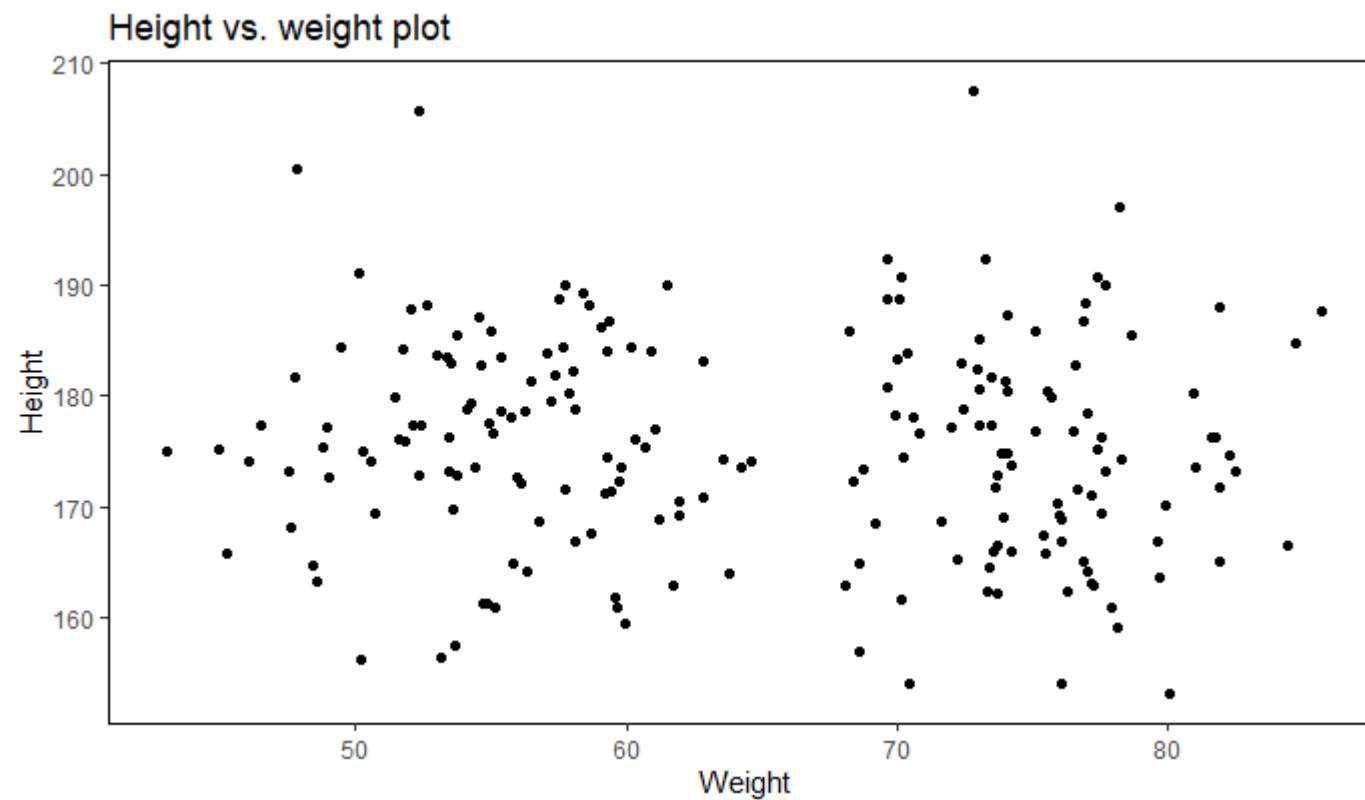
```
g + geom_density(aes(x = Weight),  
                  color = "grey", size = 3) +  
  labs(title = "Density of human weight") +  
  theme(panel.border = element_rect(color = "black", fill = NA),  
        panel.background = element_rect(fill = "white"))
```



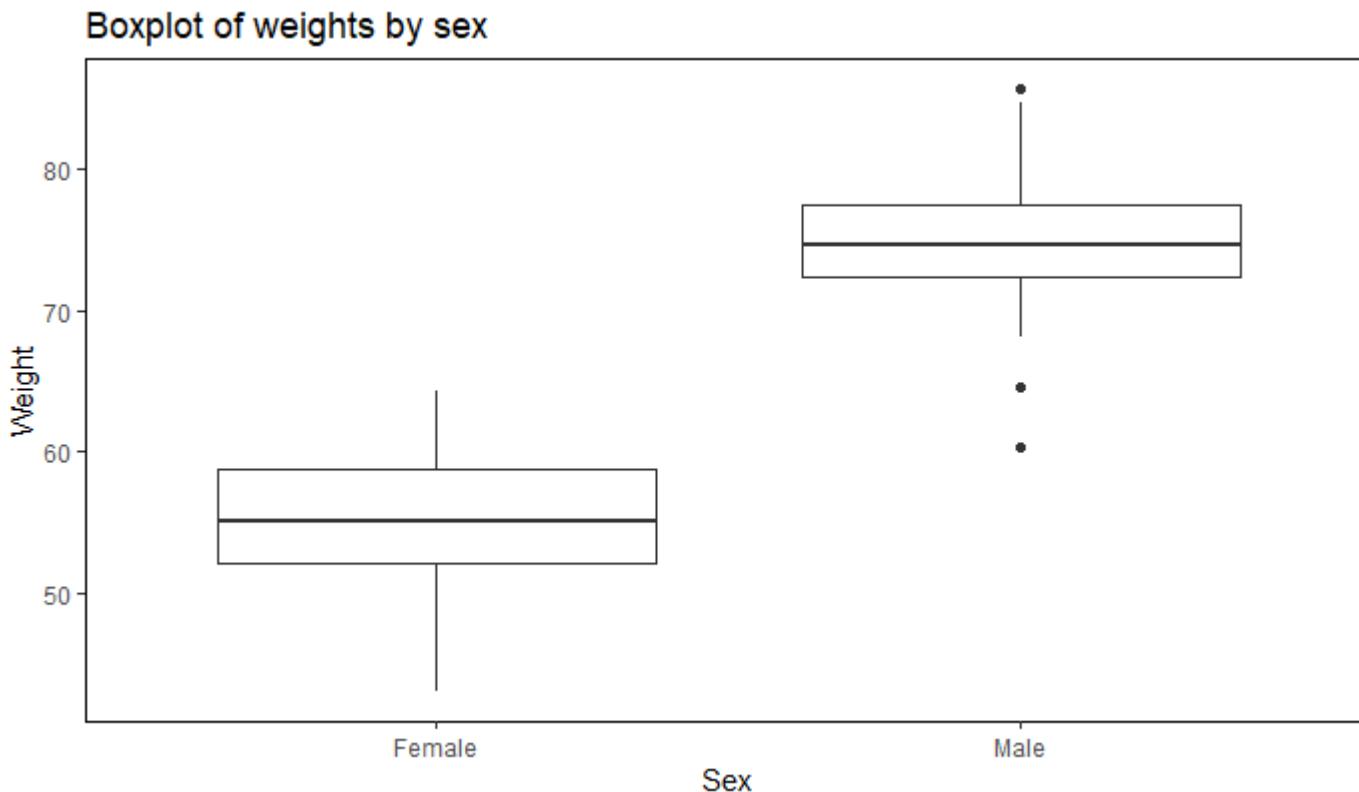
```
g <- ggplot(dat) +  
  theme(panel.border = element_rect(color = "black", fill = NA),  
        panel.background = element_rect(fill = "white"))  
g + geom_bar(aes(x = Sex)) + labs(title = "Number of sexes")
```



```
g + geom_point(aes(x = Weight, y = Height)) +  
  labs(title = "Height vs. weight plot")
```



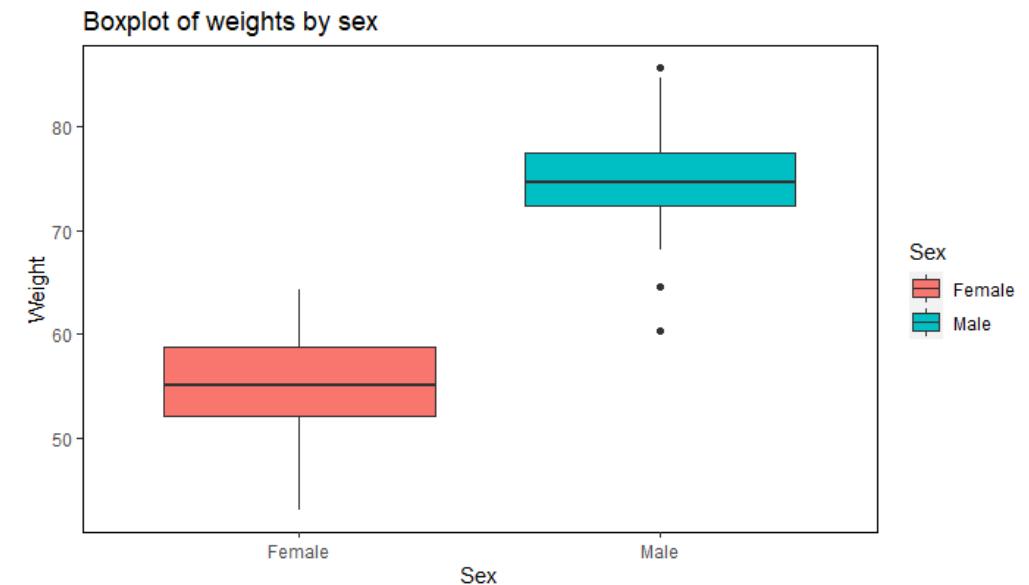
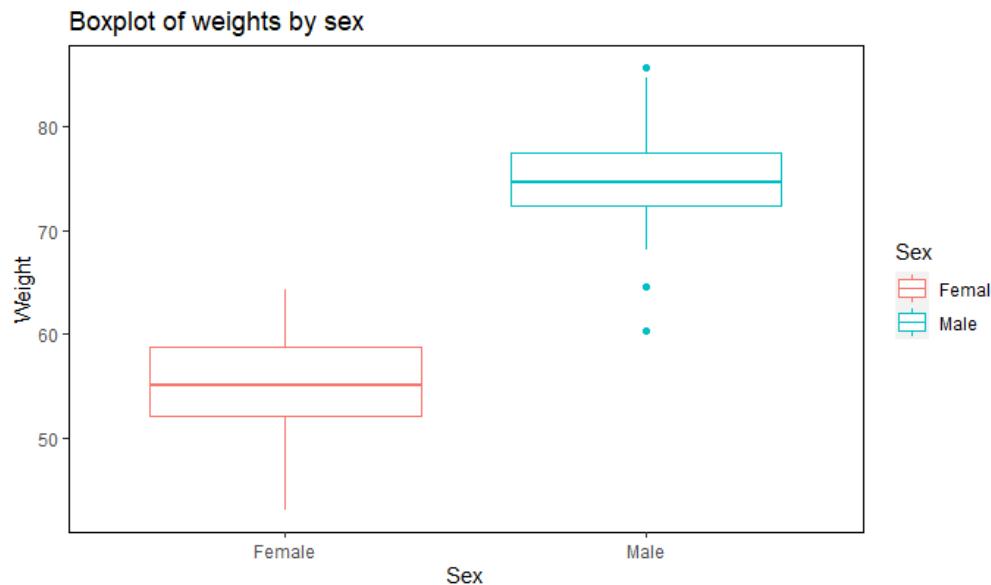
```
g + geom_boxplot(aes(x = Sex, y = Weight)) +  
  labs(title = "Boxplot of weights by sex")
```



Parameter: color vs. fill

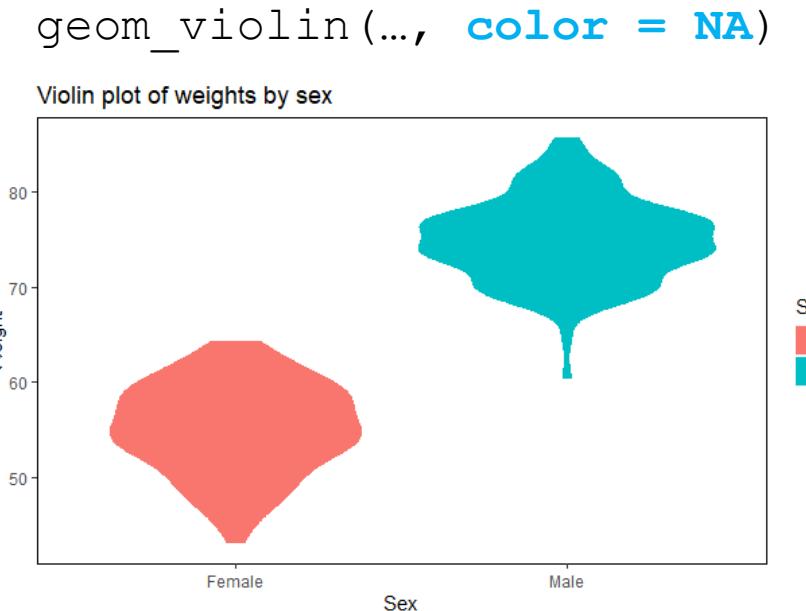
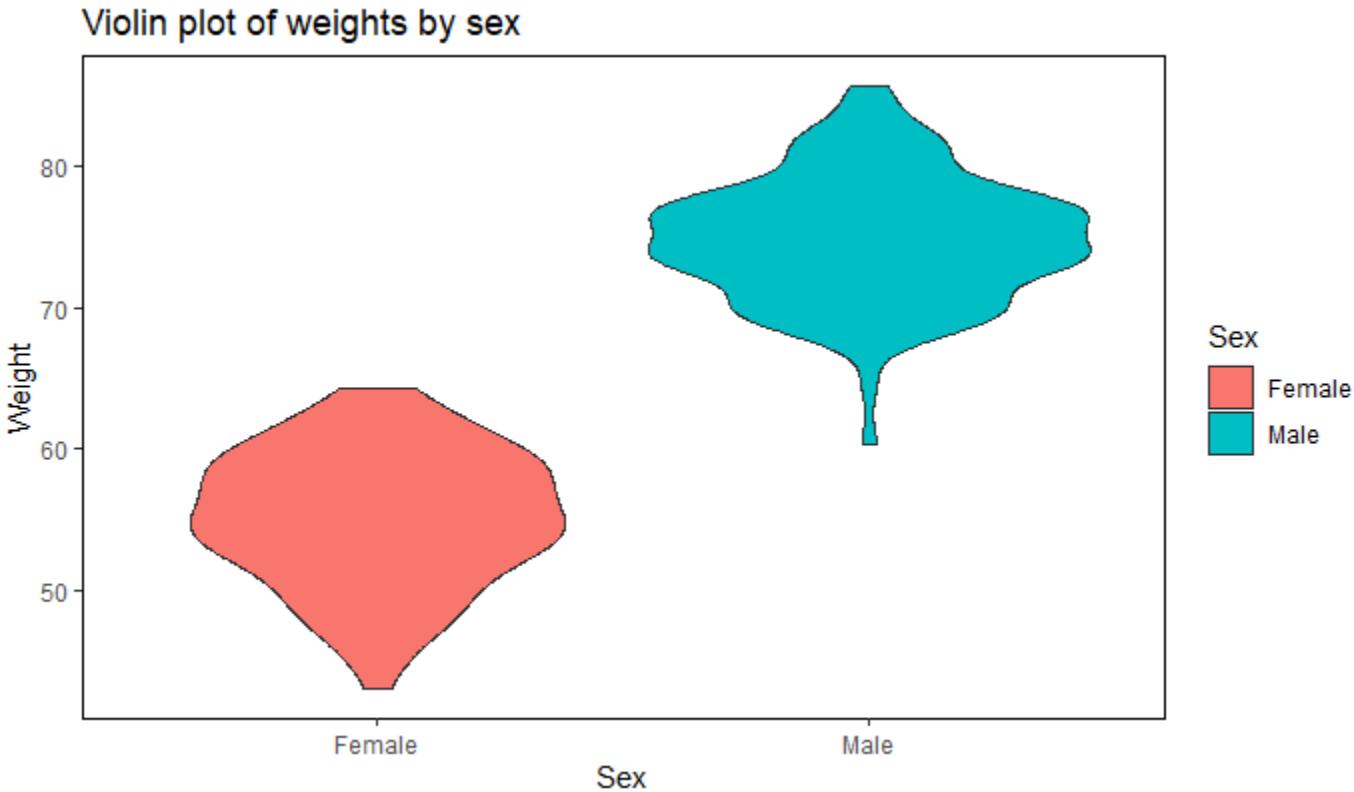
```
g + geom_boxplot(aes(Sex, Weight, color = Sex) ) +  
  labs(title = "Boxplot of weights by sex" )
```

```
g + geom_boxplot(aes(Sex, Weight, fill = Sex) ) +  
  labs(title = "Boxplot of weights by sex" )
```



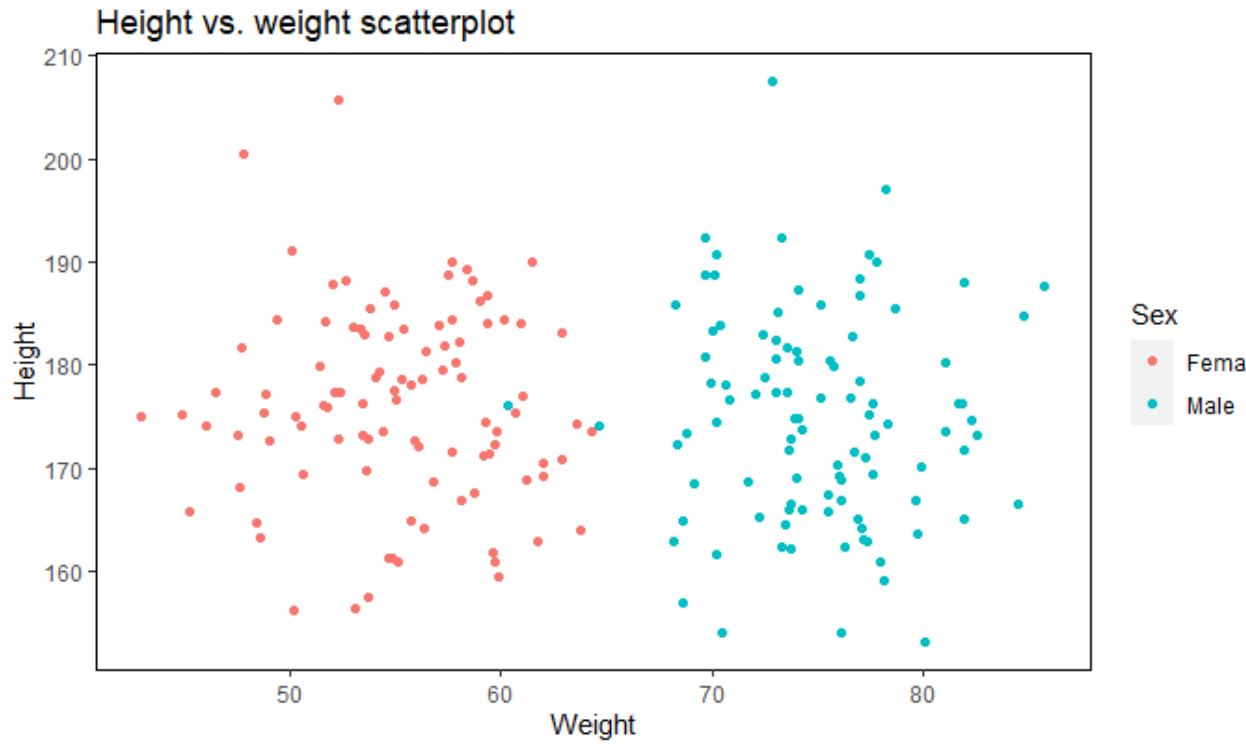
Violin plot

```
g + geom_violin(aes(x = Sex, y = Weight, fill = Sex)) +  
  labs(title = "Violin plot of weights by sex")
```

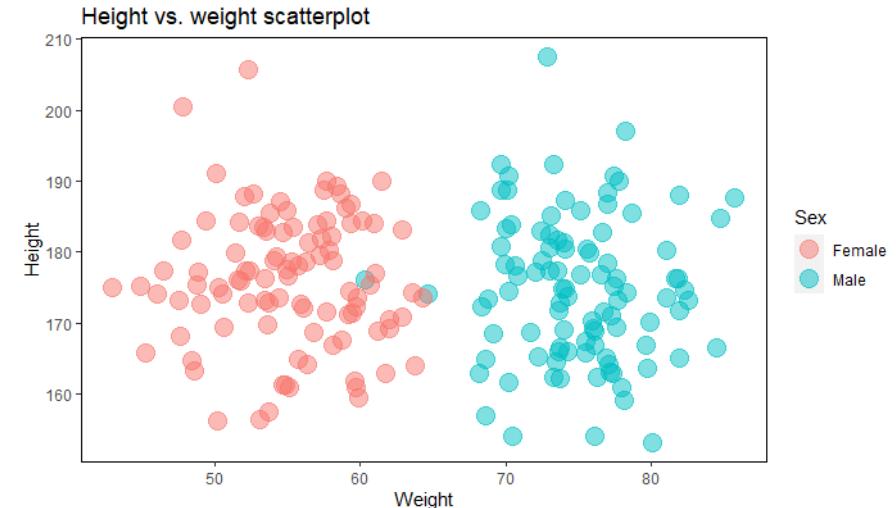


Scatterplot w/ a 3rd variable

```
g + geom_point(aes(x = Weight, y = Height, color = Sex)) +  
  labs(title = "Height vs. weight scatterplot")
```

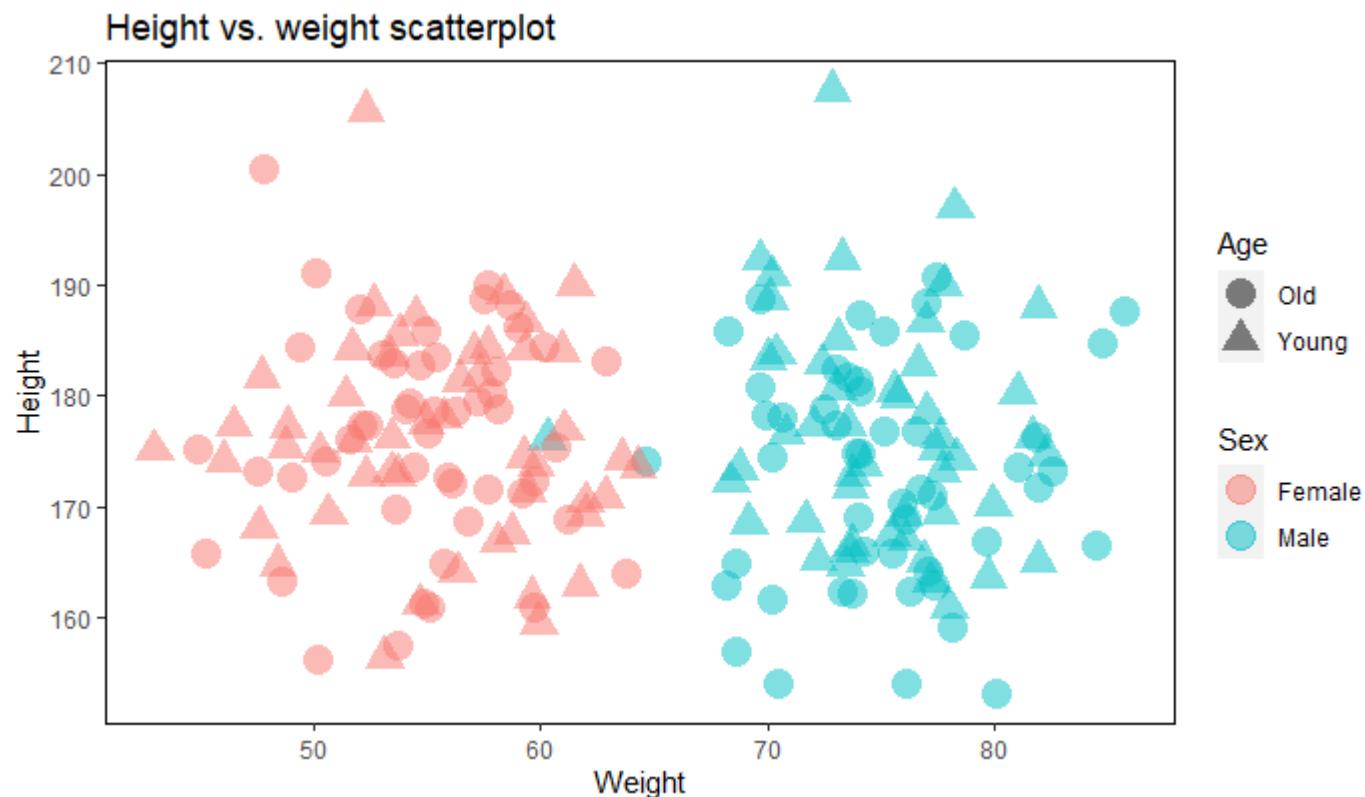


geom_violin(..., size = 5, **alpha = 0.5**)

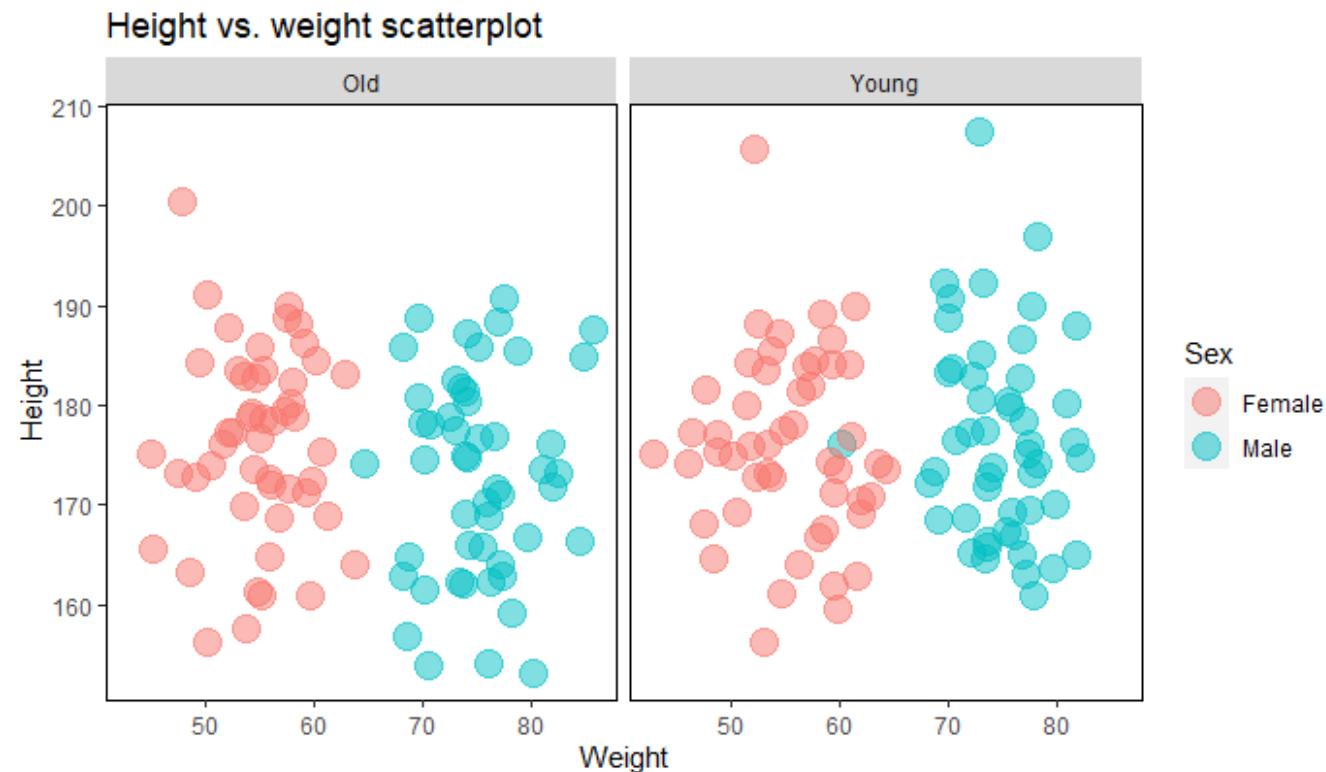


4-D data: aesthetics

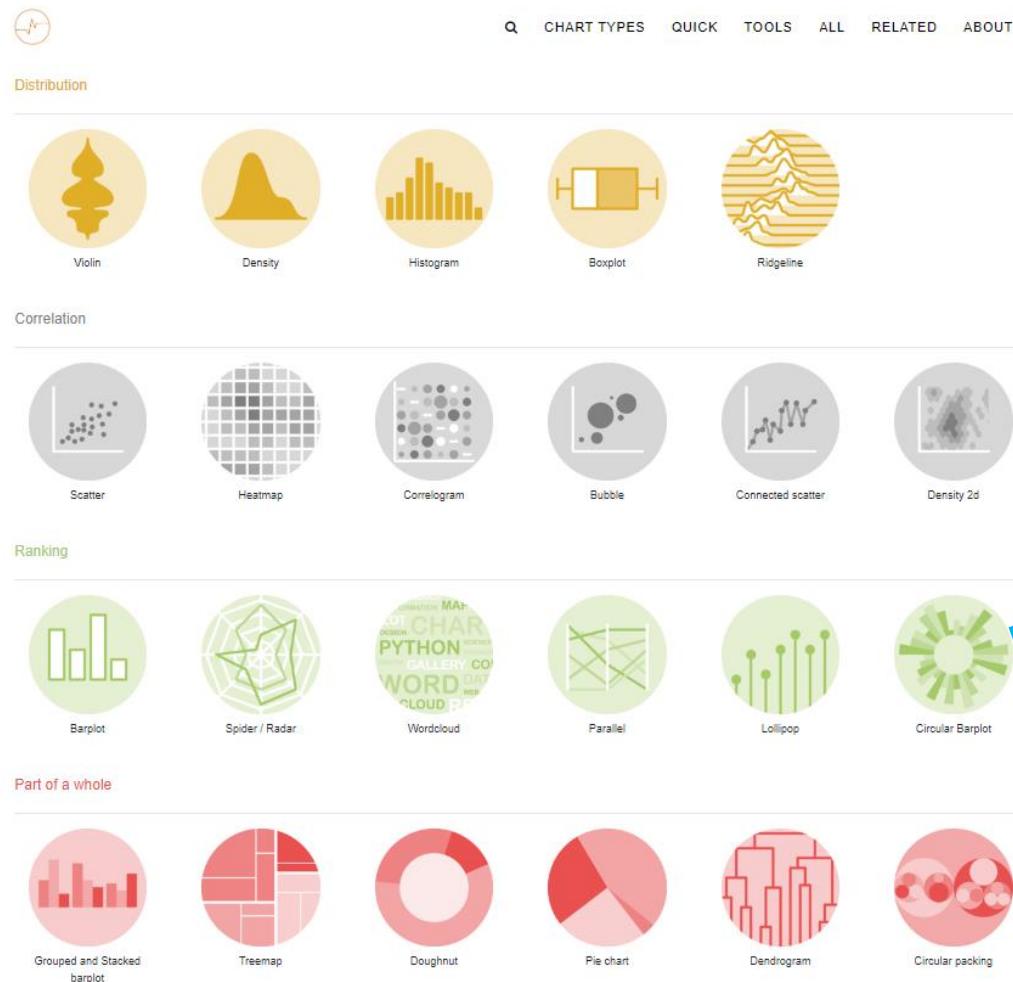
```
g + geom_point(aes(x = Weight, y = Height,  
                    color = Sex, shape = Age),  
                    size = 5, alpha = .5) +  
  labs(title = "Height vs. weight scatterplot")
```



```
g + geom_point(aes(x = Weight, y = Height, color = Sex),  
               size = 5, alpha = .5) +  
  labs(title = "Height vs. weight scatterplot") +  
facet_wrap(facets = vars(Age))
```



More plots in R gallery



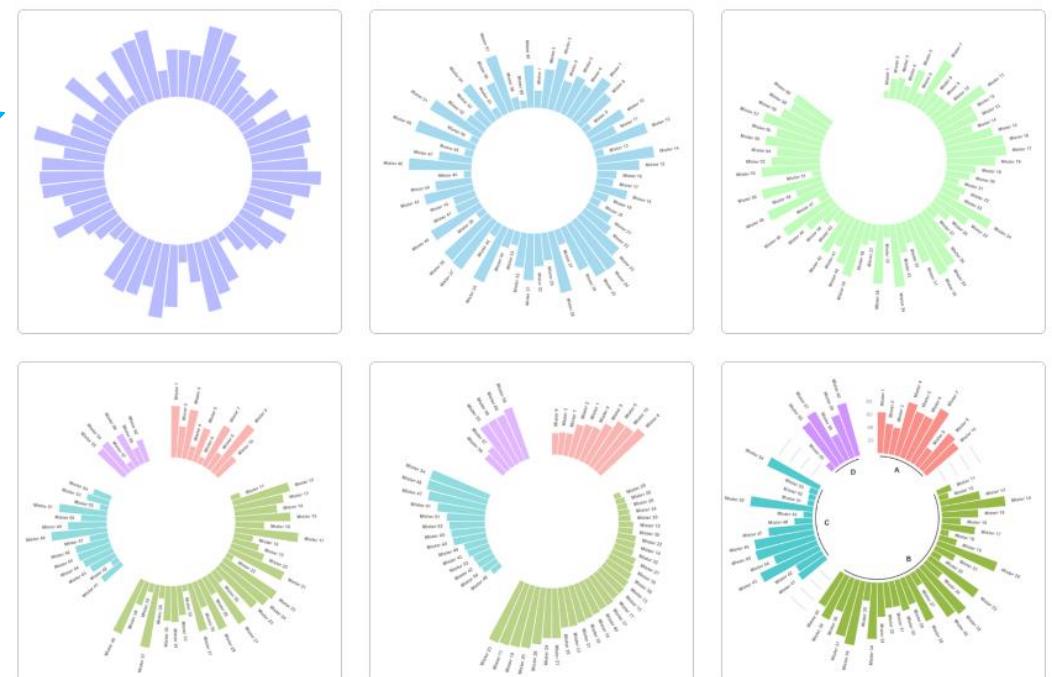
Circular barplot



This is the [circular barplot](#) section of the gallery, a variation of the well known [barplot](#). Note that even if visually appealing, [circular barplot](#) must be used with care since groups [do not share the same Y axis](#). It is very adapted for cyclical data though. Visit [data-to-viz.com](#) for more info.

STEP BY STEP

Here is a set of examples leading to a proper circular barplot, step by step. The first [most basic circular barchart](#) shows how to use `coord_polar()` to make the barchart circular. Next examples describe the next steps to get a proper figure: [gap](#) between groups, [labels](#) and customization.



Part II

A case study

Case I. Lettuce domestication

nature
genetics

ARTICLES

<https://doi.org/10.1038/s41588-021-00831-0> Check for updates

Whole-genome resequencing of 445 *Lactuca* accessions reveals the domestication history of cultivated lettuce

Tong Wei^{1,11}, Rob van Treuren^{1,2,11}✉, Xinjiang Liu^{1,11}, Zhaowu Zhang^{1,3}, Jiongjiong Chen⁴, Yang Liu¹, Shanshan Dong⁵, Peinan Sun⁴, Ting Yang¹, Tianming Lan^{1,6}, Xiaogang Wang⁷, Zhouquan Xiong⁷, Yaqiong Liu⁸, Jinpu Wei⁸, Haorong Lu^{1,8}, Shengping Han⁸, Jason C. Chen⁸, Xuemei Ni¹, Jian Wang^{1,9}, Huanming Yang^{1,9}, Xun Xu^{1,10}, Hanhui Kuang⁴, Theo van Hintum², Xin Liu^{1,11}✉ and Huan Liu¹✉

Lettuce (*Lactuca sativa*) is an important vegetable crop worldwide. Cultivated lettuce is believed to be domesticated from *L. serriola*; however, its origins and domestication history remain to be elucidated. Here, we sequenced a total of 445 *Lactuca* accessions, including major lettuce crop types and wild relative species, and generated a comprehensive map of lettuce genome variations. In-depth analyses of population structure and demography revealed that lettuce was first domesticated near the Caucasus, which was marked by loss of seed shattering. We also identified the genetic architecture of other domestication traits and wild introgressions in major resistance clusters in the lettuce genome. This study provides valuable genomic resources for crop breeding and sheds light on the domestication history of cultivated lettuce.

Case I. Lettuce population

- As a vegetable crop
 - A rich source of vitamin K and vitamin A, and a moderate source of folate and iron
 - Consumed worldwide
- As a model Asteraceae plant
 - A short life cycle; easy for transformation
 - Well-maintained germplasms
 - Various agronomic traits
 - Potential bioreactor



var. *longifoliaf*



var. *angustanairish*

Case I. Lettuce crop types



Crisp



Butterhead



Cos



Stalk



Oilseed



L. serriola
proposed wild progenitor

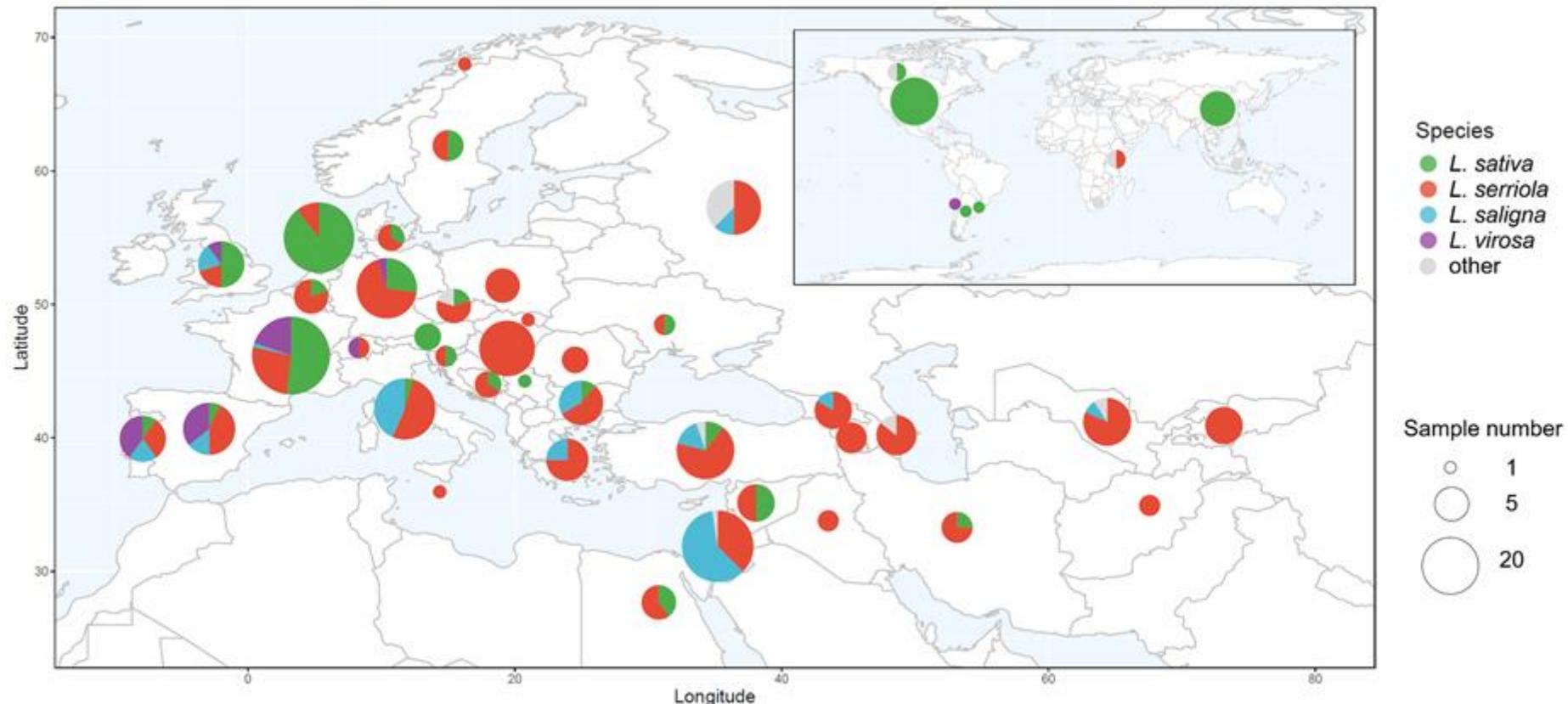
Case I. Scientific questions



1. **Where** was lettuce domesticated?
2. **When** was lettuce domesticated?
3. What are the **key events** during lettuce domestication?
4. What are the **genetic determinants** for the domestication traits?

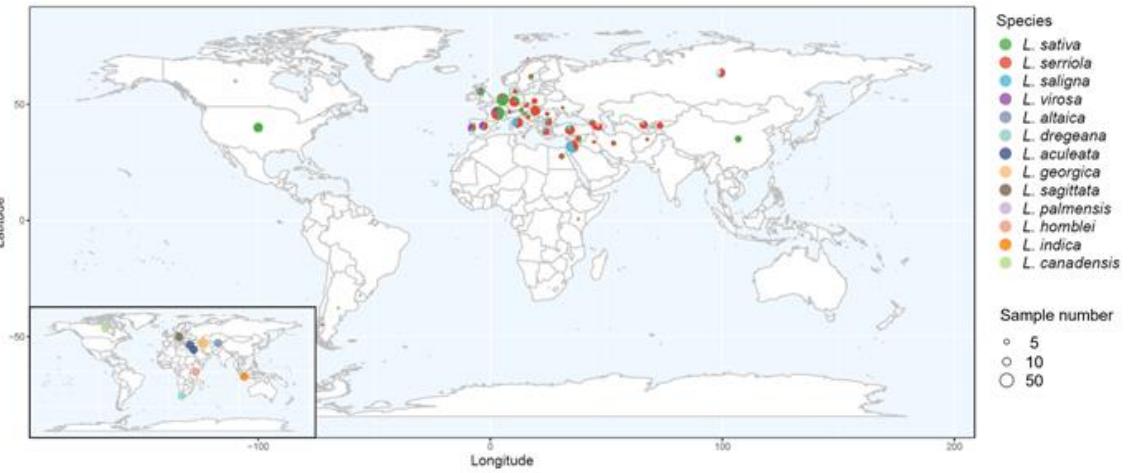
Case I. Global representation in world map

445 lines from 47 countries, mainly composed of GP breeding materials of *L. sativa*, *L. serriola*, *L. saligna* and *L. virosa*

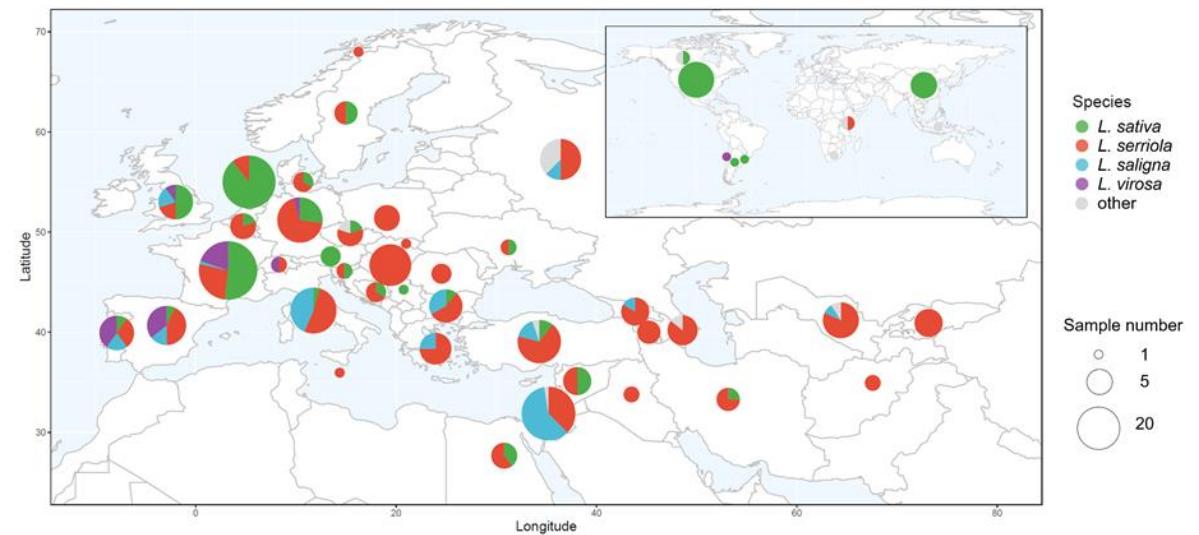


Tips: emphasize on data

a

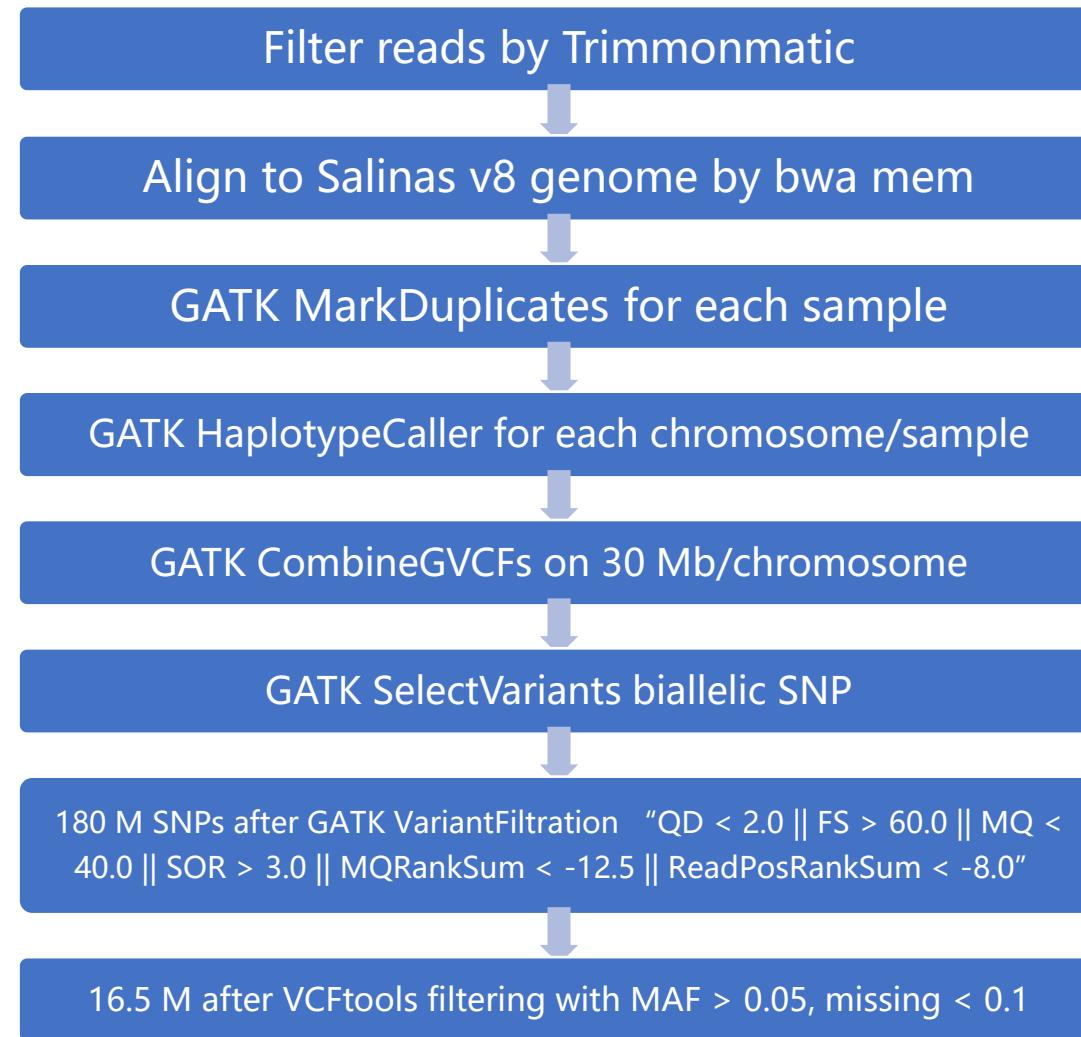


In submission

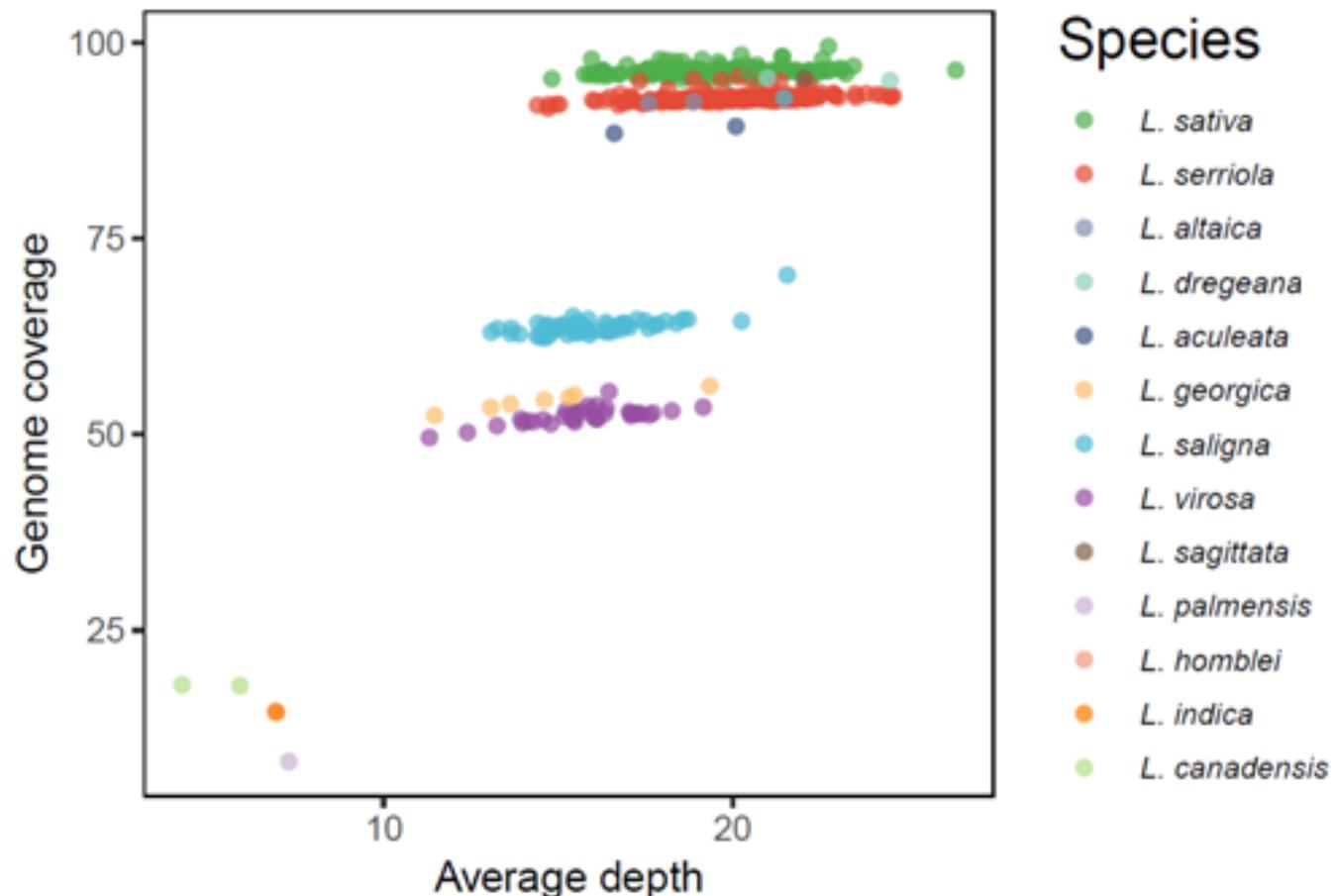


In revision

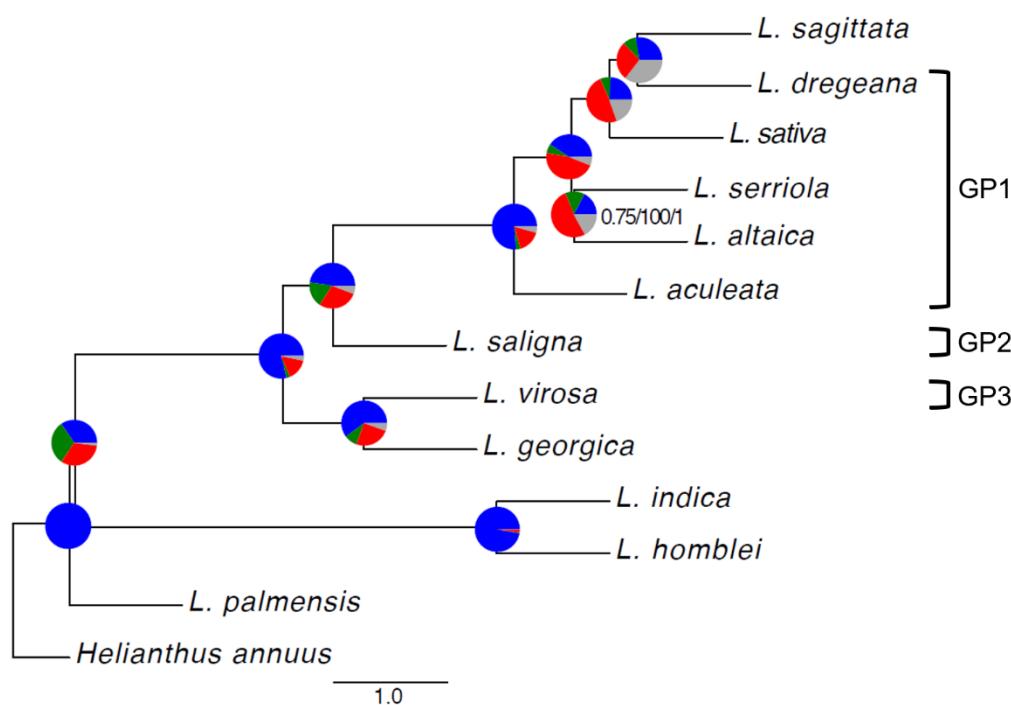
Case I. Pipeline in flow chart



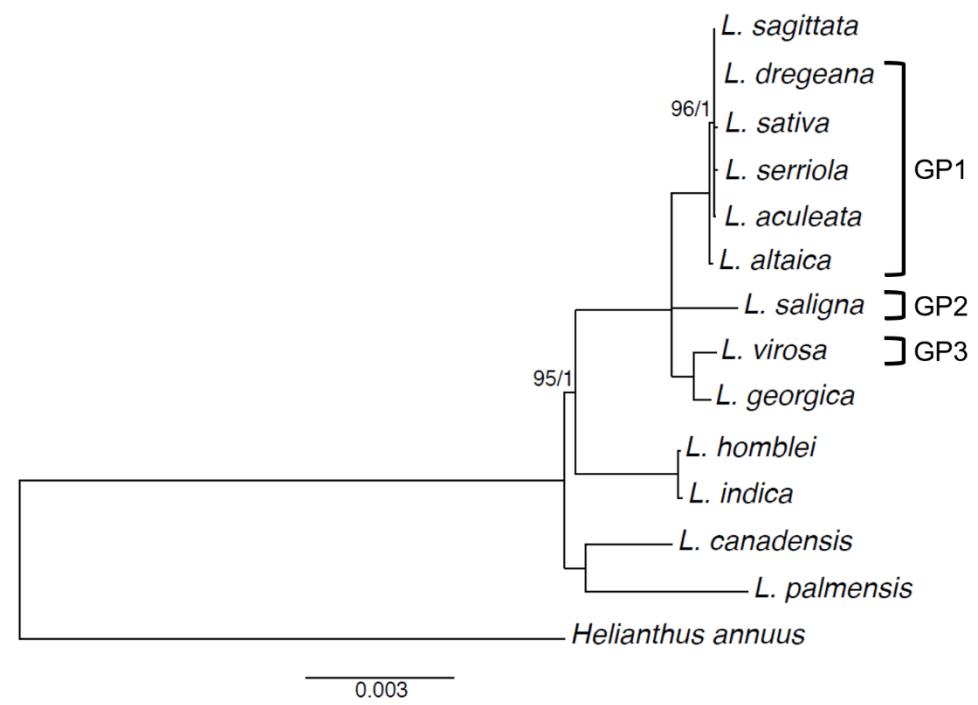
Case I. Read alignment in scatter plot



Case I. Phylogeny in tree plot

a

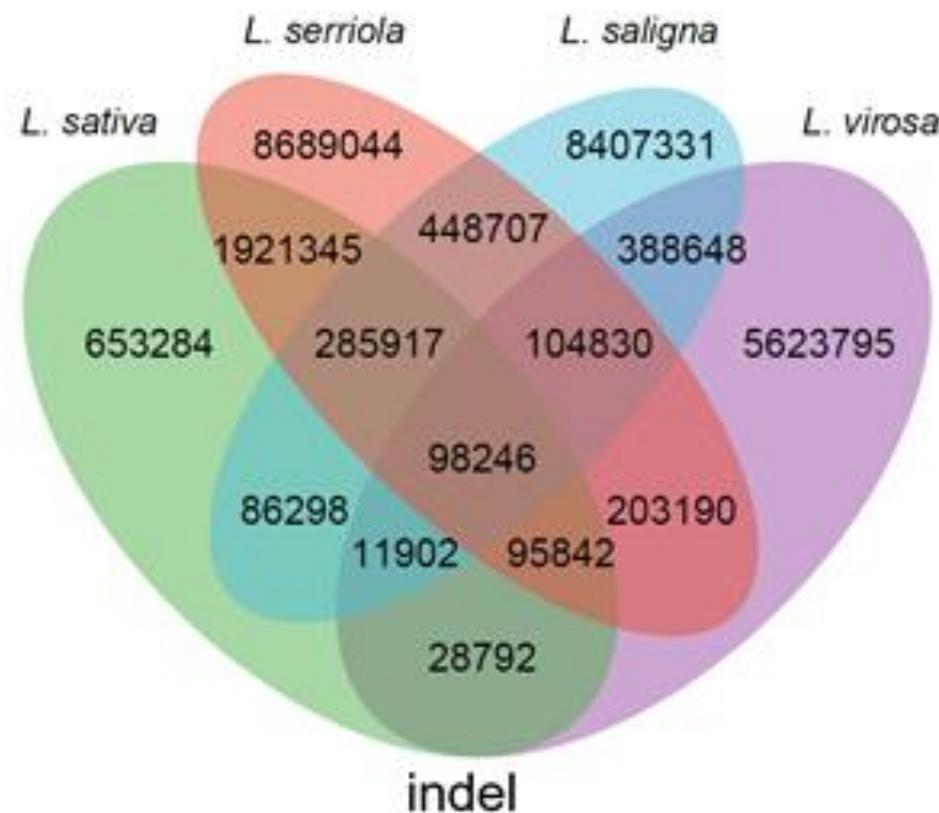
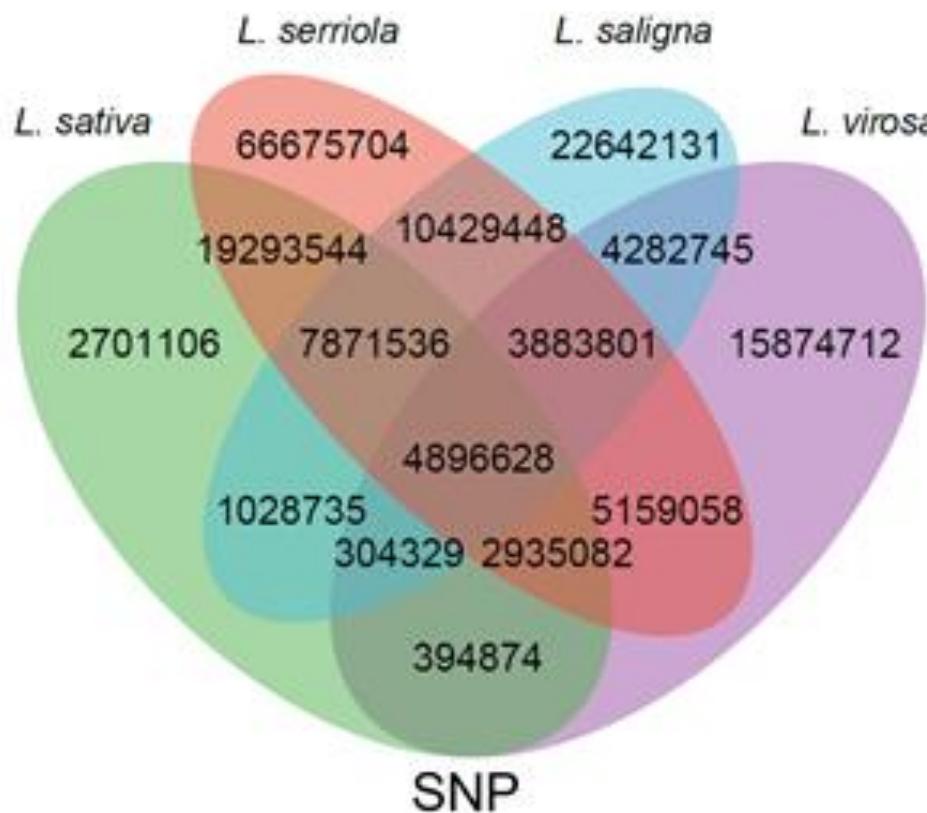
Coalescence-based tree from
4,513 nuclear genes

b

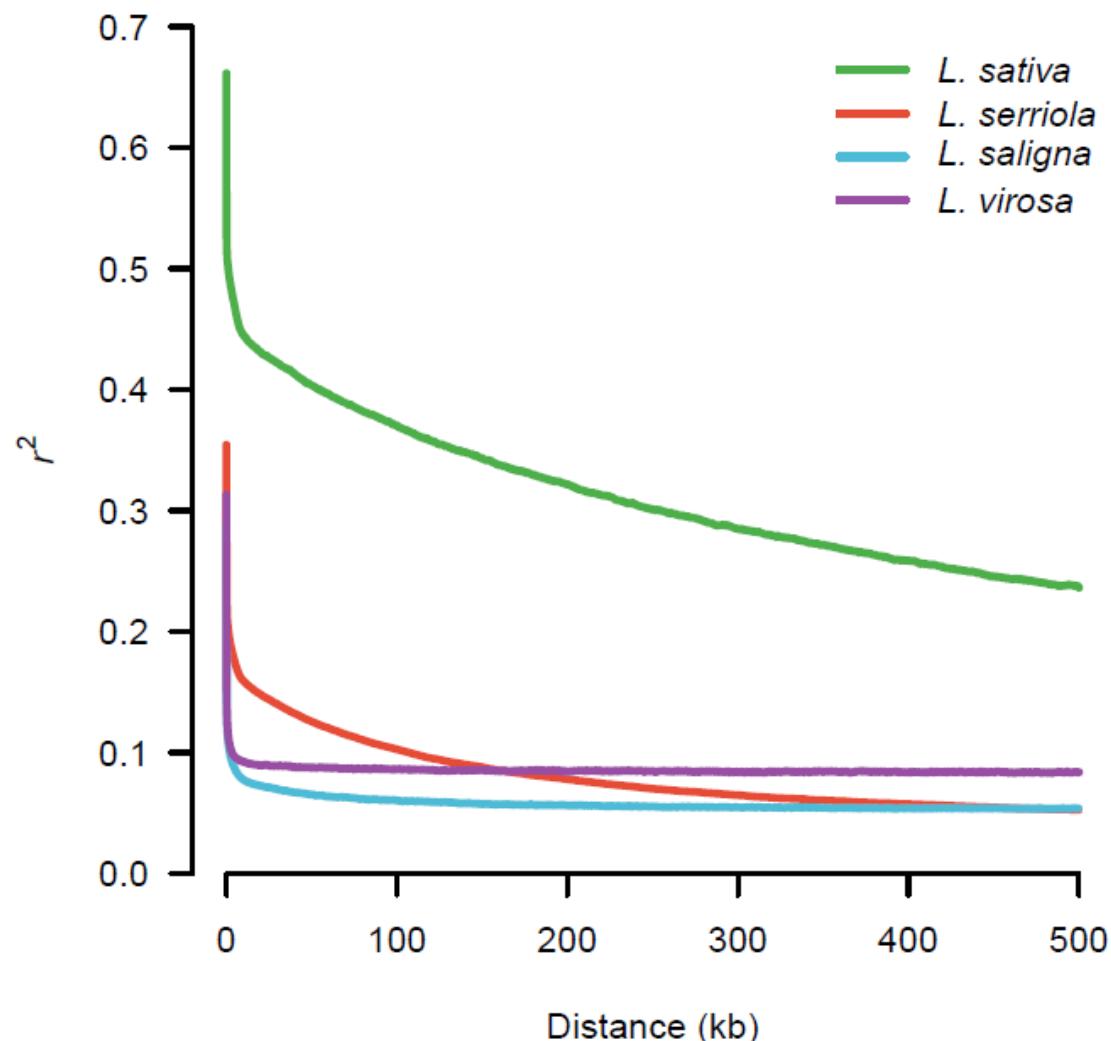
Concatenation-based tree from
75 plastid genes

Case I. SNP/indel stats in Venn diagram

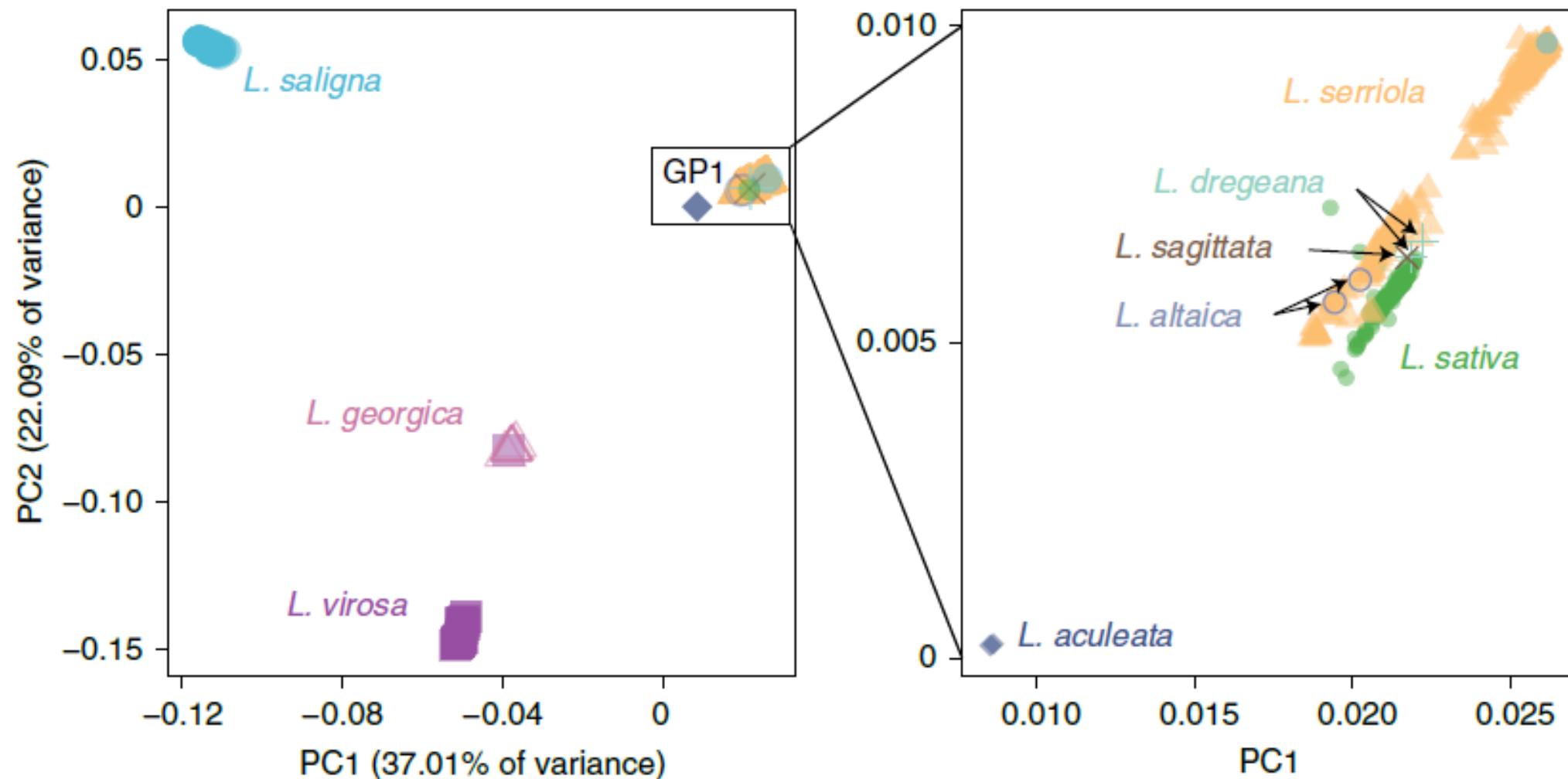
88.51% SNPs in *L. sativa* are shared with *L. serriola*.



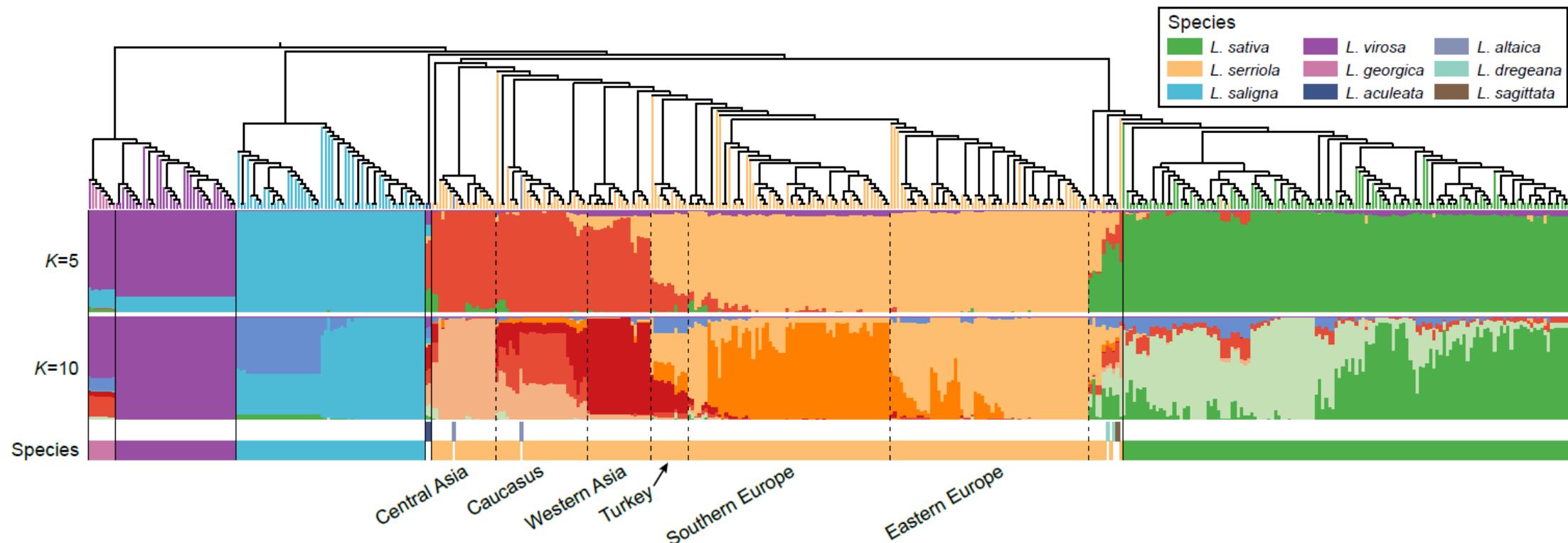
Case I. Linkage disequilibrium in line chart



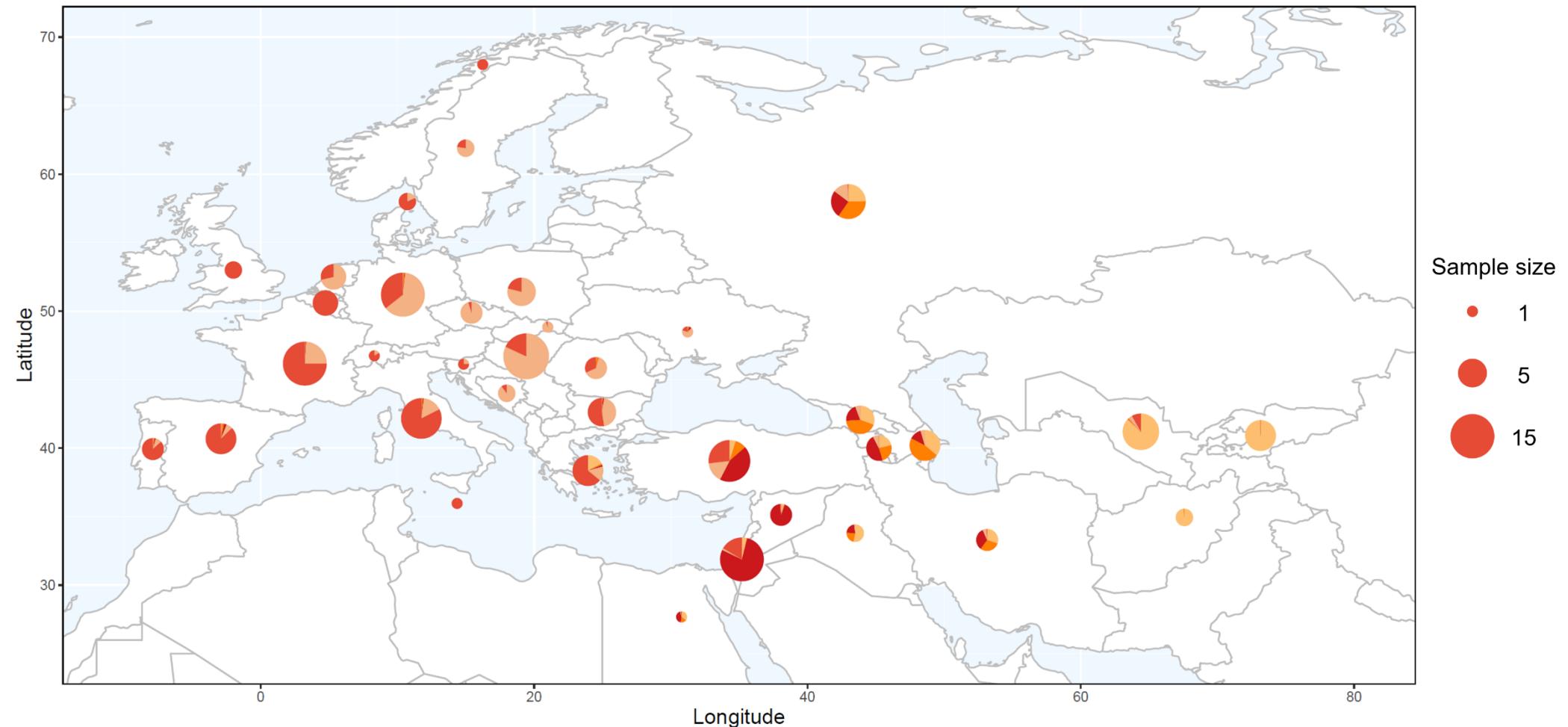
Case I. PCA in scatter plot



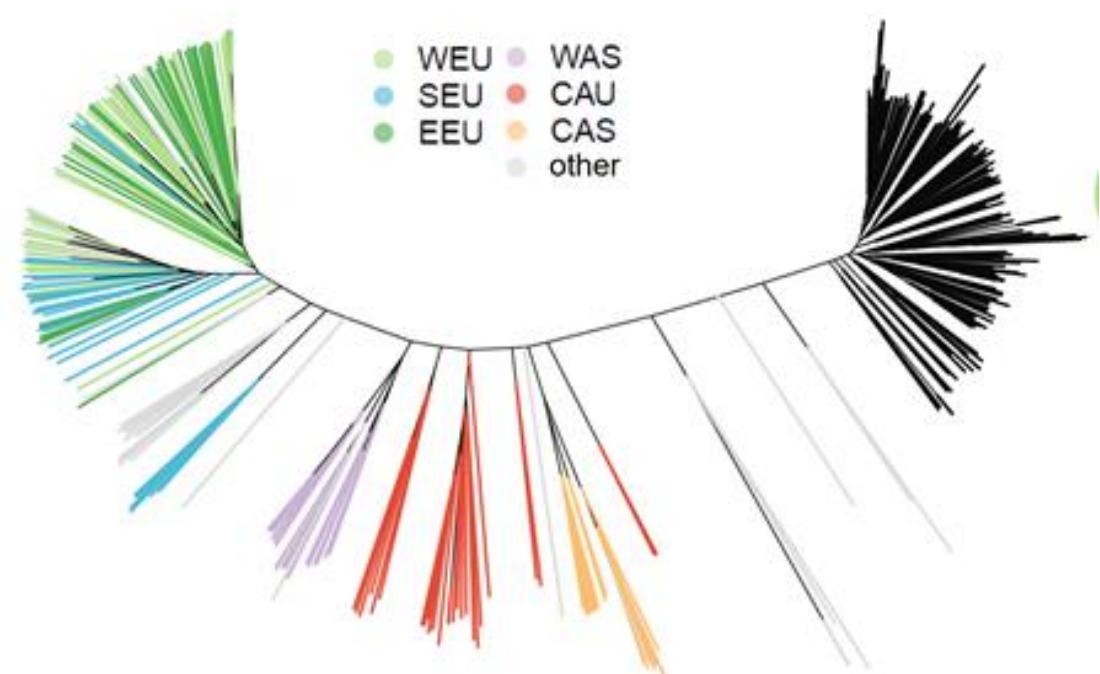
Case I. Structure in barplot



Case I. Genetic composition in piechart



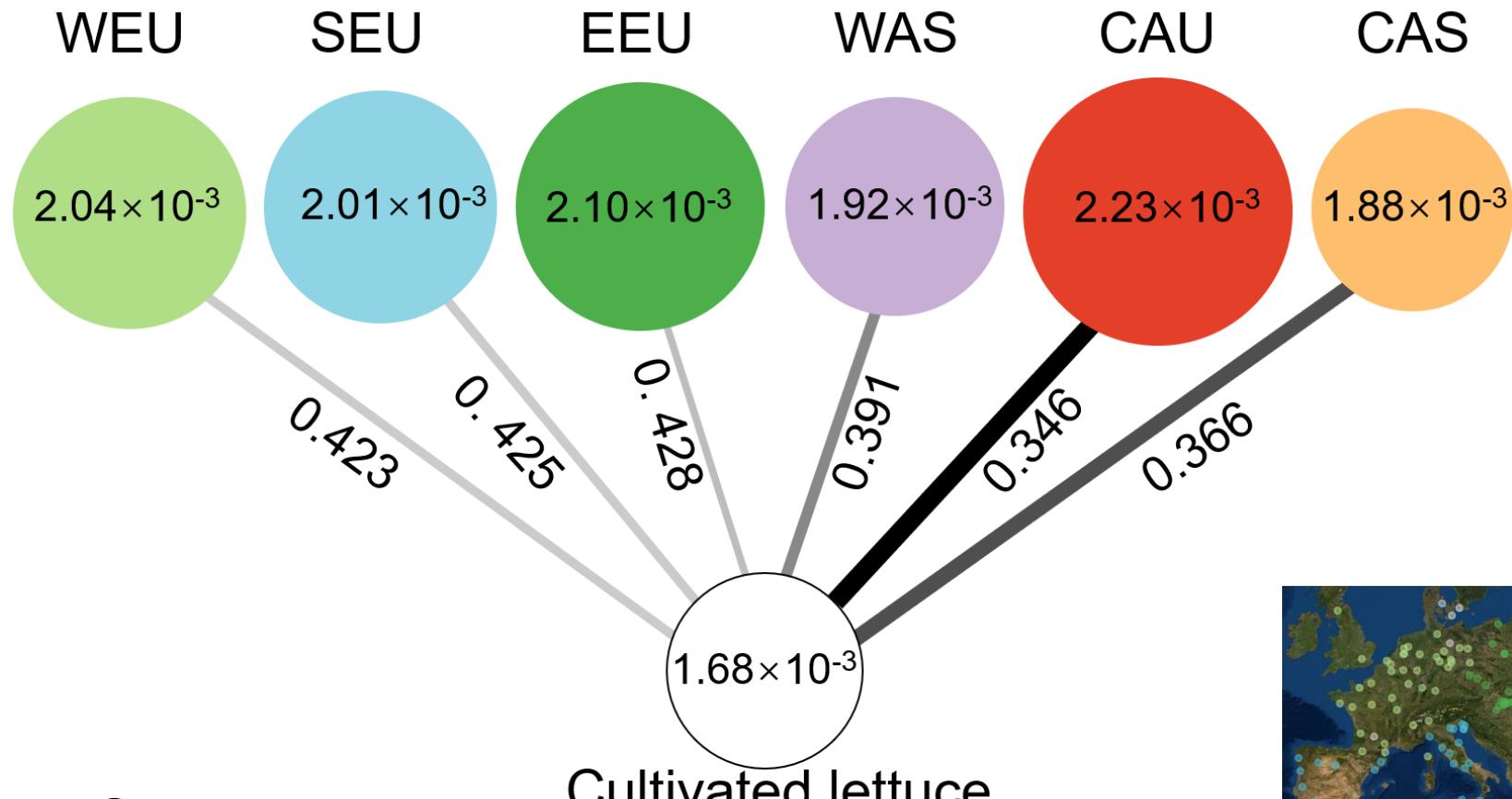
Case I. Genetic & geographic grouping



3 European groups
WEU: western Europe
SEU: southern Europe
EEU: eastern Europe

3 Asian groups
WAS: western Asian
CAU: Caucasian
CAS: central Asian

Case I. Population statistic in infographics

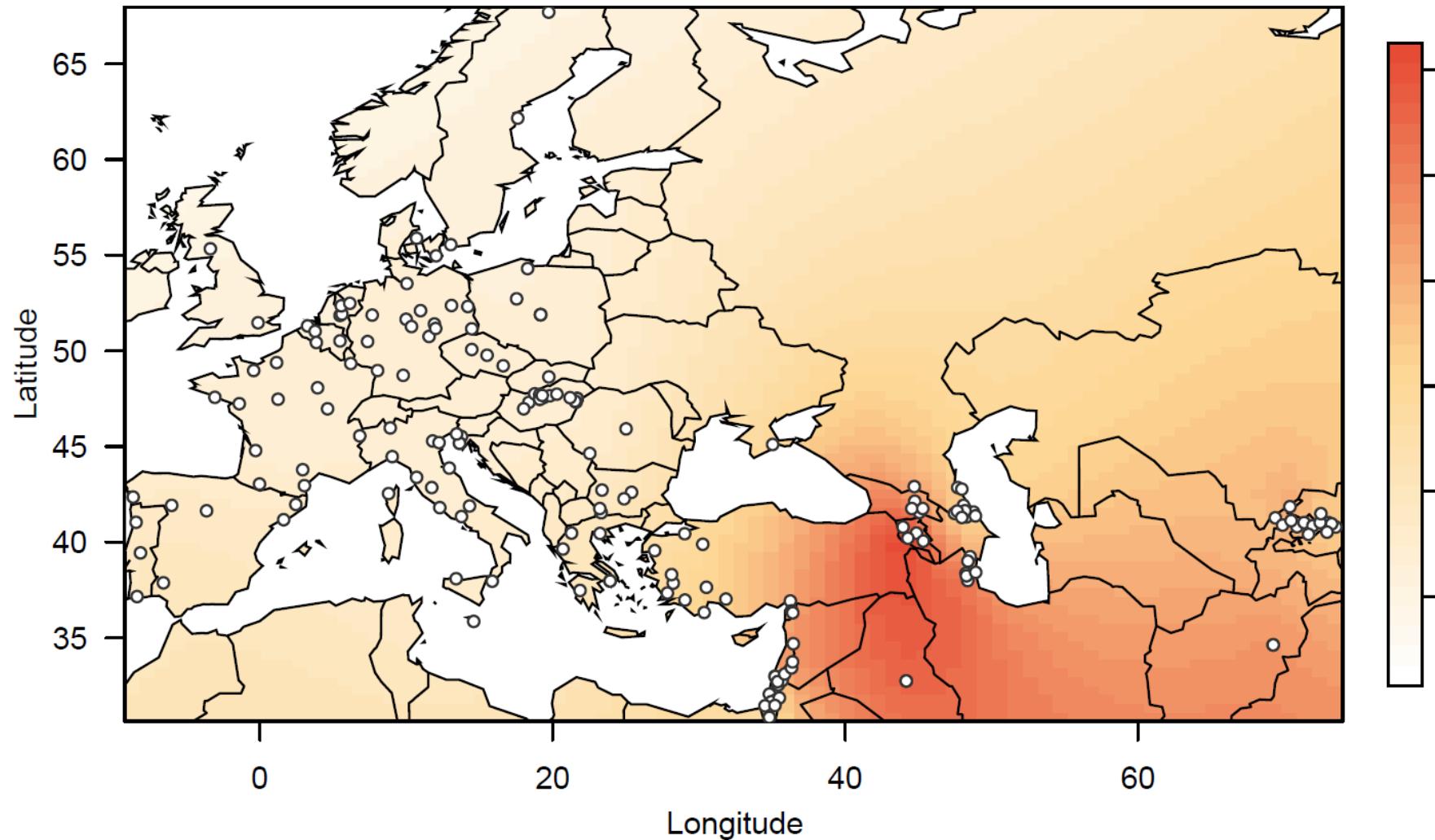


The highest π in CAU, and
the smallest F_{ST} between CAU & lettuce



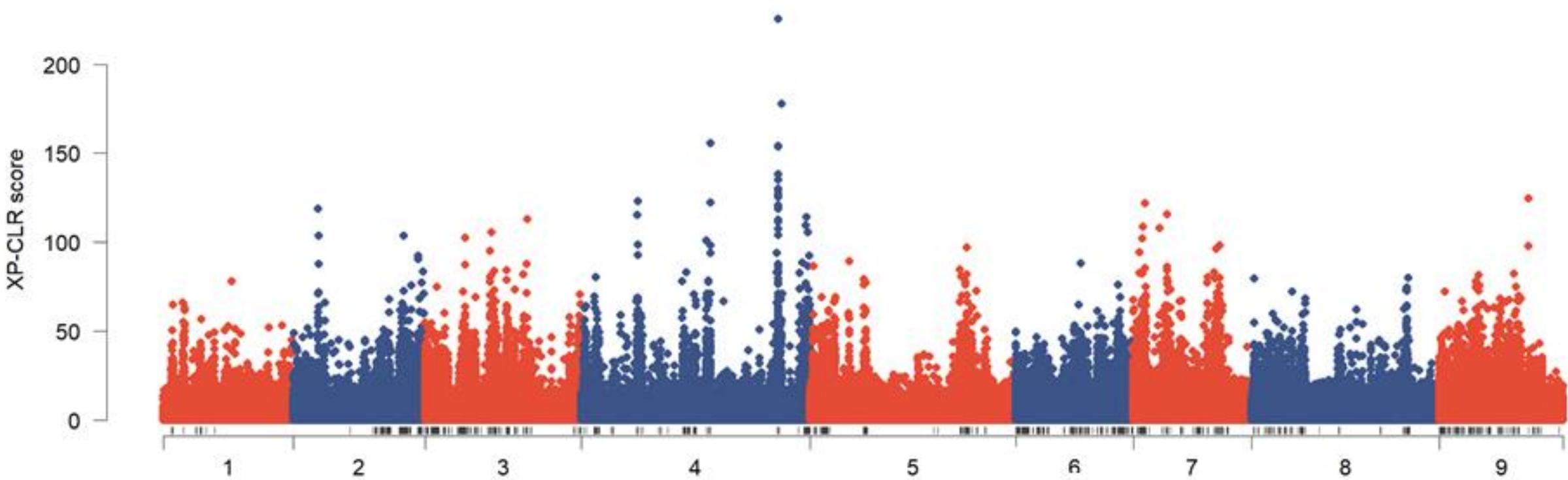
Case I. Singleton heatmap on map

The highest singletons in CAU



Case I. Selective sweeps in Manhattan plot

107.7 Mb and 2,304 genes within 4,089 selective sweeps



Case I. GWAS in Manhattan plot

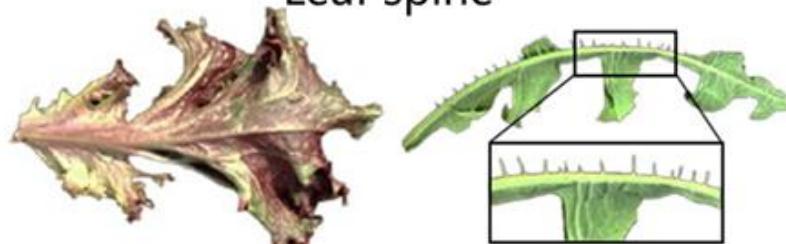
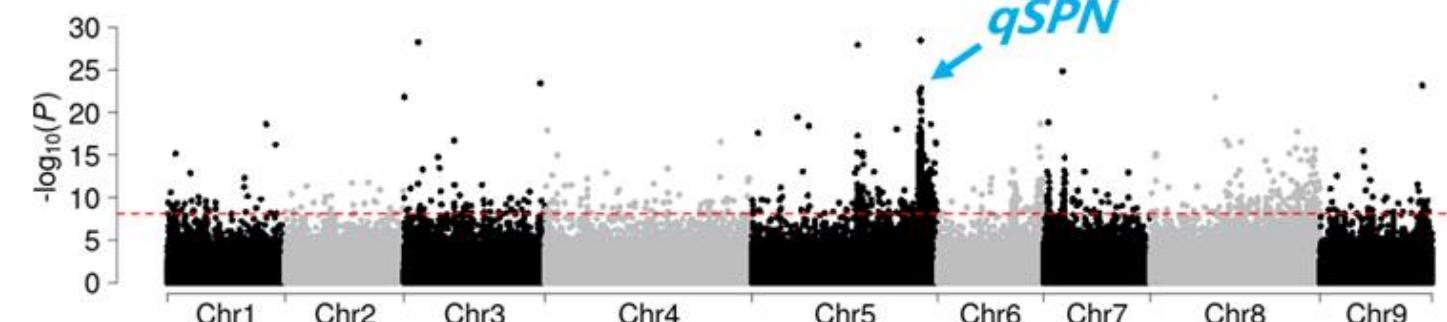
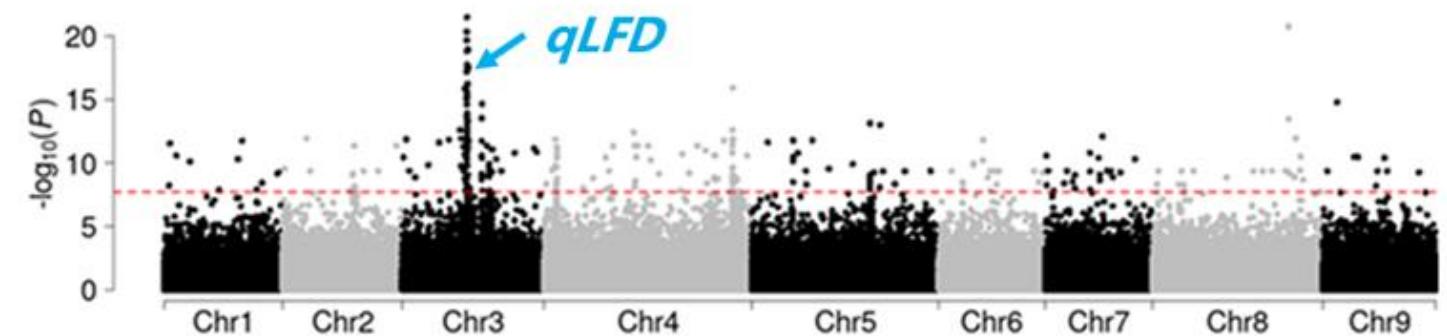
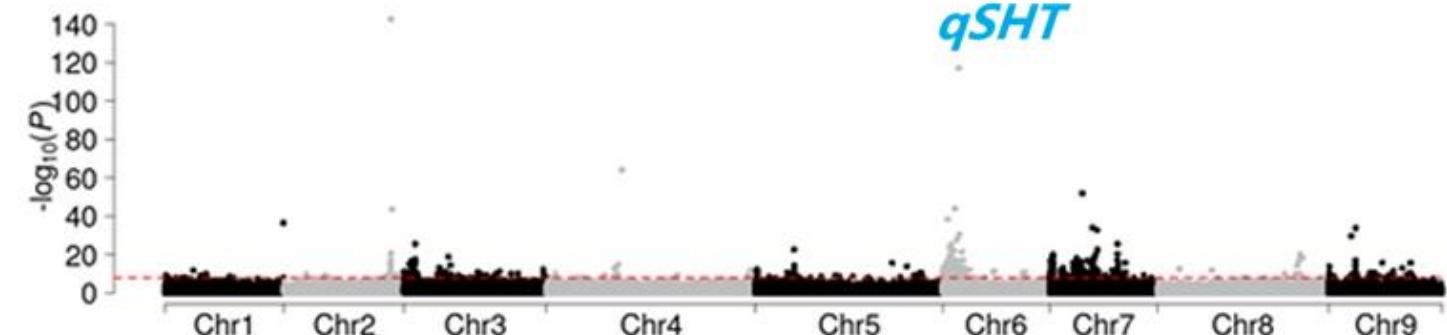
Seed shattering



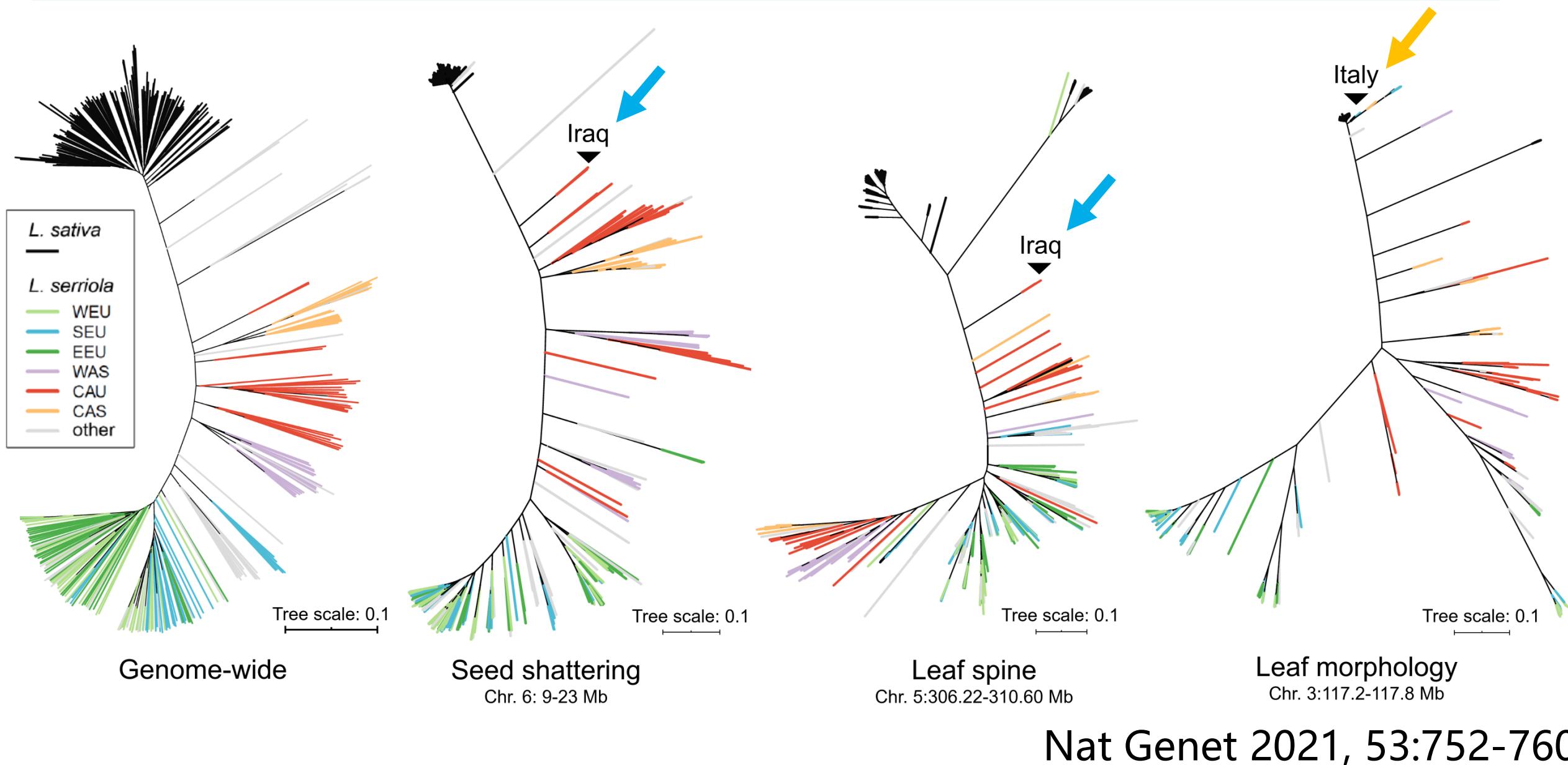
Leaf morphology



Leaf spine

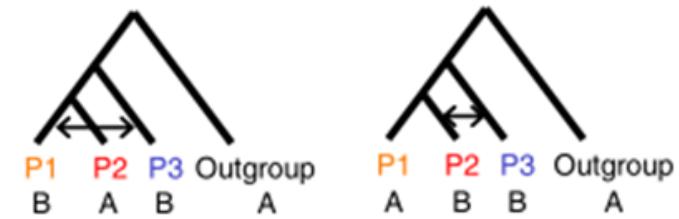
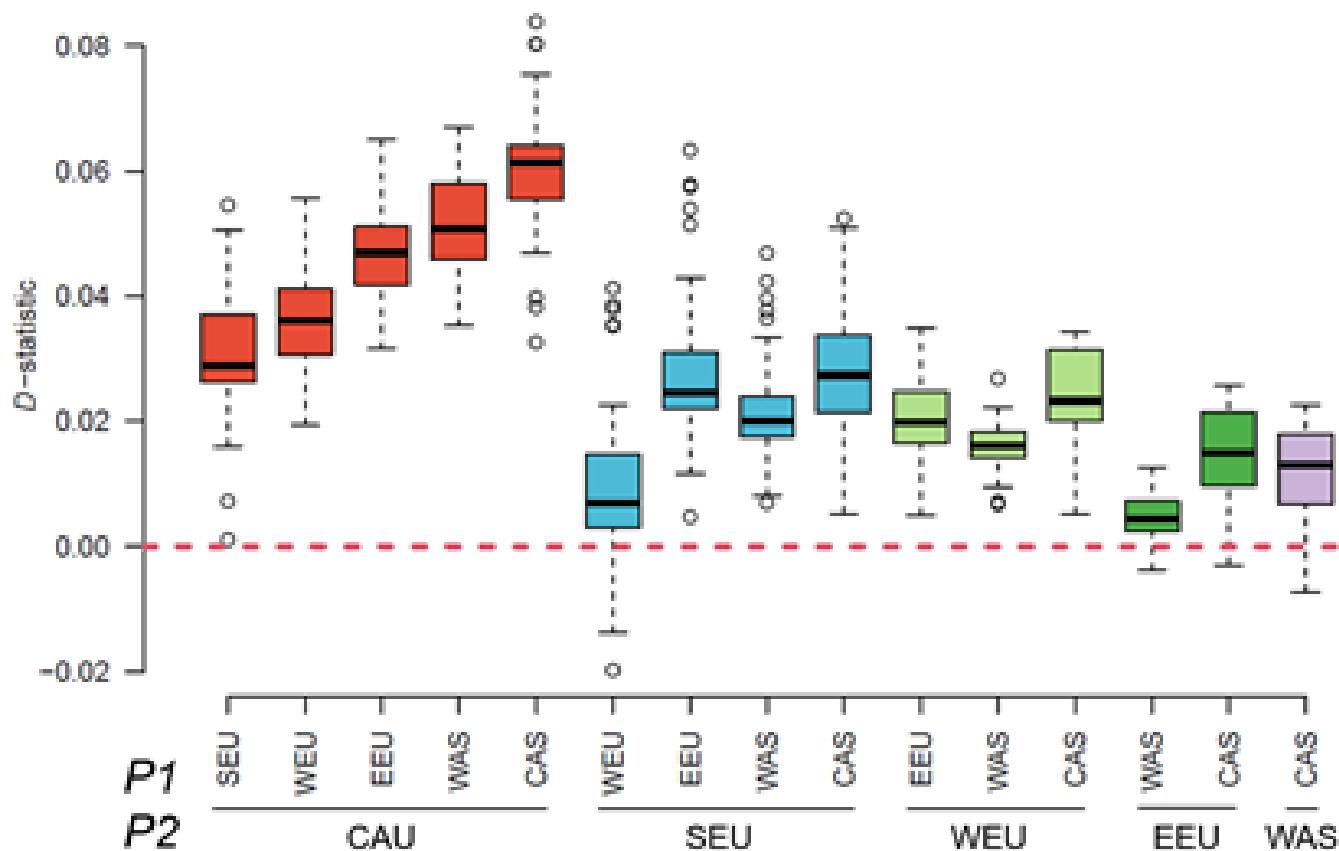
*L. sativa**L. serriola*

Case I. Trait origins deduced in tree plot



Case I. ABBA replicates in boxplot

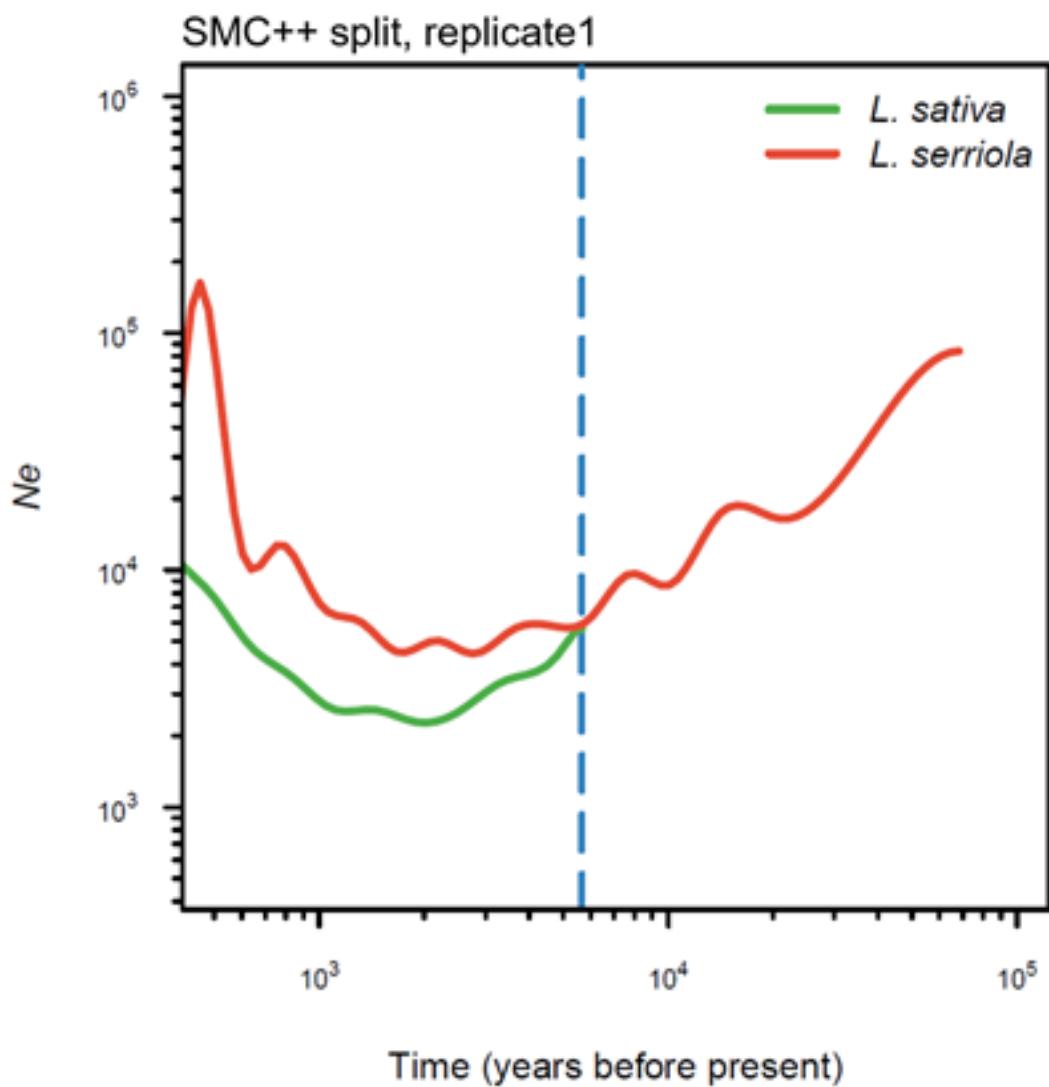
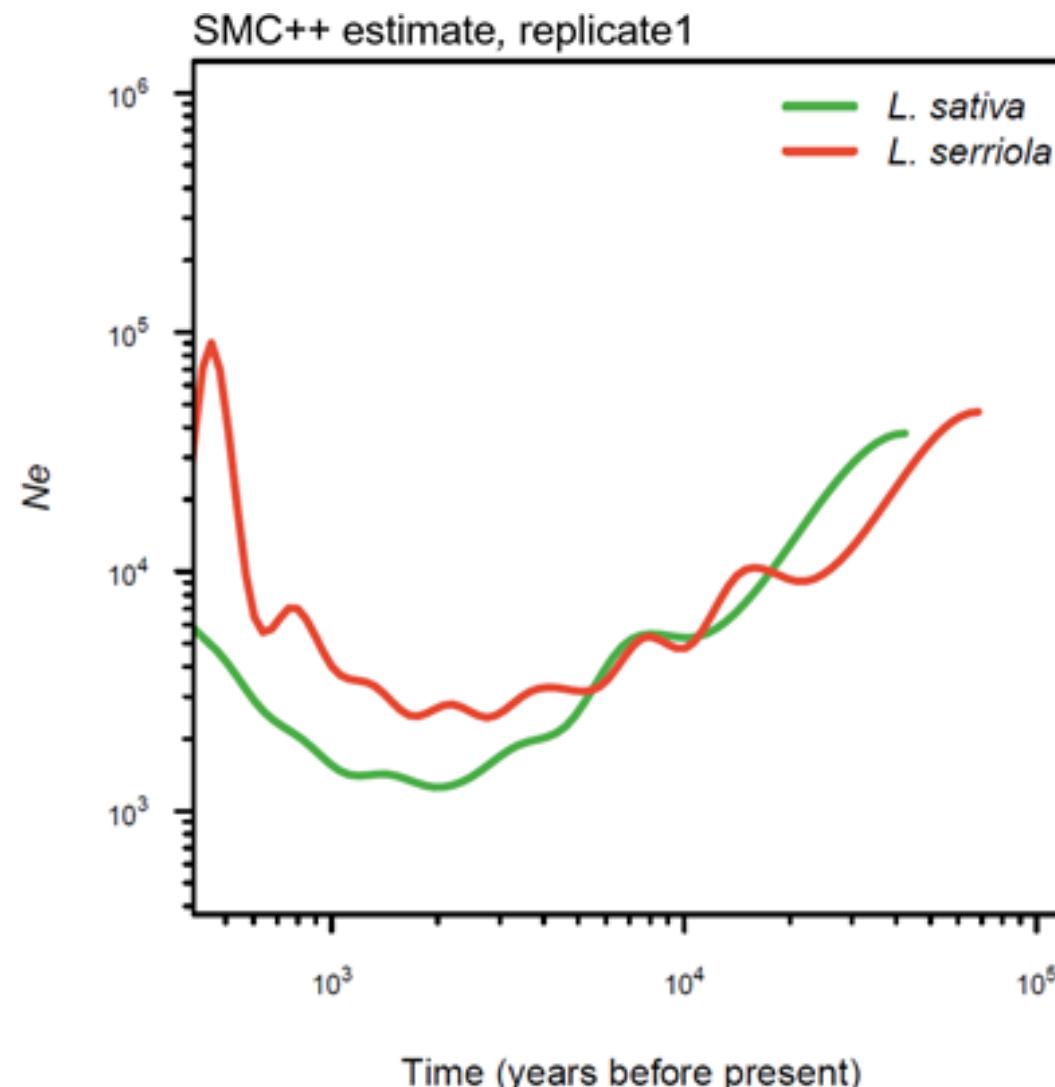
A close relationship in CAU and SEU



Outgroup: *L. saligna*
P3: *L. sativa*
P1 & P2: *L. serriola* groups



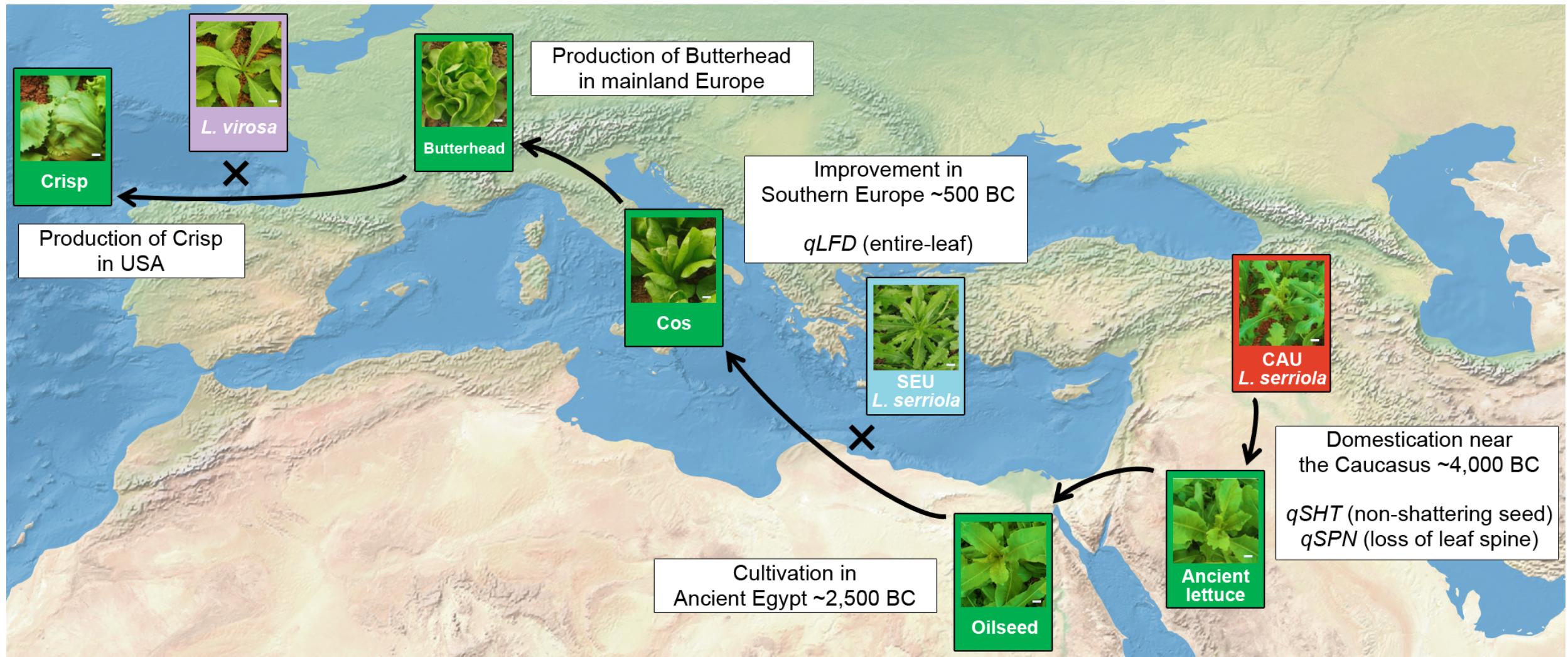
Case I. Effective population size in line chart



Case I. Answers to scientific questions

- Q: **Where and when** was lettuce was domesticated?
- A: Lettuce was domesticated **near Caucasus approximately 6,000 years ago.**
- Q: What are the **major events** and traits during lettuce domestication and improvement?
- A: **Non-shattering seeds** marked lettuce domestication; the **entire-leaf trait was introduced later** from a Southern European wild population.

Case I. Domestication history in infographics



Case I. Main figures about population

Fig. 1 | Phylogeny and population structure of cultivated lettuce and its wild relative species.

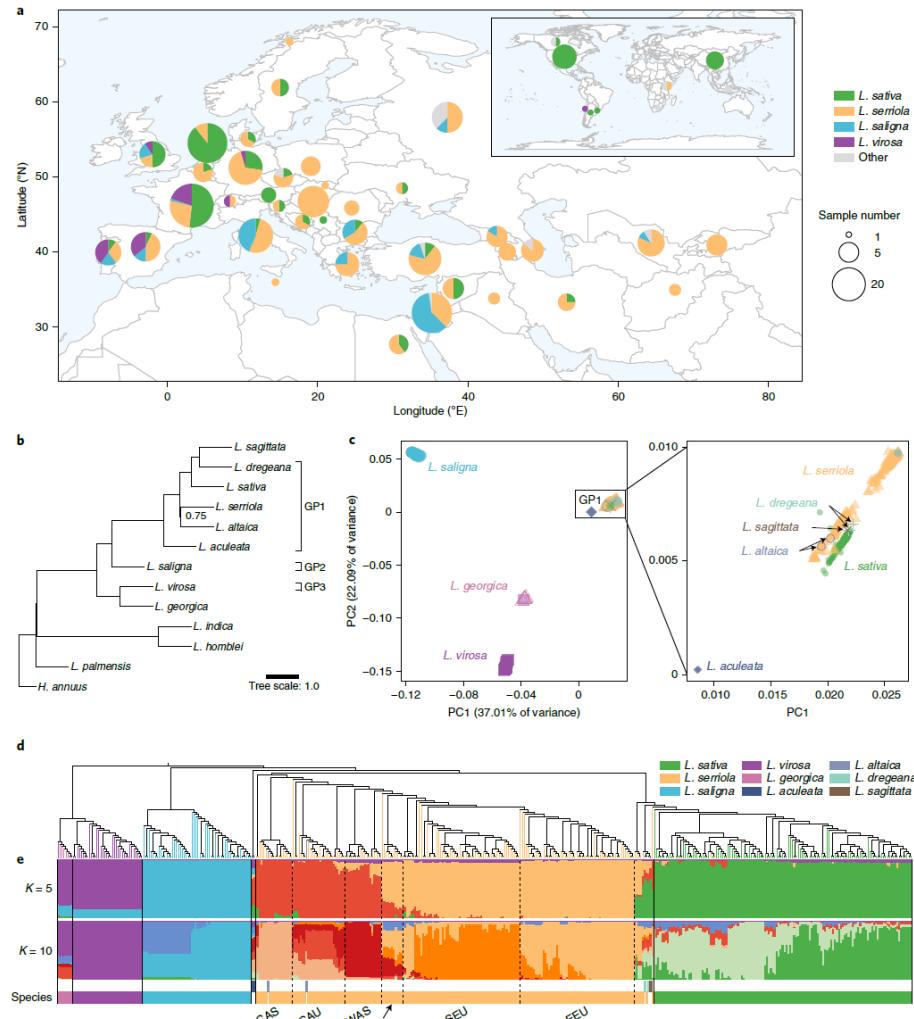
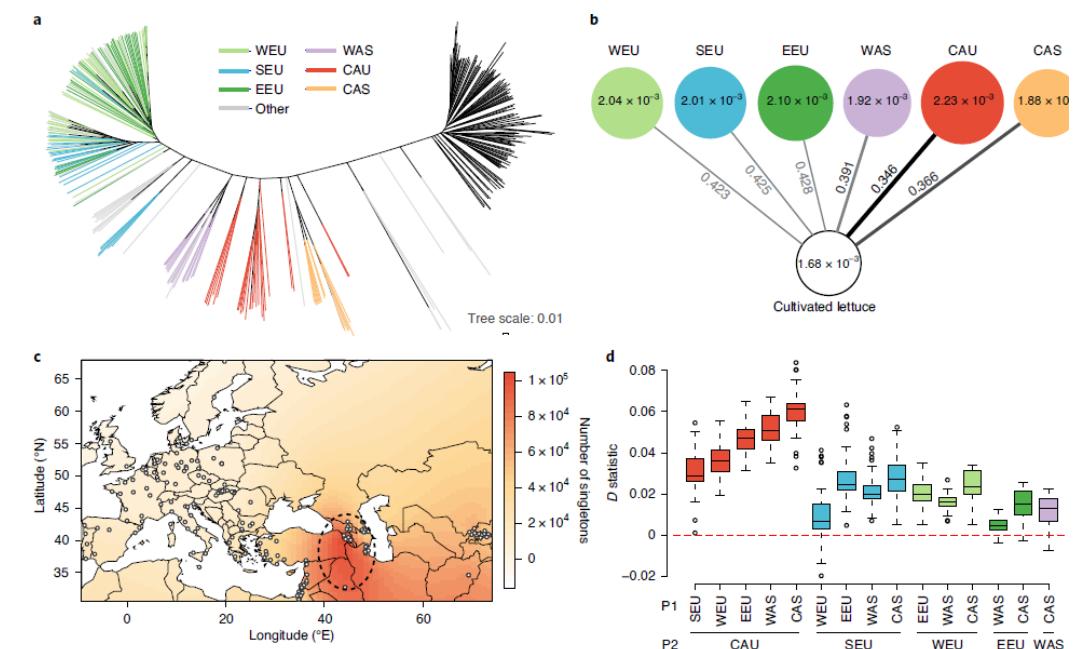


Fig. 2 | Proposed domestication center of cultivated lettuce near the Caucasus.



Case I. Main figures about function

Fig. 3 | Identification of selective sweeps associated with domestication traits in cultivated lettuce.

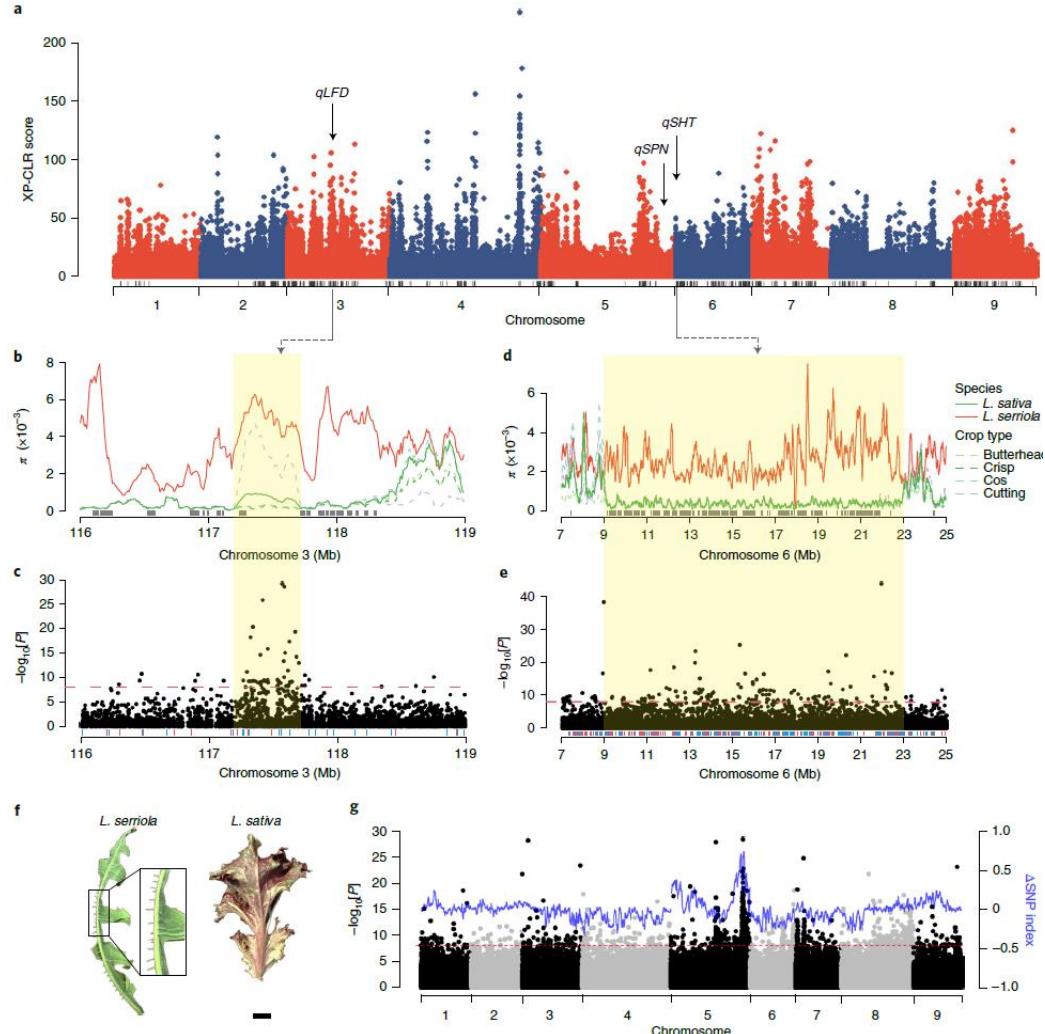
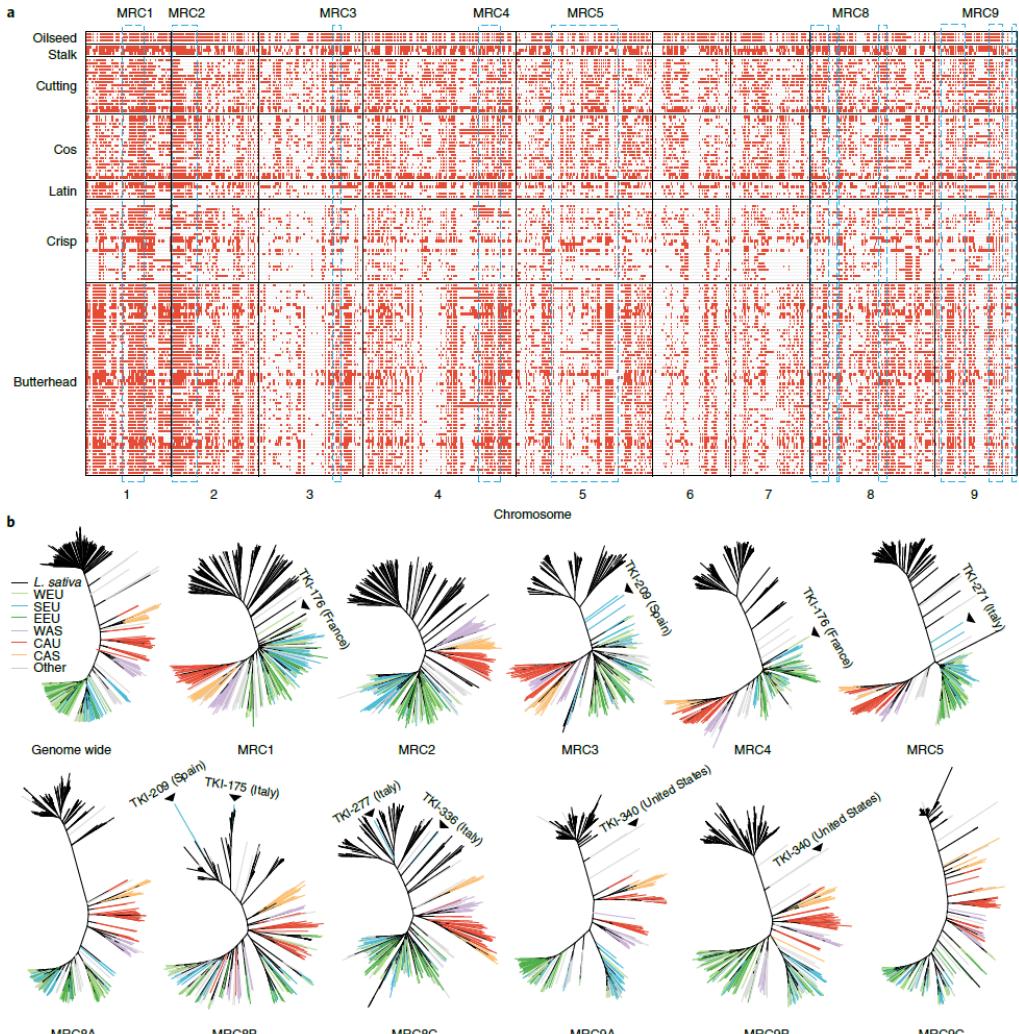
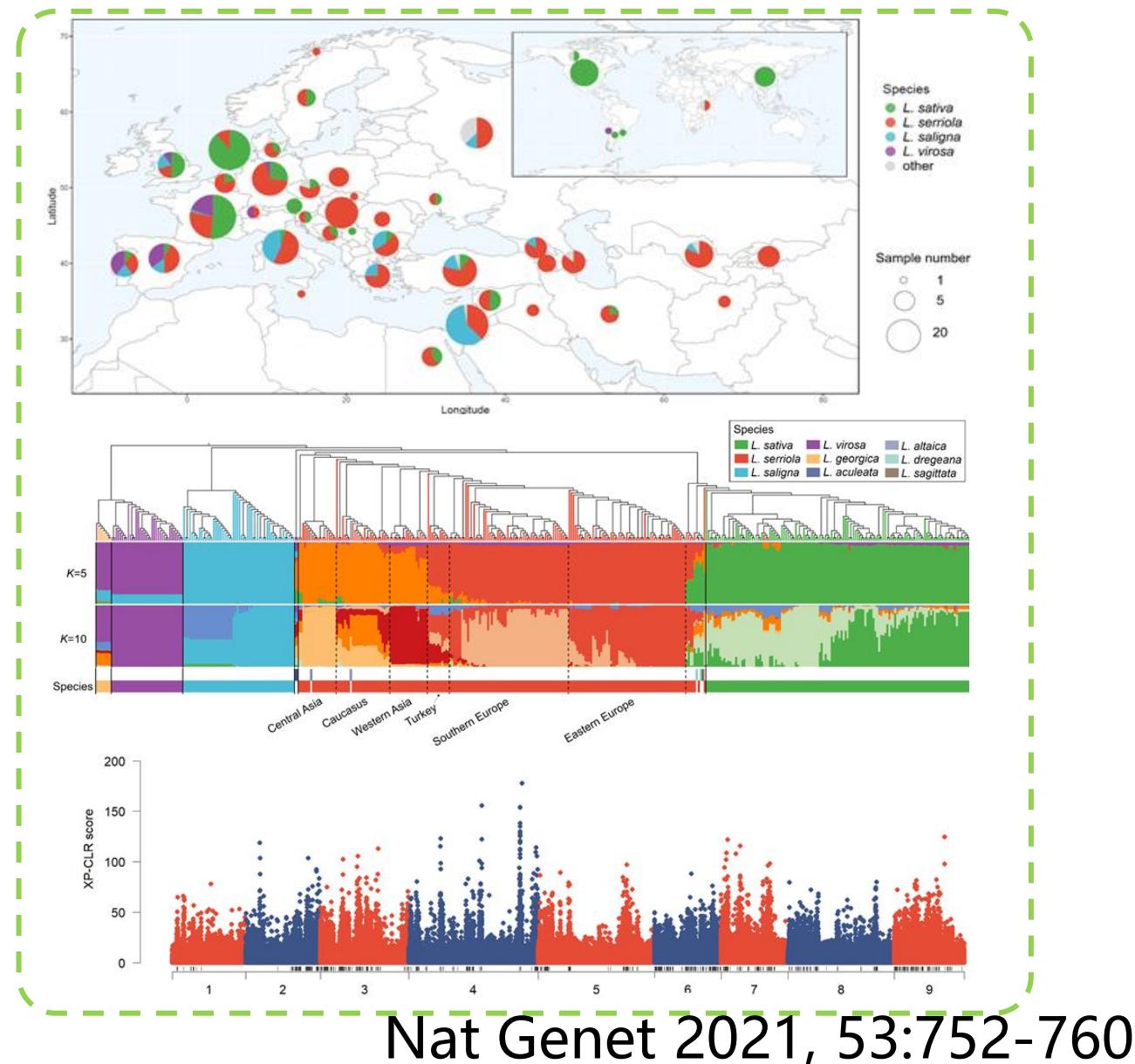


Fig. 4 | Introgressive contribution of *L. serriola* to lettuce resistance breeding.



Case I summary

- Population genomics of 445 cultivated and wild lettuce lines.
- Generated a comprehensive variation map.
- Curated GP status.
- Identified selective sweeps and major genetic determinants of domestication traits.
- Revealed lettuce domestication history.



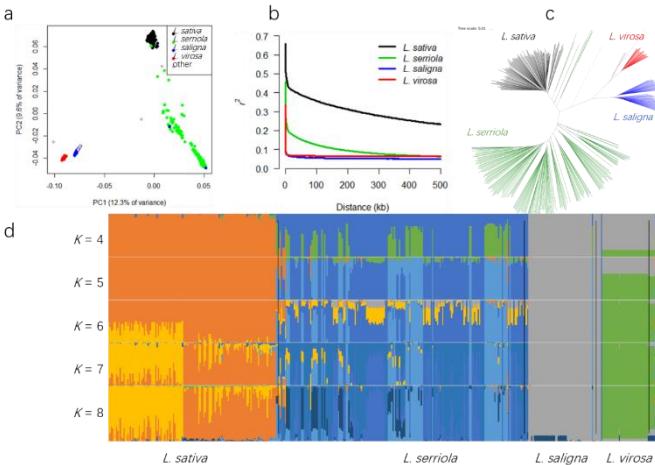
Part III

Figures for Results

Tips for writing result sections

- Each paragraph should have **ONLY one topic sentence.**
- Paragraphs should be **logically organized** around headings or sub-headings.
- Do not count words closely.

Result example: text for figure



- Our phylogenetic result showed that all the *L. sativa* accessions formed a monophyletic clade...
- Model-based clustering analysis revealed **additional** inter- and intra-species relationships.
- The phylogenetic relationships were **also** revealed by the principal component analysis (PCA)
- LD went to supplementary information.

Questions?