Quipu tutorial (Version 1.9.0)

Reinhard Simon, Jorge Nuñez

April 22, 2014

Contents

1 Uses

Genebanks increasingly use molecular markers for routine characterization of ex-situ collections and farmer managed diversity. CIP's (International Potato Center) genebank presently uses a SSR marker-kit to produce molecular profiles for potato accessions. We have been searching for a compact graphical representation that shows both molecular diversity and accession characteristics - thus permitting biologists and collection curators to have a simpler way to interpret high-volume data. Inspired by the ancient Andean quipus we devised a graph that allows for standardized representation while leaving room for updates of the marker kit and the collection of accessions. The graph has been used in several CIP publications.

Currently, the graph is designed for summarizing molecular diversity using SSR data. It shows the alleles present in an individual genotype and it allows to show the rare alleles using a combination of different sizes and colors for allele frequency classes. The allele classes have been fixed to four but cutoffs between classes may be defined freely. Each SSR locus is represented by a vertical line (which also corresponds to a gel lane) and the y range is set by default to the typical minimum and maximum ranges of SSR alleles. The actual range size of alleles in bp is than marked by a thicker grey line - where the actual range is derived either from the dataset directly or an external allele reference table. Individual alleles are then plotted on top - larger sized alleles are plotted first to allow better visibility of close alleles. Allele frequencies are calculated based on the given dataset using simple counting or can be given using a pre-defined table. This is specially useful when sets of genotypes are published in different batches but all must refer to the same main database.

The main routine has been designed so that it can be used in different settings: for interactive exploratory visualization or for creation of images in batch.

2 Quick examples

2.1 Demo

The package comes with an html-based demo. Copy the following code into your R console:

```
if (interactive()) {
    quipu::runDemo()
}
```

This demo shows the two main chart layouts and the effects of the principal parameters.

2.2 Minimal interactive example

The package *quipu* comes with a small sample dataset.

```
library(quipu)

## Loading required package:  agricolae
## Loading required package:  stringr
## Loading required package:  pixmap
## Loading required package:  shiny
## Loading required package:  xtable

library(xtable)
data("potato.quipu", package = "quipu")
dat = potato.quipu
```

The allele table has the following format (showing only the first few records):

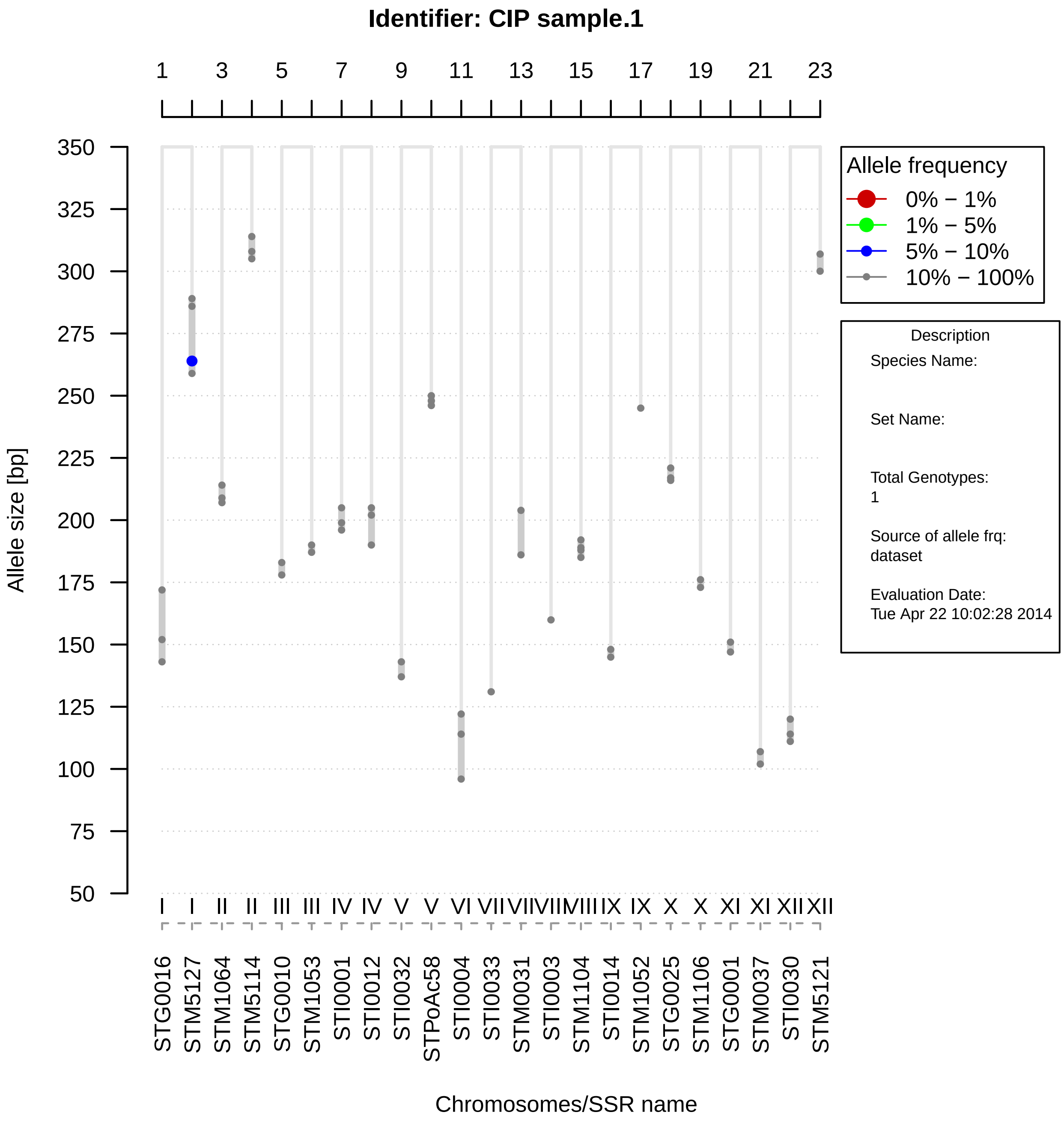
	accession_id	primer_name	marker_size	map_location
1	sample.1	STM1064	207	II
2	sample.1	STM1064	209	II
3	sample.1	STM1064	214	II
4	sample.1	STPoAc58	246	V
5	sample.1	STPoAc58	248	V
6	sample.1	STPoAc58	250	V

Columns in the table must be named as shown above. The following snippet shows the most simple usage which will switch in interactive mode from image to image for each genotype (copy into your console).

```
if (interactive()) rquipu(dat)
```

To produce a single graph the following settings should be used:

```
rquipu(dat, a.subset = "sample.1")
```



The 'a.subset' parameter recognizes also a vector of accession id's; e.g. `c("sample.1", "sample.3")`.

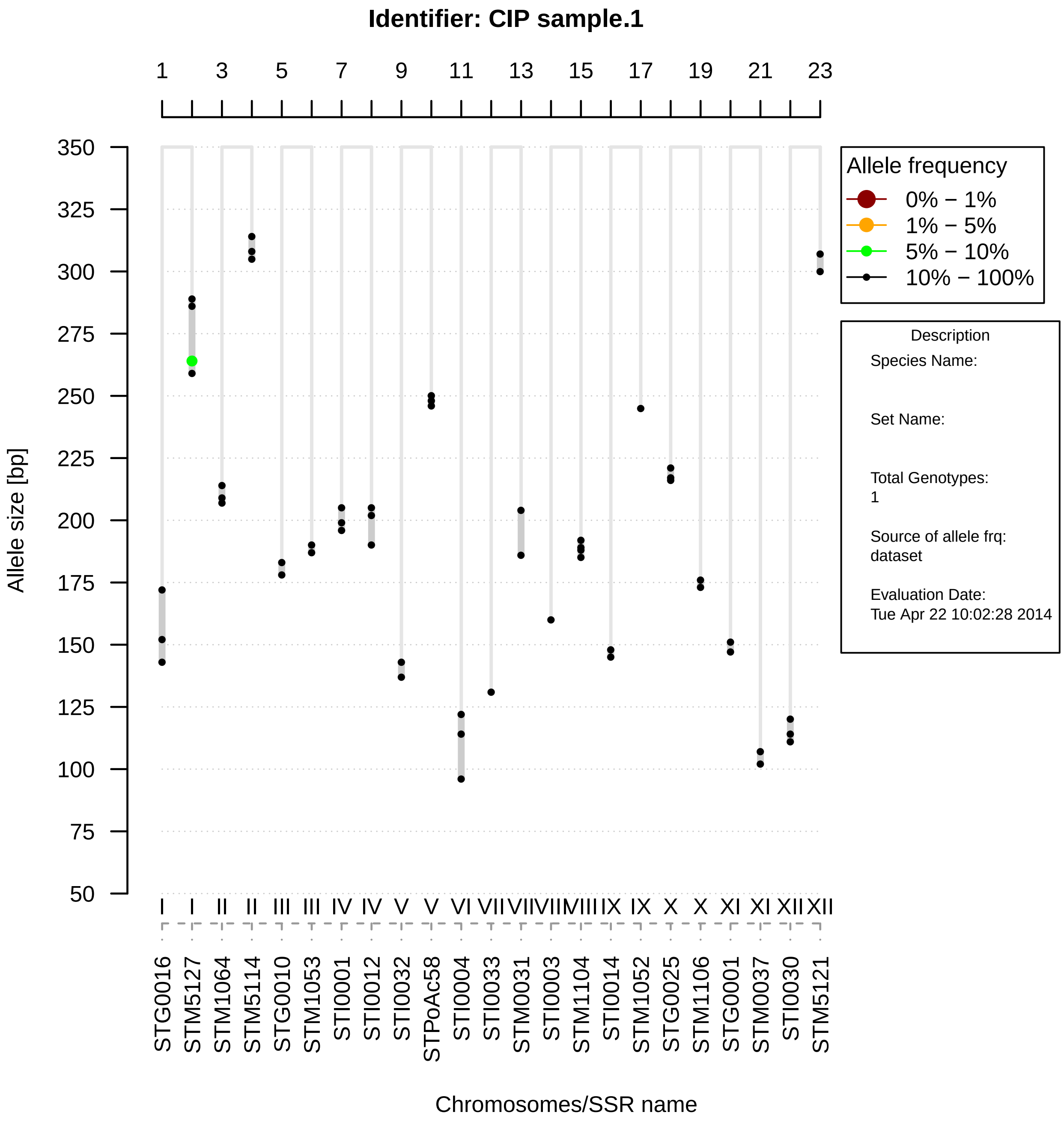
3 Modifications

The `rquipu` function has several parameters so the graph can be highly customized. The options can be grouped roughly into four groups by purpose: a) visualization, b) classification, c) database reference, and d) fine tuning legend options.

3.1 Changing visualization options: allele sizes and colors

The colors of the four node type representing the allele classes can be changed by setting the parameter `col.node`. It has the following default values: `c("red3","green","blue","gray50")`. It can be changed as in the following example:

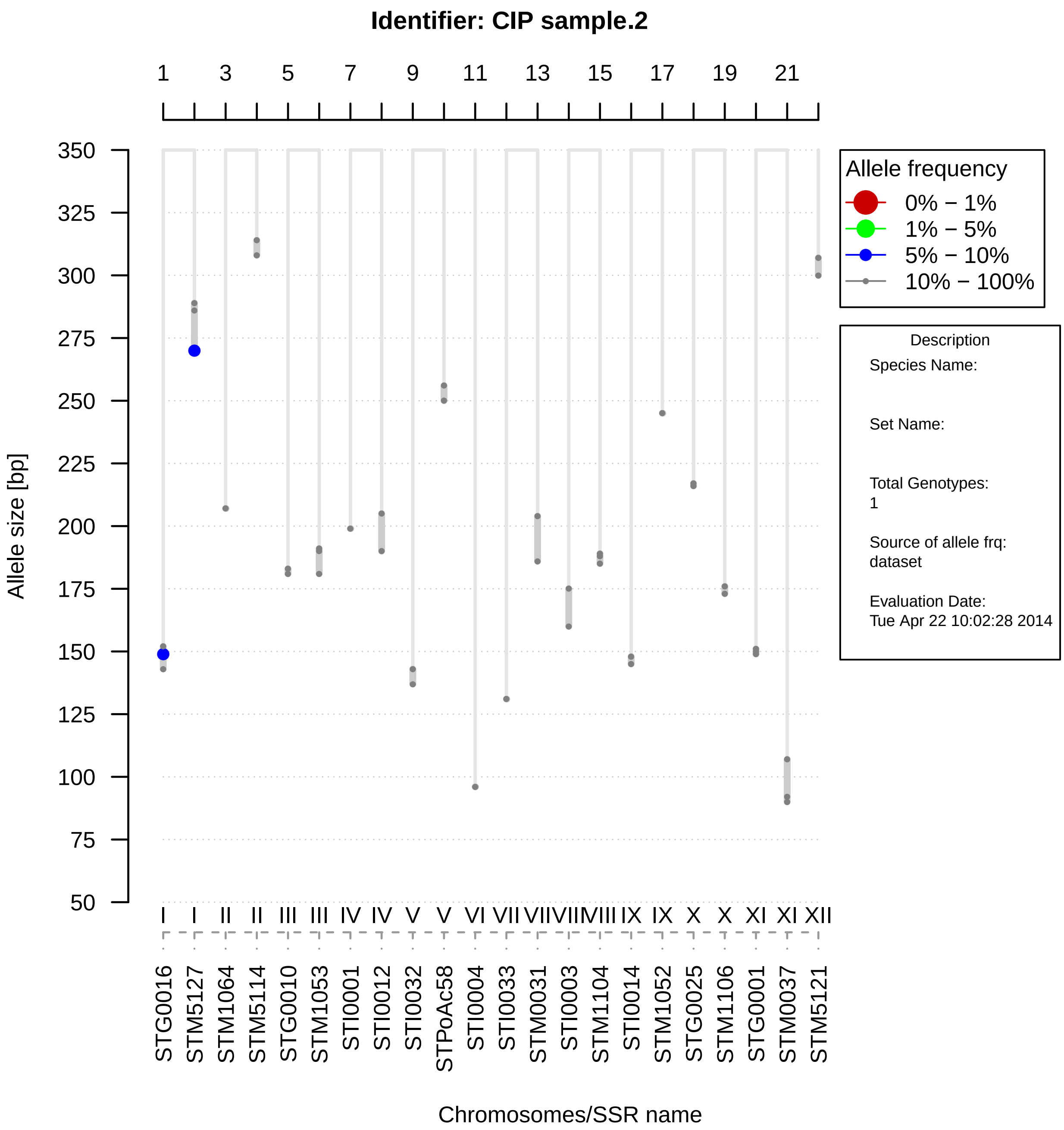
```r
rquipu(dat, a.subset = "sample.1", col.node = c("darkred", "orange",
    "green", "black"))
```



A list of possible color names can be obtained in R using the command `colors()`. The `rquipu` function checks for allowed values.

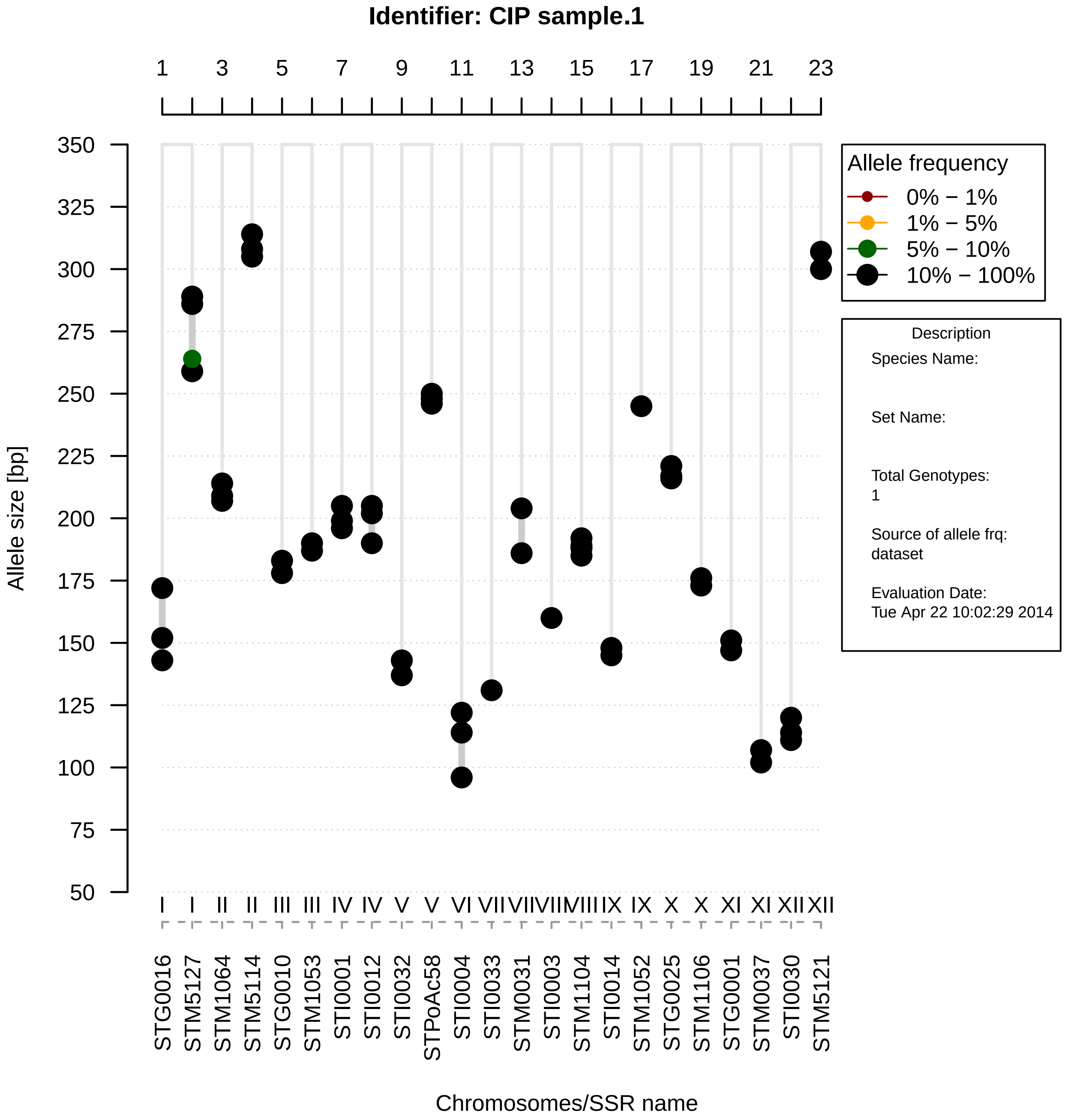
Likewise, the node size has a parameter `node.size` with the following standard values: `c(1.5, 1.2, 0.9, 0.6)`. Values may be changed as follows:

```
rquipu(dat, a.subset = "sample.2", node.size = c(2, 1.5, 1, 0.5))
```



The size order is by default inverse to the rareness of an allele, but this can be changed. One may also combine both color assignmnet and node size together:

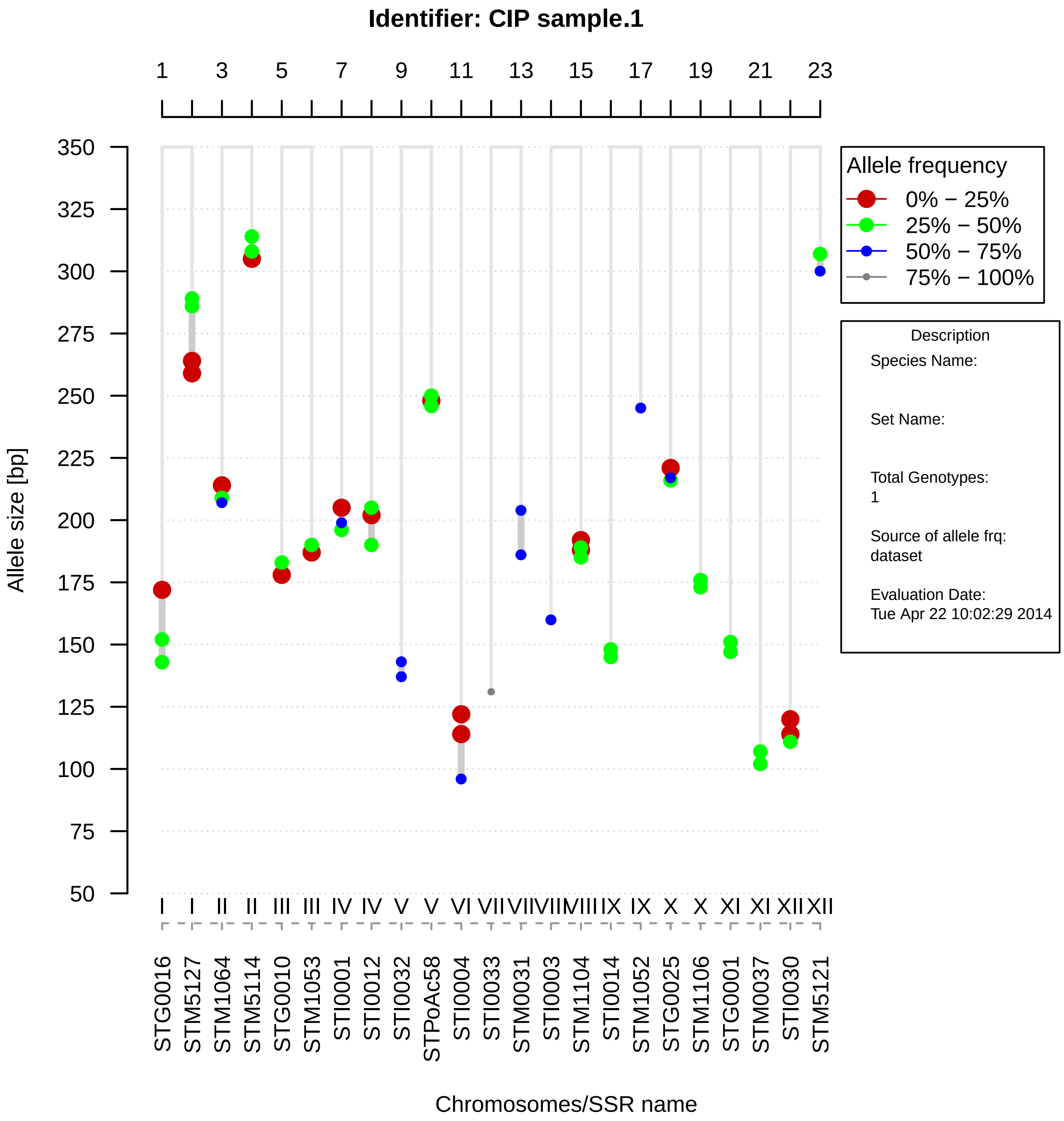
```
rquipu(dat, a.subset = "sample.1", col.node = c("darkred", "orange",
    "darkgreen", "black"), node.size = c(0.9, 1.2, 1.5, 1.8))
```



3.2 Changing classification options: allele classification cutoffs

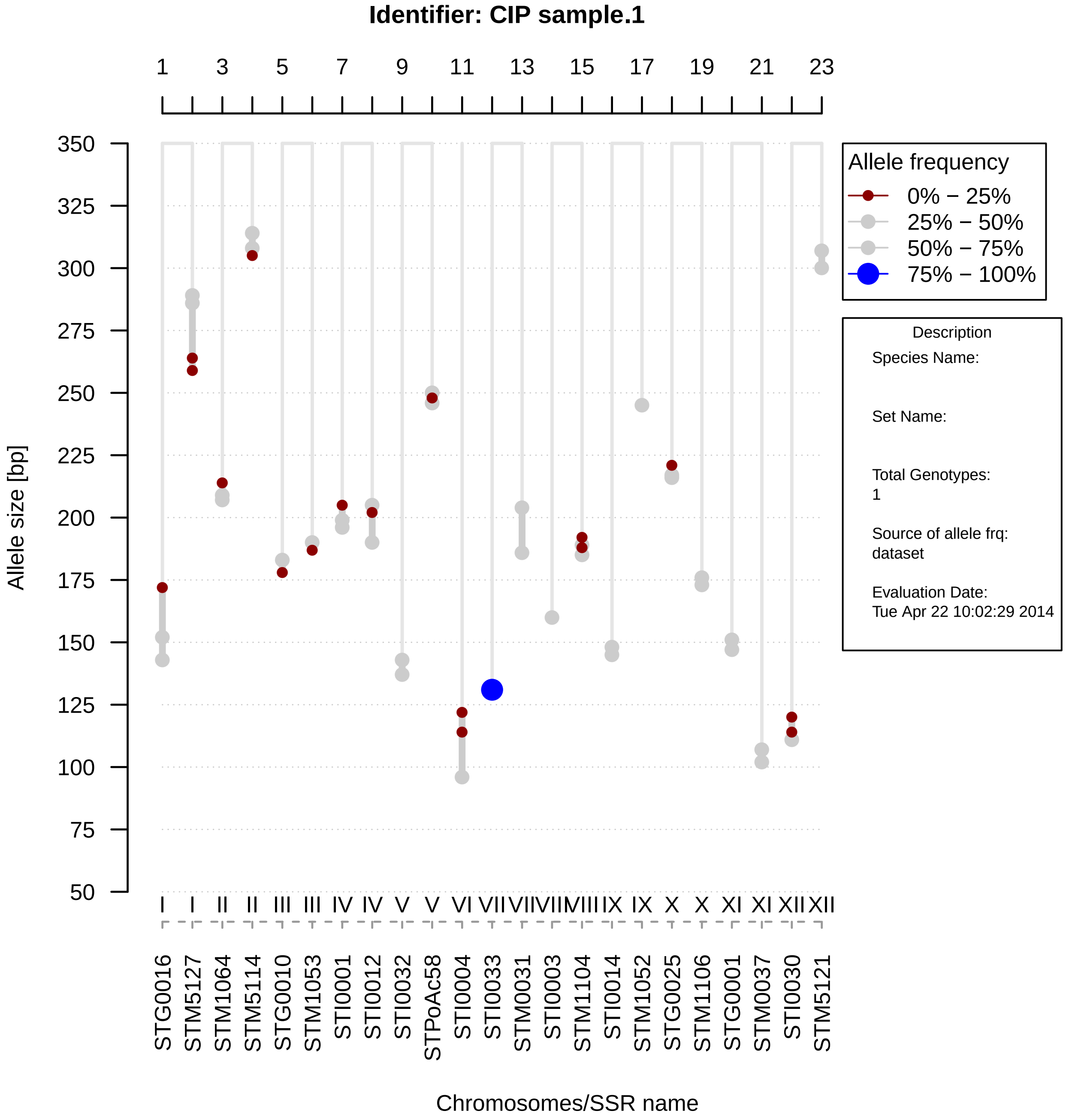
The default cut-off values between allele frequency classes were chosen to reflect the typical cut-off values used for p-values (`c(0.01, 0.05, 0.1)`). However, this can be changed using the parameter `grp.brks`

```
rquipu(dat, a.subset = "sample.1", grp.brks = c(0.25, 0.5, 0.75))
```



Again, this parameter can be freely combined with others as in the following example to visualize the extreme 25 percent (this also shows that the node size parameters can be of equal size):

```
rquipu(dat, a.subset = "sample.1", col.node = c("darkred", "grey80",
    "grey80", "blue"), node.size = c(0.9, 1.2, 1.2, 1.8), grp.brks = c(0.25,
    0.5, 0.75))
```



3.3 Changing database options: allele frequencies

In some circumstances the simple counting of alleles from the supplied data.framne may not be adequate. The parameter obs.alls.frq should then be used to supply a reference table. Also, for documentation purposes, another parameter obs.alls.frq.ref should be used for registering a short name to the source, e.g. '[database name] [version number]'.

```
data("allele.freqs", package = "quipu")
```

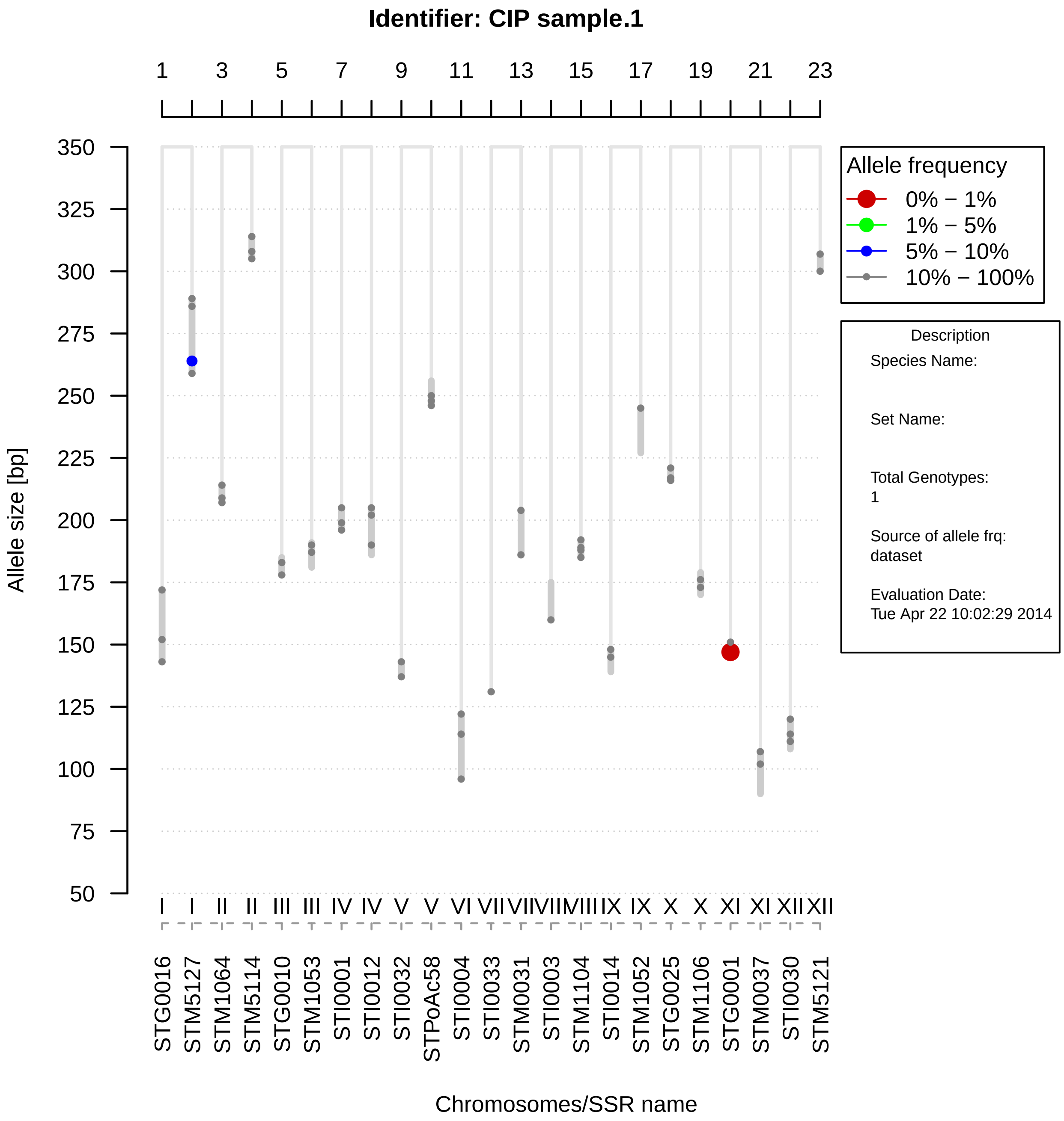
The frequency reference table has the following structure - again, column names must be given as shown. For illustration purposes we change the frequency of the third allele to a very rare one.

	marker	marker_size	frequency
1	STG0001	147	0.0010
2	STG0001	149	0.1818
3	STG0001	150	0.1818
4	STG0001	151	0.3636
5	STG0010	178	0.2222
6	STG0010	181	0.2222

```
allele.freqs[3, 3] <- 1e-04
```

	marker	marker_size	frequency
1	STG0001	147	0.0010
2	STG0001	149	0.1818
3	STG0001	150	0.0001
4	STG0001	151	0.3636
5	STG0010	178	0.2222
6	STG0010	181	0.2222

```
rquipu(dat, a.subset = "sample.1", obs.alls.frq = allele.freqs)
```

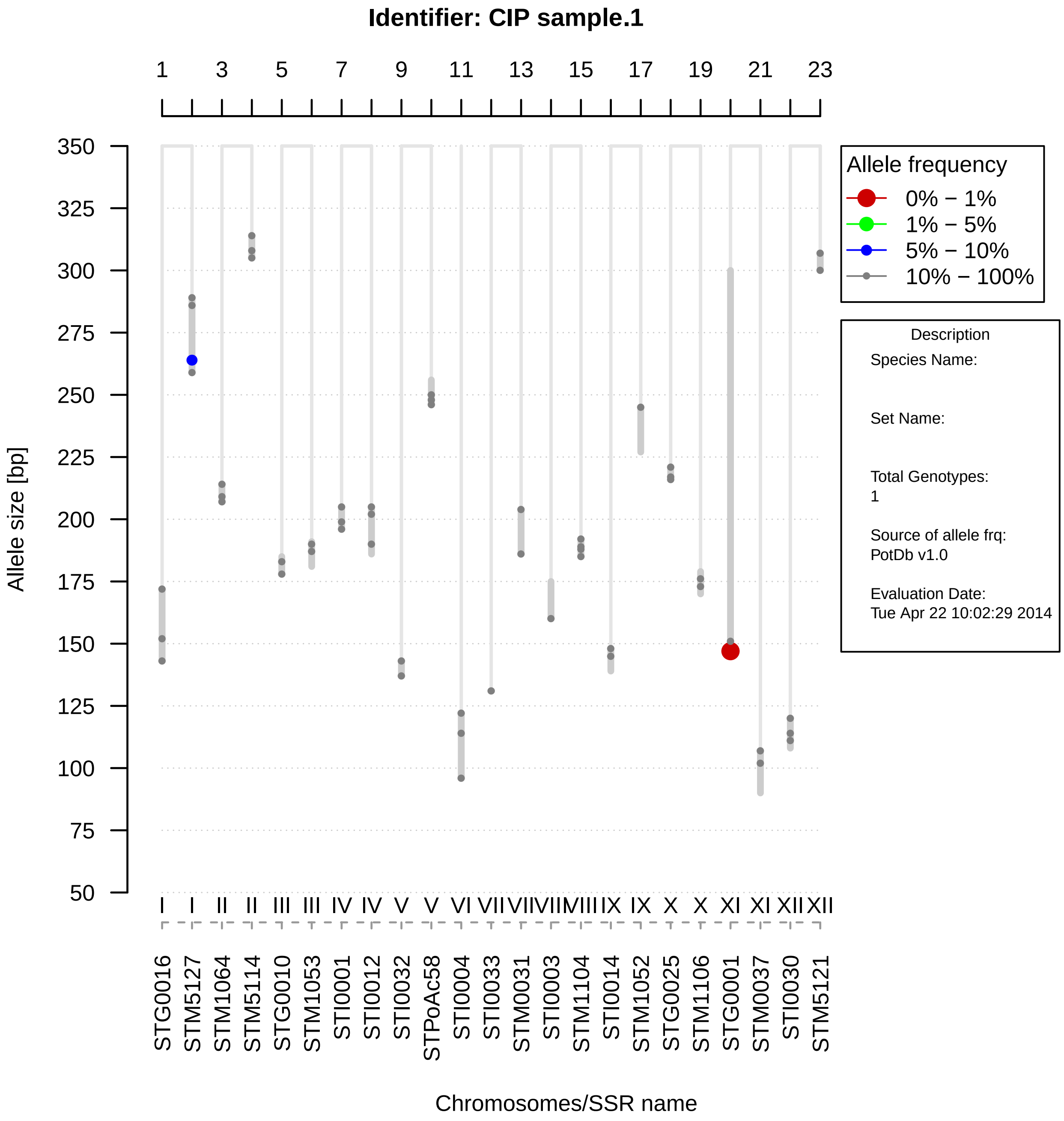


In the case of using an external reference file for allele frequencies for consistency and reporting purposes also the range of the alleles for a locus should be taken from the external database. To illustrate this point we add an additional record as in the following example.

```
xx = rbind(c("STG0001", 300, 0.01), allele.freqs)
xx$marker_size = as.integer(xx$marker_size)
xx$frequency <- as.numeric(xx$frequency)
```

```
rquipu(dat, a.subset = "sample.1", obs.alls.frq = xx, obs.alls.frq.ref = "PotDb v1.0")
```

	marker	marker_size	frequency
1	STG0001	300	0.0100
2	STG0001	147	0.0010
3	STG0001	149	0.1818
4	STG0001	150	0.0001
5	STG0001	151	0.3636
6	STG0010	178	0.2222



Description

Species Name:

Set Name:

Total Genotypes:
1

Source of allele frq:
PotDb v1.0

Evaluation Date:
Tue Apr 22 10:02:29 2014

3.4 Changing legend options: text and logo to show or hide

Some minor modifications may be used to remove text or add a logo to the legend area. By default, the 'Description' area shows the total number of accessions in the data.frame provided; this may be turned off using the parameter `show.accs.total = FALSE`. Similarly, a path to a logo image may be provided using the parameter `dir.logo`.

Another important option is to create image files and direct them to a certain directory. The parameter `img.format` can be used to produce images in either JPEG or PNG format; the parameter `dir.print` to set the output directory.

For the remaining parameters see the documentation.

3.5 Simplified charts

Lastly, the parameter `layout = "no text"` allows to draw q chart that only shows the graphic. This is meant for use in catalogues where one may want to save space. Together with the parameter resolution the overall size can then also be reduced; e.g. `res = c(600, 400)`.

```
par(mfrow = c(1, 3))
rquipu(dat, layout = "no text", res = c(600, 400))
```