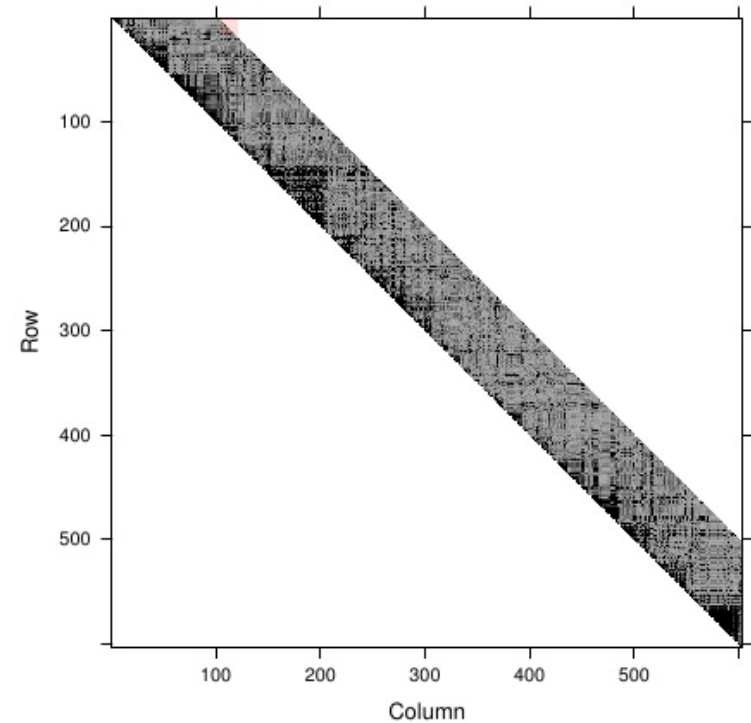**Dimensions: 603 x 603**

The important things to note are

1. there are fairly well-defined "blocks" of LD, and

2. LD is more pronounced in the Europeans than in the Africans.

The second point is demonstrated by extracting the D-prime values from the matrices (they are to be found in a slot named **x**) and calculating quartiles of their distribution:

```
> quantile(ld.ceph$D.prime@x, na.rm = TRUE)
```

4

Dimensions: 603 x 603

The important things to note are

1. there are fairly well-defined "blocks" of LD, and

2. LD is more pronounced in the Europeans than in the Africans.

The second point is demonstrated by extracting the D-prime values from the matrices (they are to be found in a slot named **x**) and calculating quartiles of their distribution:

```
> quantile(ld.ceph$D.prime@x, na.rm = TRUE)
```

4

```
      0%        25%       50%       75%      100%
0.0000000 0.1284461 0.2966520 0.6373923 1.0000000
```

```
> quantile(ld.yri$D.prime@x, na.rm = TRUE)
```

```
      0%        25%       50%       75%      100%
0.0000000 0.1066341 0.2438465 0.5098256 1.0000000
```
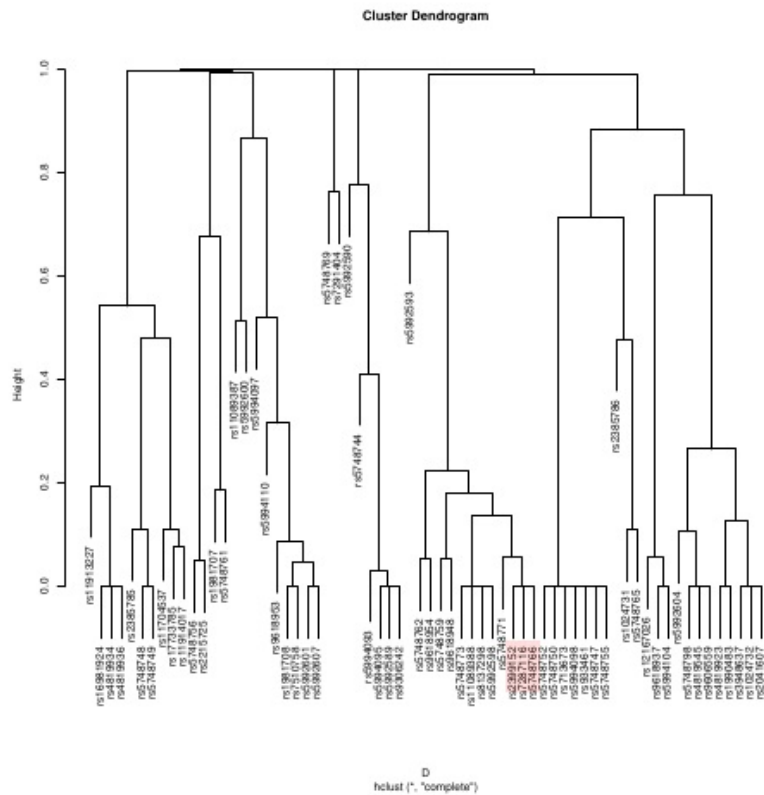
If preferred, **image** can produce colour plots. We first create a set of 10 colours ranging from yellow (for low values) to red (for high values)

```
> spectrum <- rainbow(10, start = 0, end = 1/6)[10:1]
```

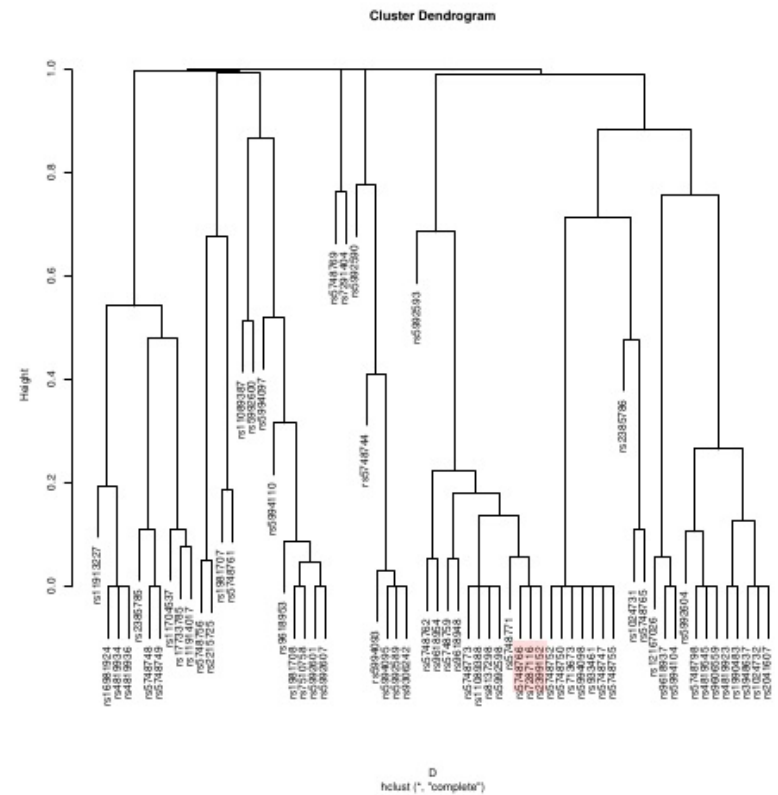and plot the image, with a colour key down its right hand side

```
> image(ld.ceph$D.prime, lwd = 0, cuts = 9, col.regions = spectrum,
+     colorkey = TRUE)
```

Cluster Dendrogram

D
hclust (*, "complete")

The interpretation of this dendrogram is that, if we were to draw a horizontal line at a "height" of 0.5, then this would divide the SNPs into clusters in such a way that the value of $(1 - R^2)$ between any pair of SNPs in a cluster would be no more than 0.5 (so that R^2 would be at least 0.5). This is carried out using the **rect.hclust** function, which returns a list giving the names of the members of each cluster together with their positions in the rows and columns of the distance matrix[2]

```
> clusters <- rect.hclust(hc, h = 0.5)
```
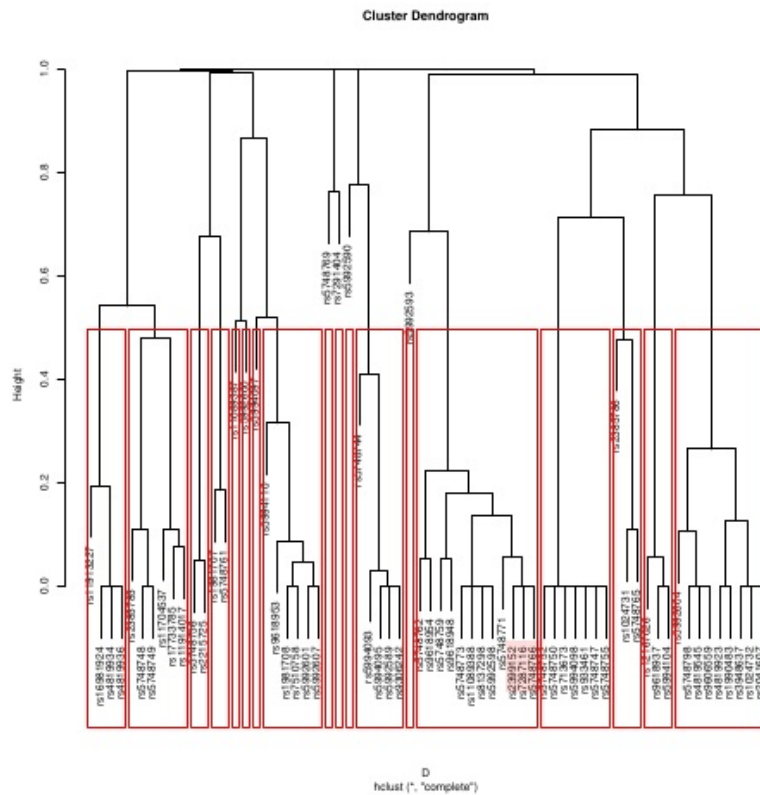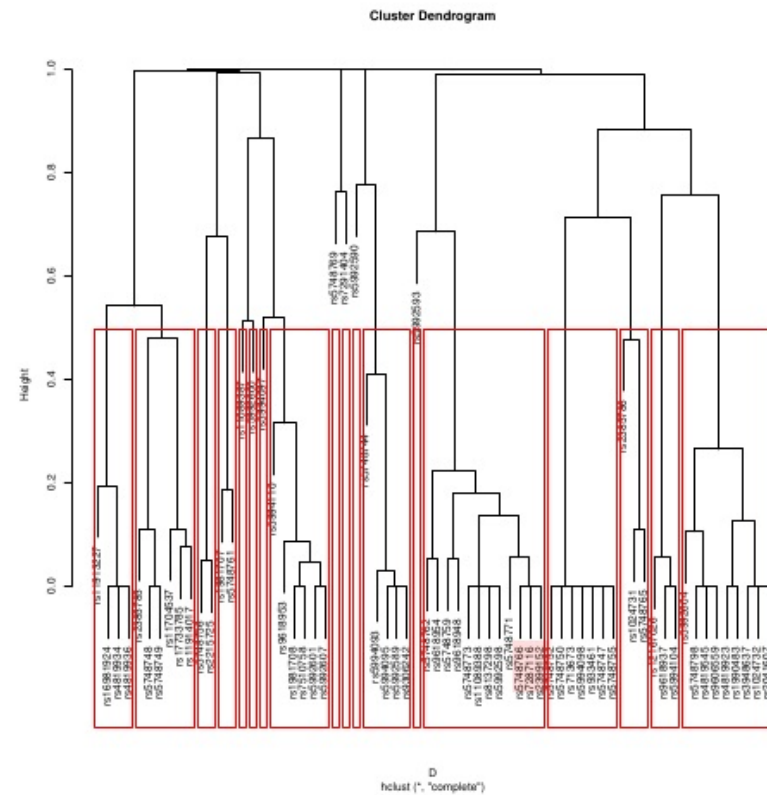
[2]The command also superimposes rectangles corresponding to these clusters on an existing dendrogram plot

Cluster Dendrogram

D
hclust (*, "complete")

It can be seen that there are 18 clusters. To have a reasonable chance of picking up an association with the SNPs in this 80kb region, we would need to type a SNP from each one of these clusters. Of these, 7 SNPs would only tag themselves!

A threshold R^2 of 0.5 might seem rather low. However, this is a "worst case" figure and most values of R^2 would be substantially better than this, particularly if an effort is made to choose tag SNPs which are in the center of clusters rather than on their edges. Also, this process has only considered tagging by single SNPs; it can be that two or more tag SNPs, taken together, can provide substantially better prediction than any one of them alone.