

## Individual assignment using STRUCTURE

$$P(i|k) = \frac{P(x_i|\gamma_k)}{\sum_k P(x_i|\gamma_k)}$$

$x_i$  = genotype of individual  $i$

$\gamma_k$  = genotype frequencies in population  $k$

For example, if  $A_1A_1$  is labeled as 1,  $A_1A_2$  as 2,  $A_2A_2$  as 3, and we assume that genotypes are in Hardy-Weinberg, then

$$P((1, 2, 2, 1, 3)|(p_{k1}, p_{k2}, p_{k3}, p_{k4}, p_{k5})) = (p_{k1}^2)(2p_{k2}q_{k2})(2p_{k3}q_{k3})(p_{k4}^2)(q_{k5}^2)$$

## Using STRUCTURE in barberry

*Berberis thunbergii*

- 85 feral, 7 horticultural, 4 cultivated
- 147 polymorphic AFLP markers

K	Mean L(K)
2	-2553.2
3	<b>-2331.9</b>
4	-2402.9
5	-2476.3

Table 1: Mean log probability of the data for  $K = 2, 3, 4, 5$  in the *Berberis thunbergii* data

# Using STRUCTURE in barberry

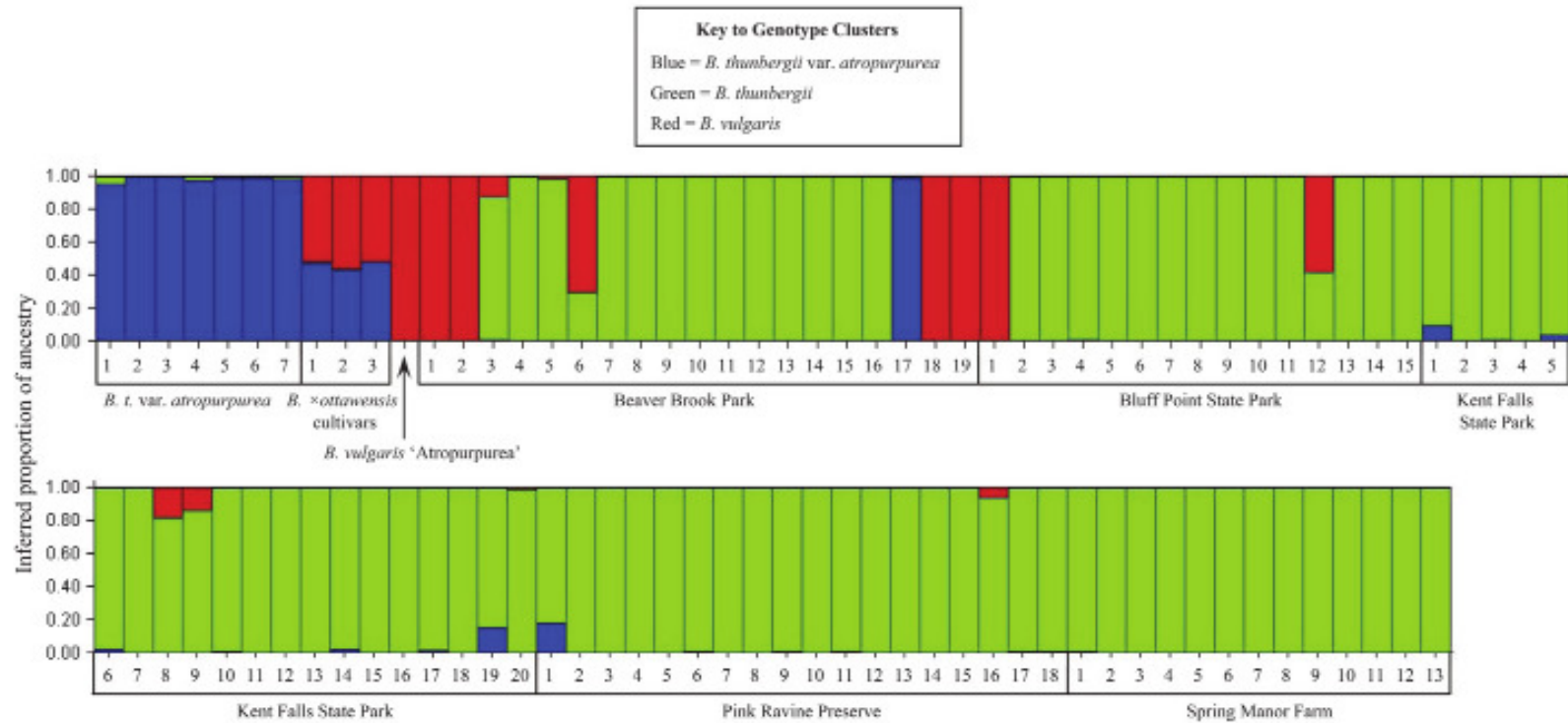
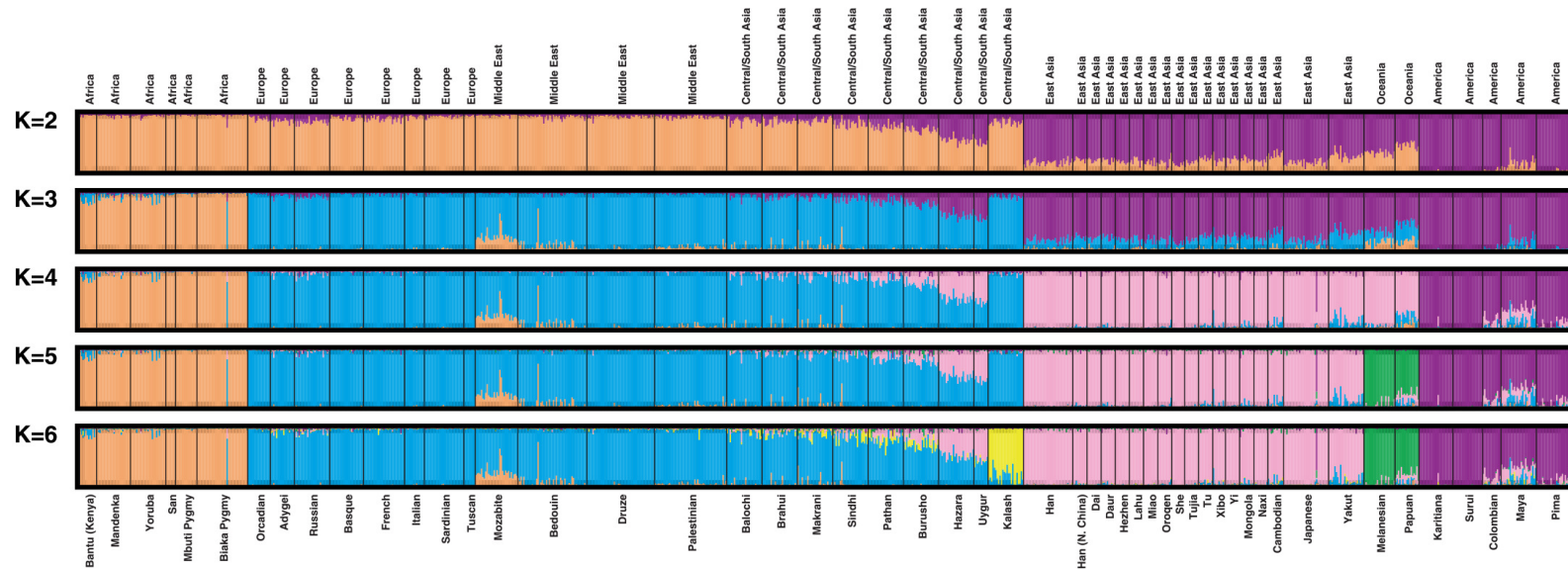


Figure 1: Analysis of AFLP data from *Berberis thunbergii*

# Using STRUCTURE in humans

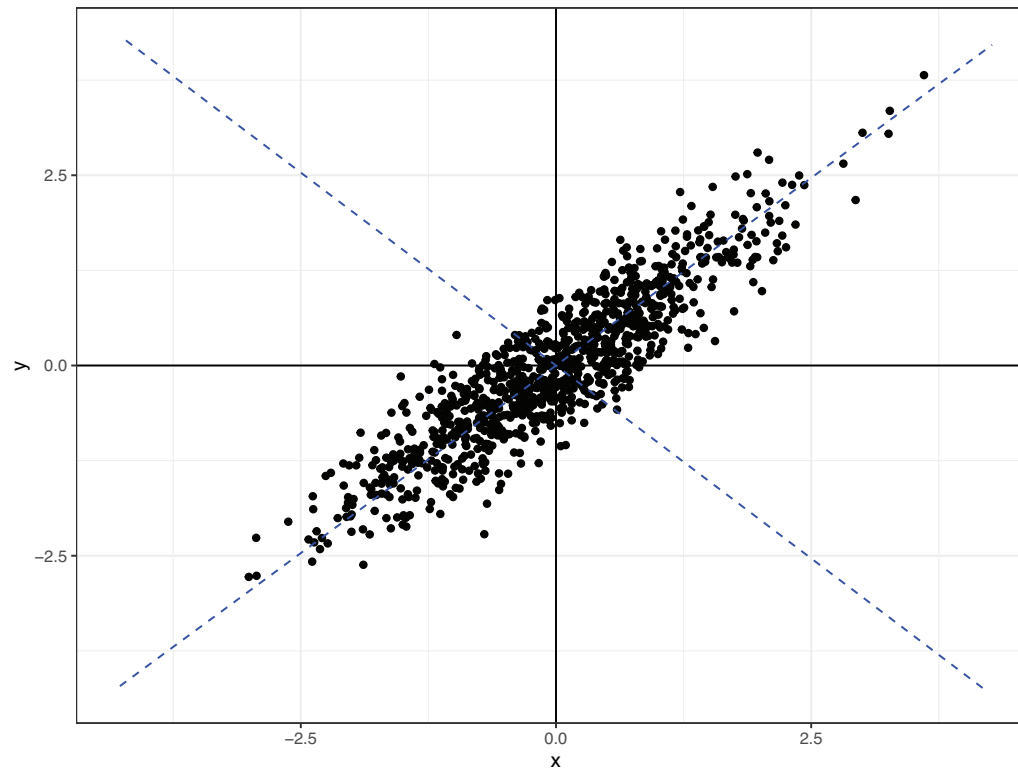
- Human Genome Diversity Cell Line Panel (HGDP-CEPH)
- 1056 individuals, 52 geographic populations, 377 autosomal microsatellite loci





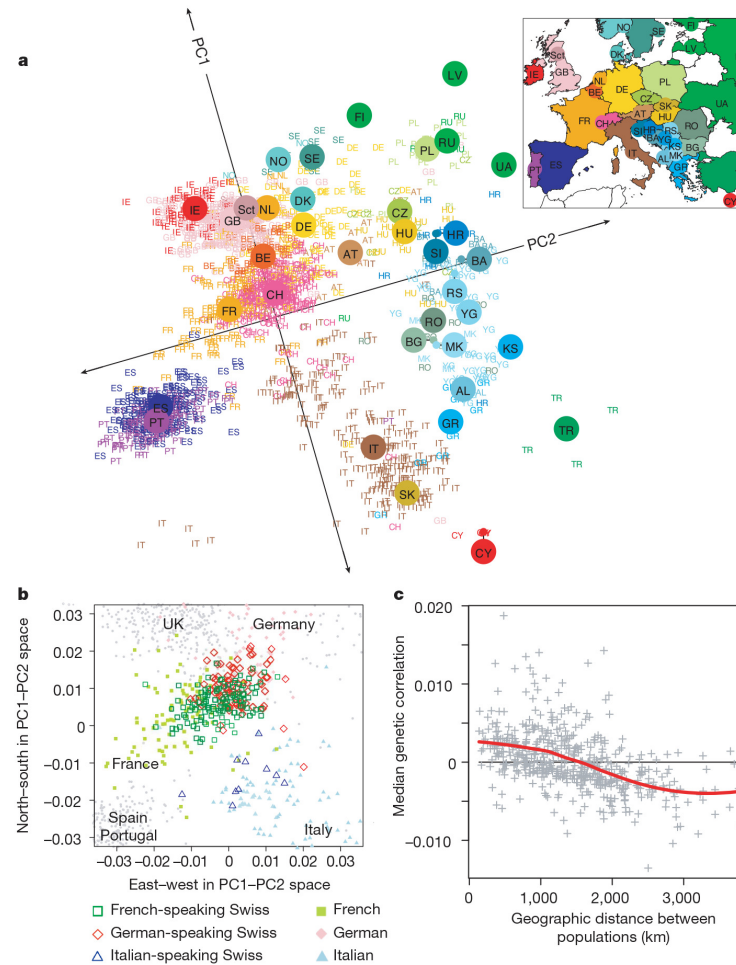
# Principal components analysis of genotypes

Principal components analysis is a “dimension reduction” method, a way of reducing a very large number of variables to a smaller, more manageable number for interpretation and analysis.

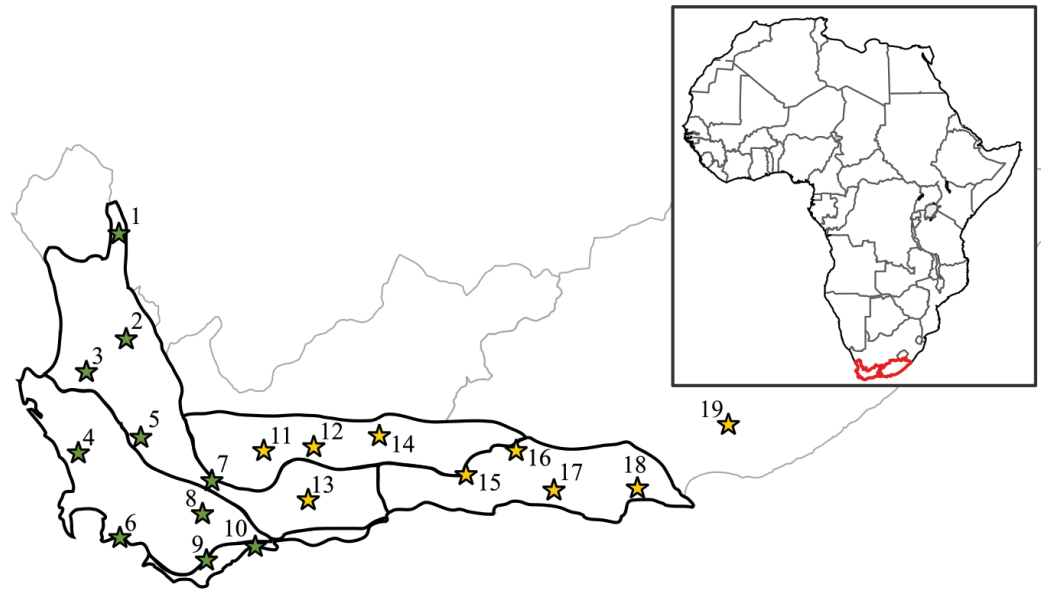
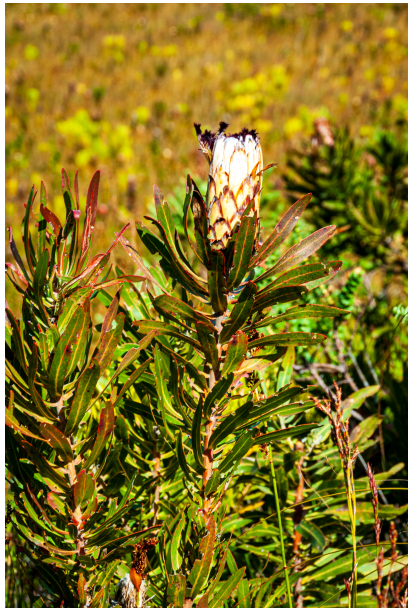


# Principal components analysis of genotypes

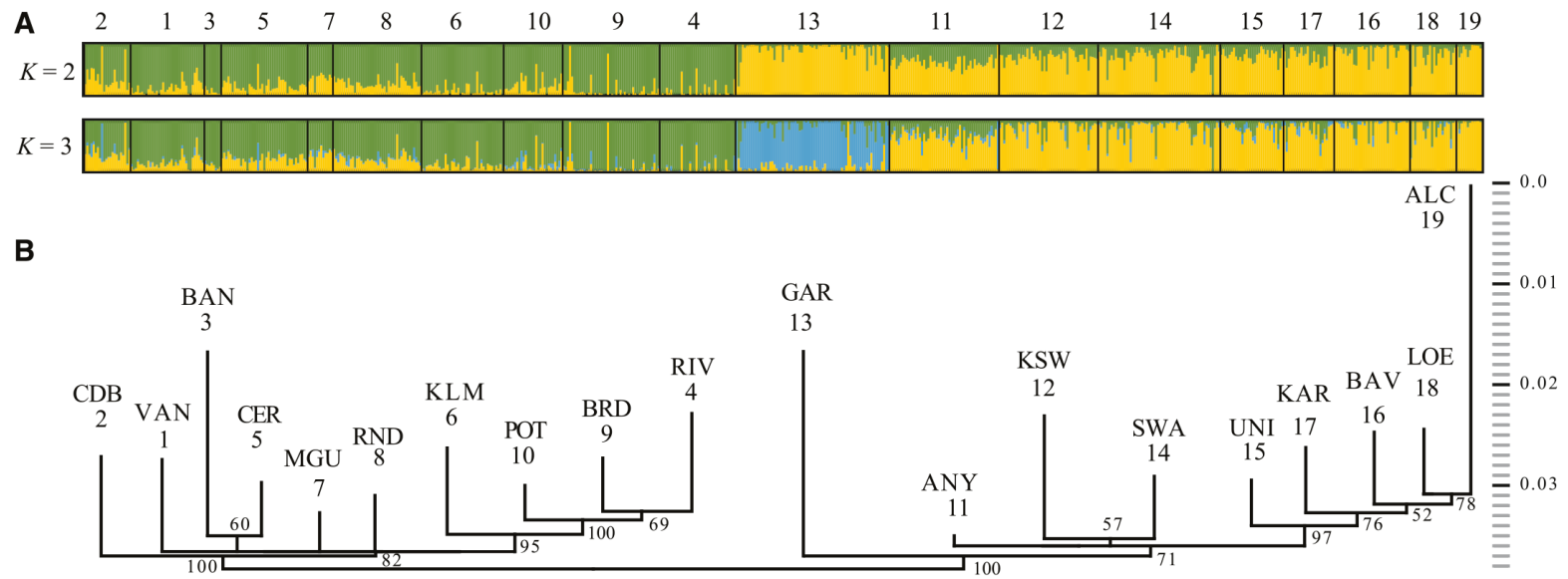
3129 Europeans, 500,568 SNP loci



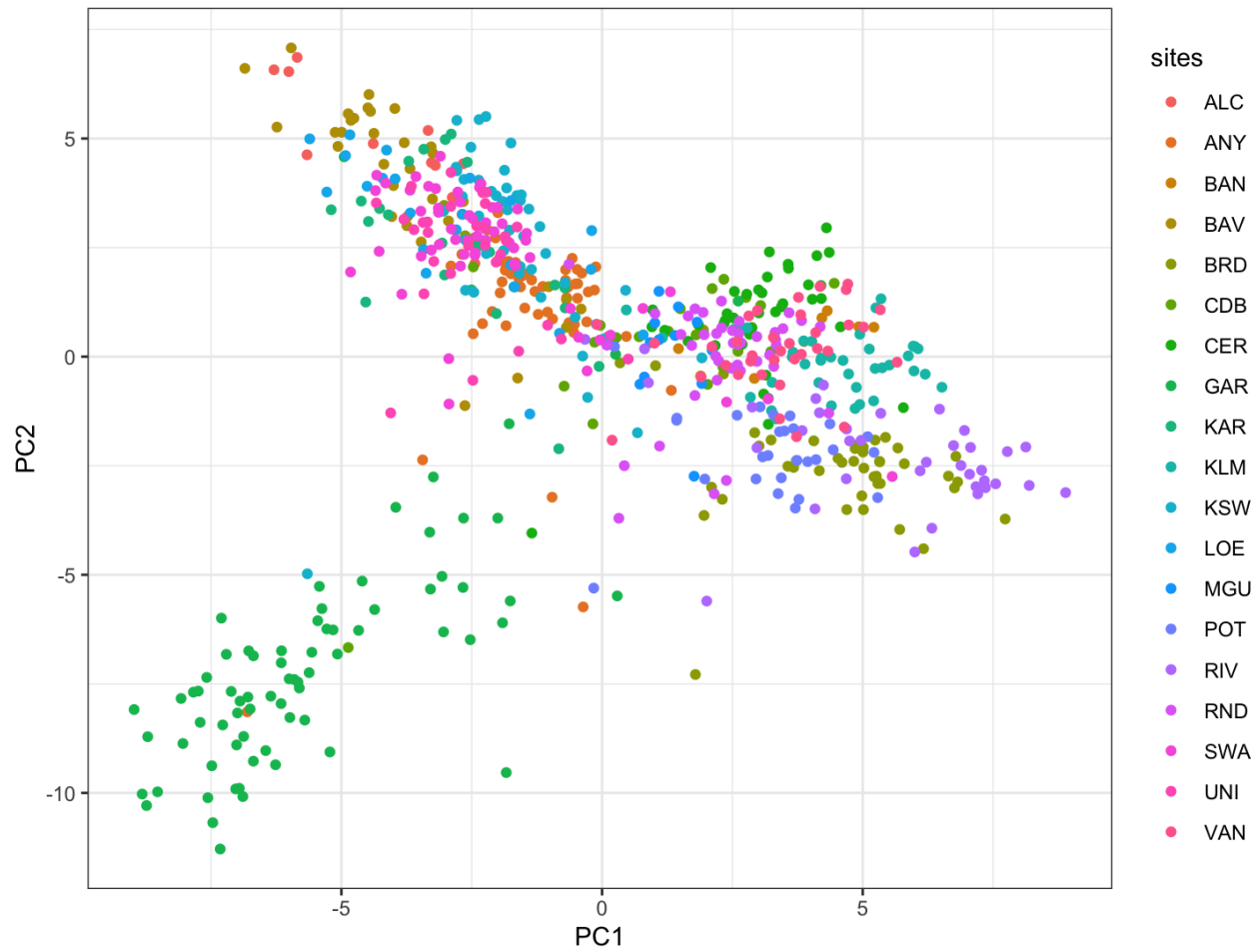
## Combining approaches: *Protea repens*



# Structure



# Principal Components Analysis



# Principal Components Analysis

