Principal Components Analysis (PCA) and Multi-Dimensional Scaling (MDS)

- SNP Data is multi-dimensional; each SNP site can be considered an axis

- PCA and MDS techniques are ways to visualize this multi-dimensional data in 2D

- PCA:

- Find the vector through the data that explains the most variance. This is the first principal component

- Find the vector that explains the most remaining variance. This is the second principal component.

- repeat...

- MDS

- Project from multiple dimensions onto 1 or 2 dimensions in a way that preserves distances present in multi-dimensional space

- (Board examples)

- (3D example)

Principal Components Analysis (PCA) and Multi-Dimensional Scaling (MDS)

- SNP Data is multi-dimensional; each SNP site can be considered an axis
- PCA and MDS techniques are ways to visualize this multi-dimensional data in 2D
- PCA:
  - Find the vector through the data that explains the most variance. This is the first principal component
  - Find the vector that explains the most remaining variance. This is the second principal component.
  - repeat...
- MDS
  - Project from multiple dimensions onto 1 or 2 dimensions in a way that preserves distances present in multi-dimensional space
  - (Board examples)
  - (3D example)

population structure

- Project SNPs in reduced dimensional space.
- Clumps of individuals represent population structure

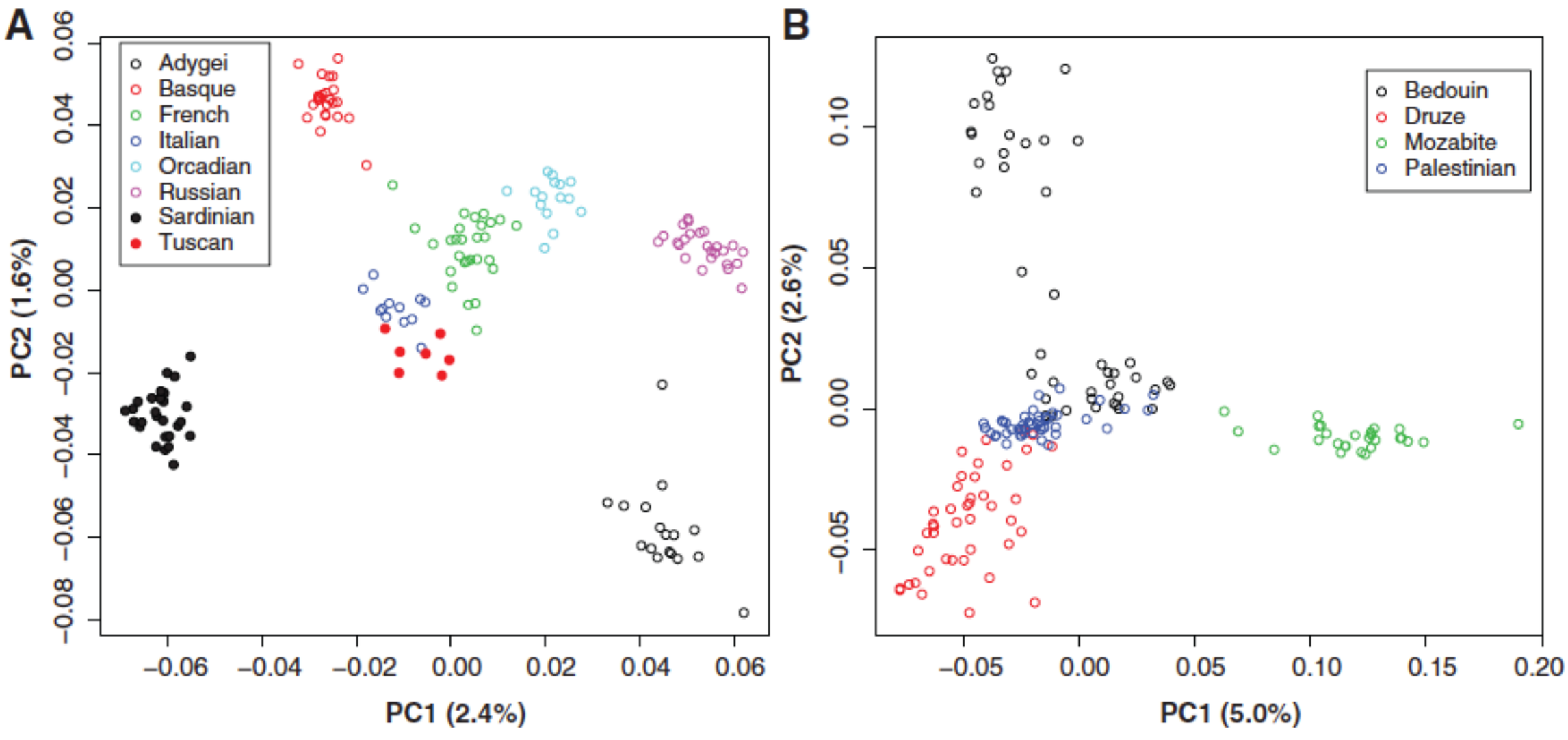

Fig. 2. Fine-scale population structure principal component analyses in two geographic regions, using all autosomal SNPs. (**A**) Europe. (**B**) The Middle East.

Li et al (2008) Science