

**BIS I 80L**  
**Rice Diversity Lab**

May 01, 2018  
Professor MALOOF  
[inmalooof@ucdavis.edu](mailto:inmalooof@ucdavis.edu)

# Outline

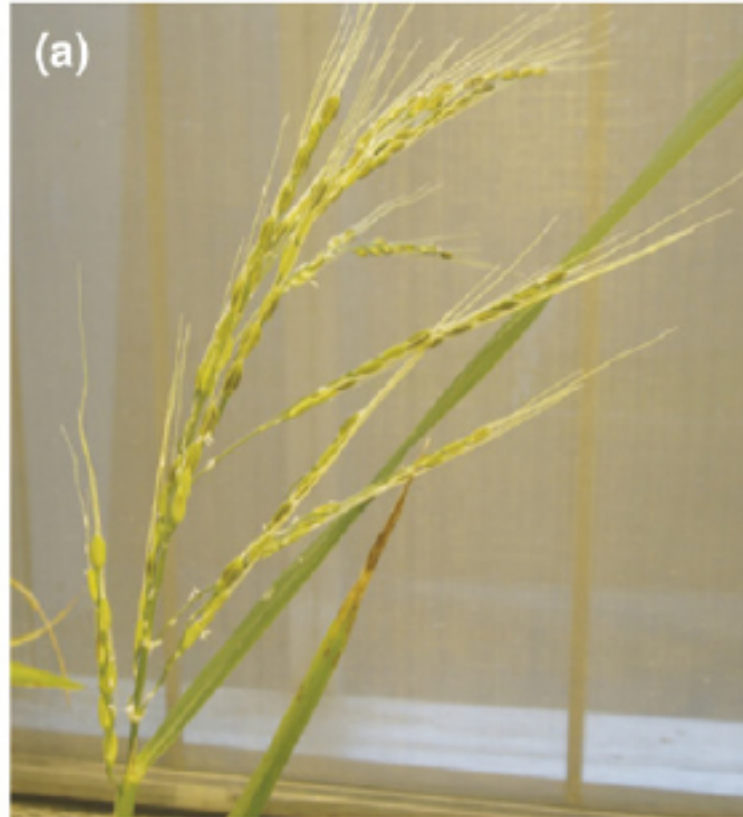
- Today:
  - Intro to rice
  - Review of population structure
- Later today or Thursday
  - Genome Wide Association Mapping (LD)

## Study system: Rice

- This Week:
  - Study diversity in ~ 400 rice strains
- Why Rice?

## Rice Domestication

Rice progenitor:  
*Oryza rufipogon*



Domesticated rice  
*Oryza sativa*



*TRENDS in Genetics*

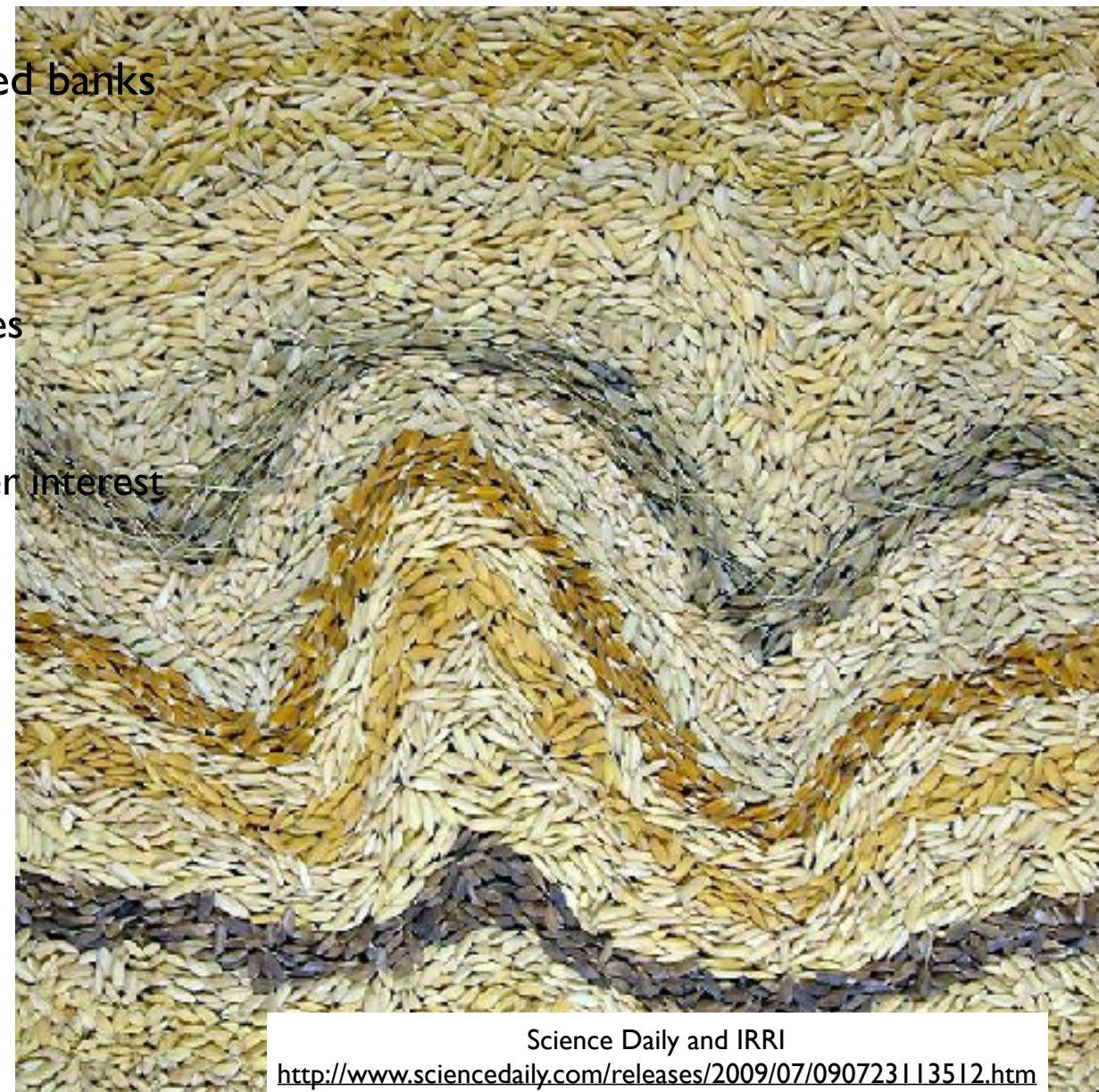
# Rice Diversity

- Two main sub populations: *Indica* and *Japonica*
  - Thought to have been domesticated independently from *rufipogon*
  - Estimated divergence time: 100,000 years ago or more.
- *Indica*
  - Grown in lowland tropical areas
  - S. and S.E. Asia and China
  - Further subdivided into *indica* and *aus*
- *Japonica*
  - Grown in lowland and high-elevation upland areas of tropical SE Asia
  - Also grown in colder temperate climates including NE Asia, Europe, Western US, Chile, Australia
  - Subdivided into *tropical japonica*, *temperate japonica*, and *aromatic*



# Using natural variation for rice improvement

- Urgent need for crop improvement:
  - World population expected to grow to 9,000,000,000 by 2050
  - Climate change and increasing unpredictability will reduce yield
  - Increasing demand for meat and biofuels put further strains on agriculture
- One path forward is to use the natural genetic diversity (“natural variation”) already present in rice
  - 120,000 different rice strains have been deposited in seed banks
  - These harbor different genetic variants
    - Many are just random
    - some provide adaptation to specific environments/stresses
      - drought, flooding, heat, pathogens, etc
    - some determine specific grain characteristics of consumer interest
      - stickiness
      - grain length
      - aromatic (basmati, jasmine)
      - color
- Identify genes or genomic locations of variants
  - Why?





## We will focus on root traits

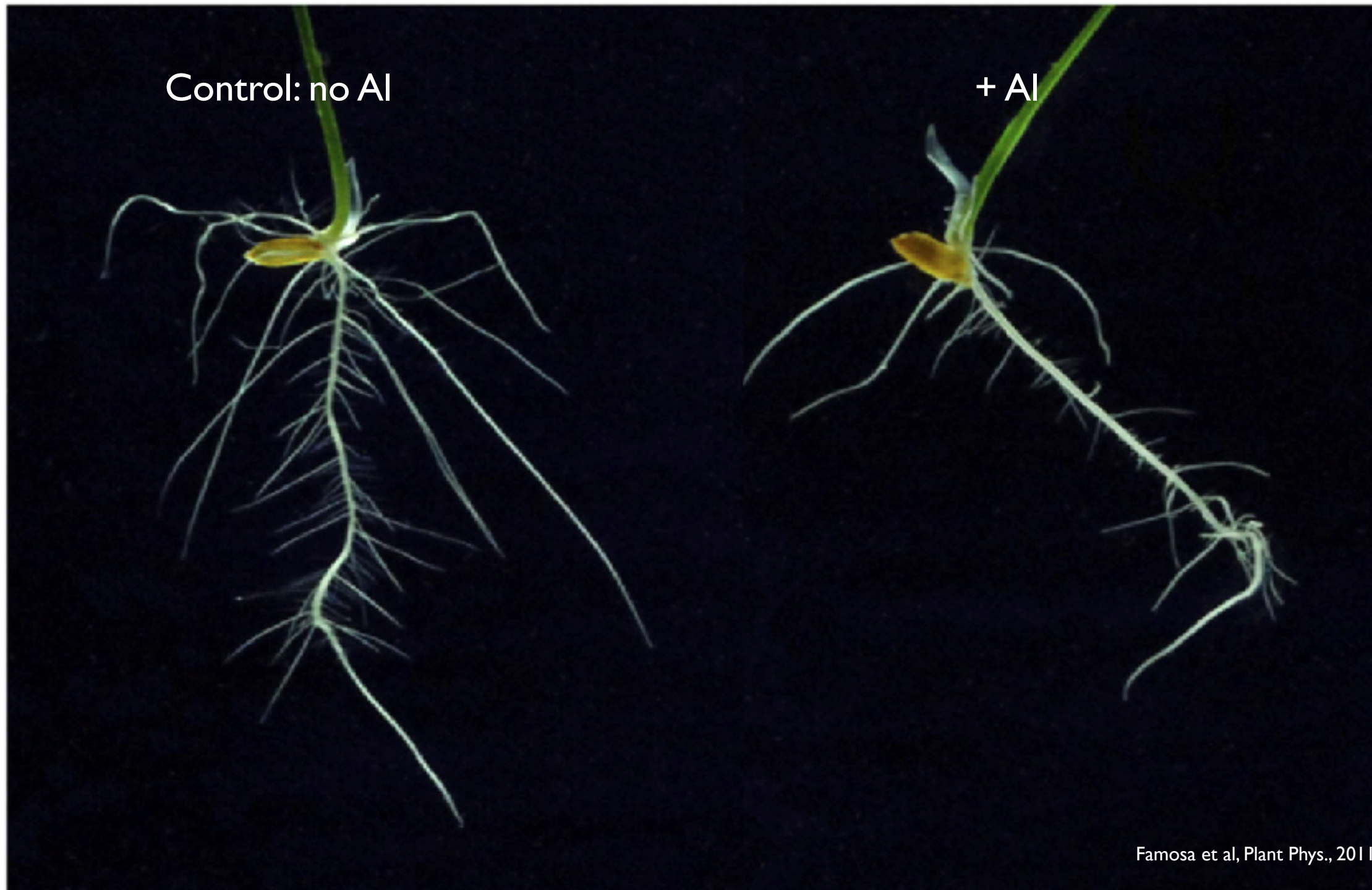
- Why Roots?





## Root Aluminum Tolerance

- Aluminum: Most abundant metal in earth's crust.
- Solubilized under highly acidic soil conditions
- Toxic to plant roots
- Can we find variation in resistance to Aluminum among rice varieties?





# Study Design

- 413 Rice varieties from 82 countries
- Phenotypes:
  - Roots in control and Al<sup>+</sup> conditions
  - Grain length
  - Amylose content
  - Flowering time
  - ...
- Genotypes:
  - Sequence 20 strains, find common SNPs
  - Design Affymetrix “SNP chip”
  - Assay 44,000 Single Nucleotide Polymorphisms on each of the 413 varieties
  - Yields ~ 1 SNP per 10kb

## Overview of Questions about Rice Data:

- Population Structure:
  - Are these samples from a single randomly mating population or are there sub-populations?
    - Multi-dimensional scaling plot of genotype data to examine diversity and relationships
    - fastStructure to assign varieties to ancestral populations
- Can we find SNPs associated with Aluminum tolerance or other traits?
  - What are the genes underlying these SNPs?

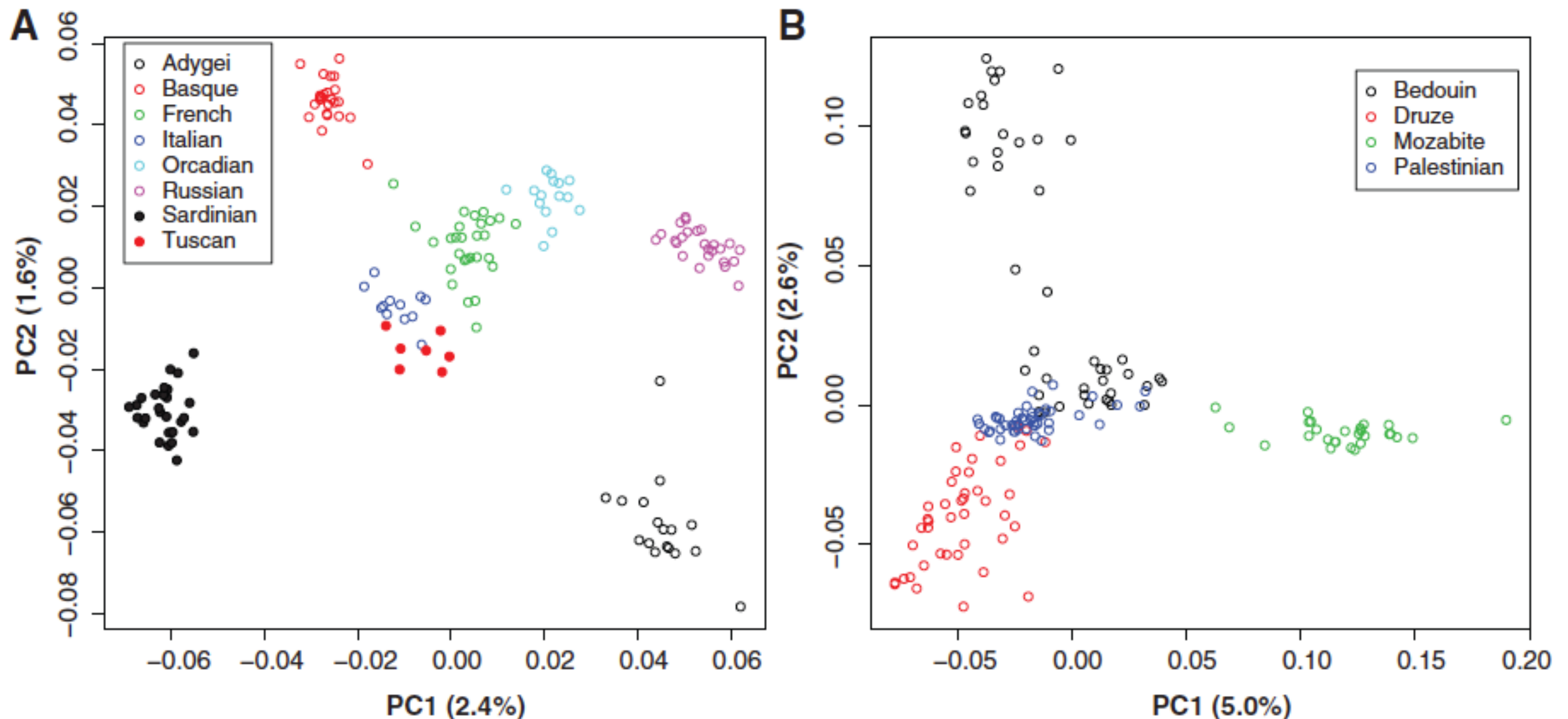
# Principal Components Analysis (PCA) and Multi-Dimensional Scaling (MDS)

- 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)



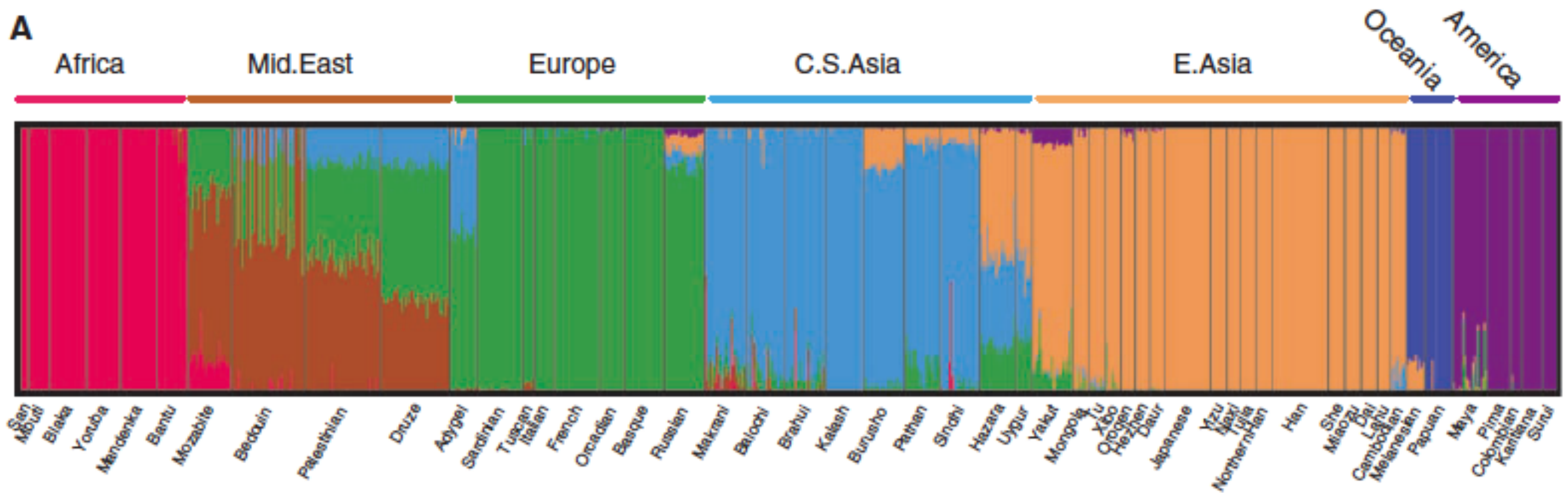
# population structure

- 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.

# Model based assignment of individuals to populations



- a priori decide on the number of likely ancestral populations
- use an evolutionary genetics model to assign the most likely ancestry to each individual

## Questions to be considered in answering assignment questions:

- Is there population structure?
  - How does structure relate to region of origin?
  - How does structure relate to amylose content?
  - How does structure relate to Aluminum tolerance
- Are there GWAS hits for the trait you are studying (TBD)?
- What are the candidate genes for your GWAS hit?