

# Species Tree Inference Lab – Part II

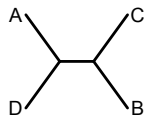
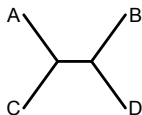
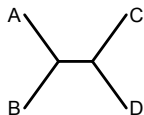
## The SVDquartets Program

Laura Kubatko  
Departments of Statistics and  
Evolution, Ecology, and Organismal Biology  
The Ohio State University  
kubatko.2@osu.edu

July 30, 2013

## Species tree estimation using algebraic statistics

**Main idea:** use the observed site pattern distribution to provide information about which of the three possible splits for a set of four taxa is the true split.



The program [SVDscores](#) computes a score for each split in a given quartet of taxa and chooses the split with the best (lowest) score.

### Algorithm

- (1) Generate all quartets (small problems) or sample quartets (large problems)
- (2) Run all quartets sampled through [SVDscores](#)
- (3) Use a quartet method to build the species tree

### Algorithm

- (1) Generate all quartets (small problems) or sample quartets (large problems)
  - (2) Run all quartets sampled through [SVDscores](#)
- Place the data in a file called [data.phy](#) in PHYLIP format. The file [splits](#) must be placed in the same directory.
    - ▶ At the end of the first line, add two numbers: the first is an indicator (1 for all quartets, 0 to sample) and the second indicates how many quartets to sample.
  - Type [SVDquartets](#) at the prompt.
    - ▶ This forms quartet data sets and runs [SVDscores](#). The output is written to [quartets.out](#).

### Algorithm

- (3) Use a quartet method to build the species tree
  - The [quartets.out](#) file is compatible with many quartet phylogeny programs.
  - We'll use the program [Quartet MaxCut](#) (Snir and Rao 2010) to build the tree estimate.

## Species tree estimation using algebraic statistics

### Example: rattlesnake data

Get the zip file `svd2013.zip` in the `/class/shared/` directory and unzip it..

```
> unzip svd2013.zip
```

Copy the file containing the rattlesnake data to a file called `data.phy`:

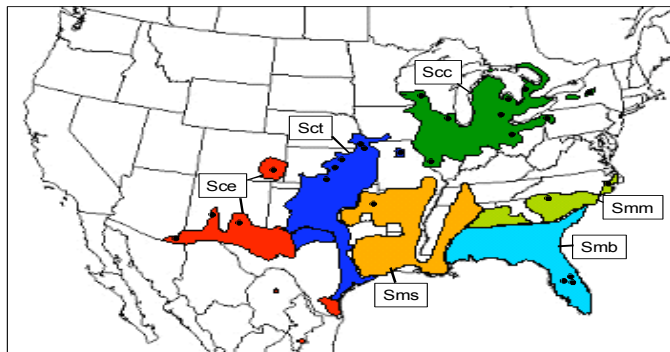
```
> cp data_snakes.phy data.phy
```

Commands to carry out inference:

```
> SVDquartets
```

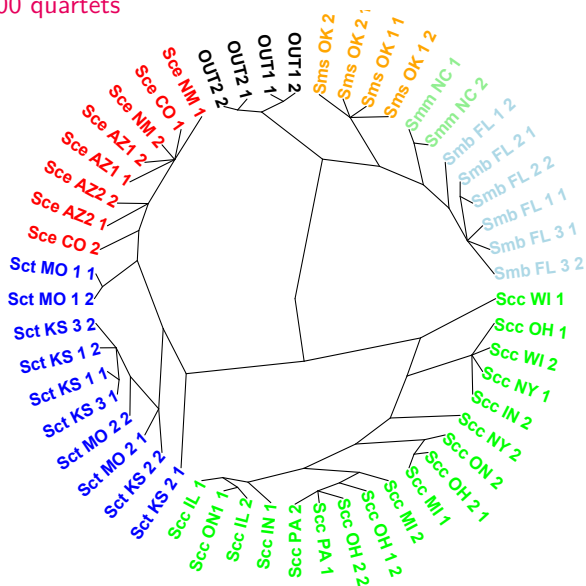
```
> findCut.linux qrtt=quartets.out otre=snakes.out
```

## Species tree estimation using algebraic statistics



# Species tree estimation using algebraic statistics

Using 3,000 quartets





If you're bored ...

[Tutorial on the use of STEM is posted on my course wiki page.](#)

- I've created four datasets under varying conditions:
  - M1 No hybridization, long intervals between speciation events.
  - M2 No hybridization, short intervals between speciation events.
  - M3 Low-levels of hybridization - B is a hybrid of A and C (species tree as in M1 and M2).
  - M4 Extensive hybridization - B is a hybrid of A and C (species tree as in M1 and M2).
- All data sets have 6 species, 2 individuals/species, and 10 loci.
- **GOAL: match the data set to the condition listed above**  
Solutions are linked to on my course wiki page.