

# MID8: Analysis II

---

3/5/15

Shannon's office hours:

Friday 3-4pm in 16-319

*Sunday 3-5pm in 528 Simmons Hall*

Monday 3-4pm in 16-319

Noreen and Leslie have OH on Monday

Noreen: 2-4pm in 16-429C ~~✓~~

Leslie: 1-2pm in 16-429C ~~✓~~

# Announcements

## Journal club

Tuesday: Meet in 16-336 at 1:30pm (speakers 1:15pm)

**ModI Abstract + Data Summary:** *Stellar*  
Due Saturday, March 14th at 5pm.

Leslie R. will give a talk on Abstract on Tuesday at 11am

**Read the wiki — there are required figures + suggestions on ways to present your data.**

**ModI Primer Design Memo:**

Due Thursday, March 19th at 10pm

Data was sent to you last night. Let me know if you did not get it.

Suggestion: complete your Abstract + Data summary first

# Announcements

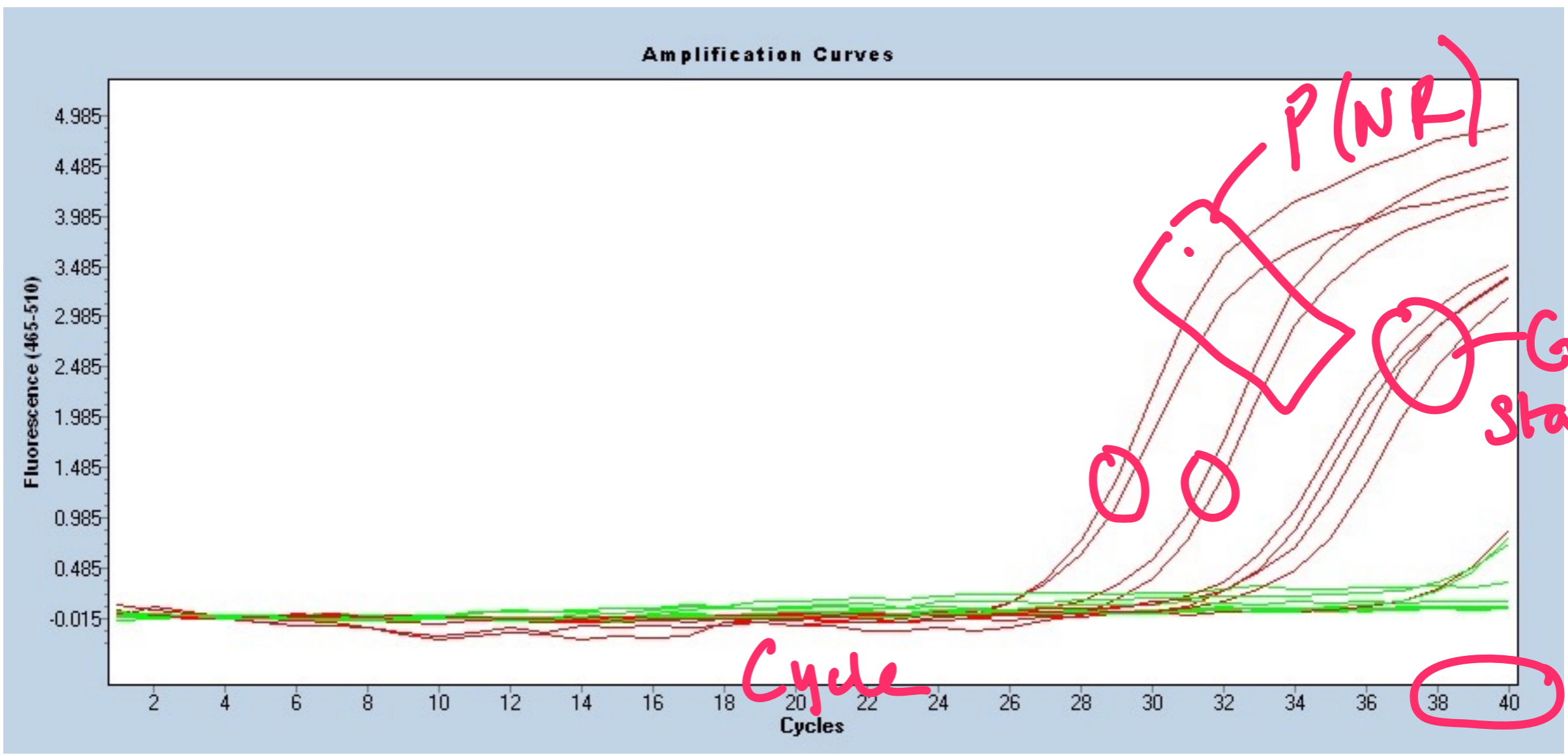
## Comments about Figure + Caption + R&D Bullets:

1. What was the overall goal of these data/figure? State this concisely in an intro sentence.
2. What was your expected result? State this concisely in a bullet point (i.e what are the expected band sizes on your gel?).
3. What was the result? State this concisely in a bullet point or two (i.e. were your bands present at the expected size? Why or why not?).
4. What evidence do you have that your result is correct or incorrect? (i.e. what controls did you perform and did they work as expected? What was the expectation?)
5. In sum, what do these data suggest or indicate? Think about how we used these data.
6. What does this motivate you to do next?

# Comments on ALV experiment:

A, B +  
C -

F

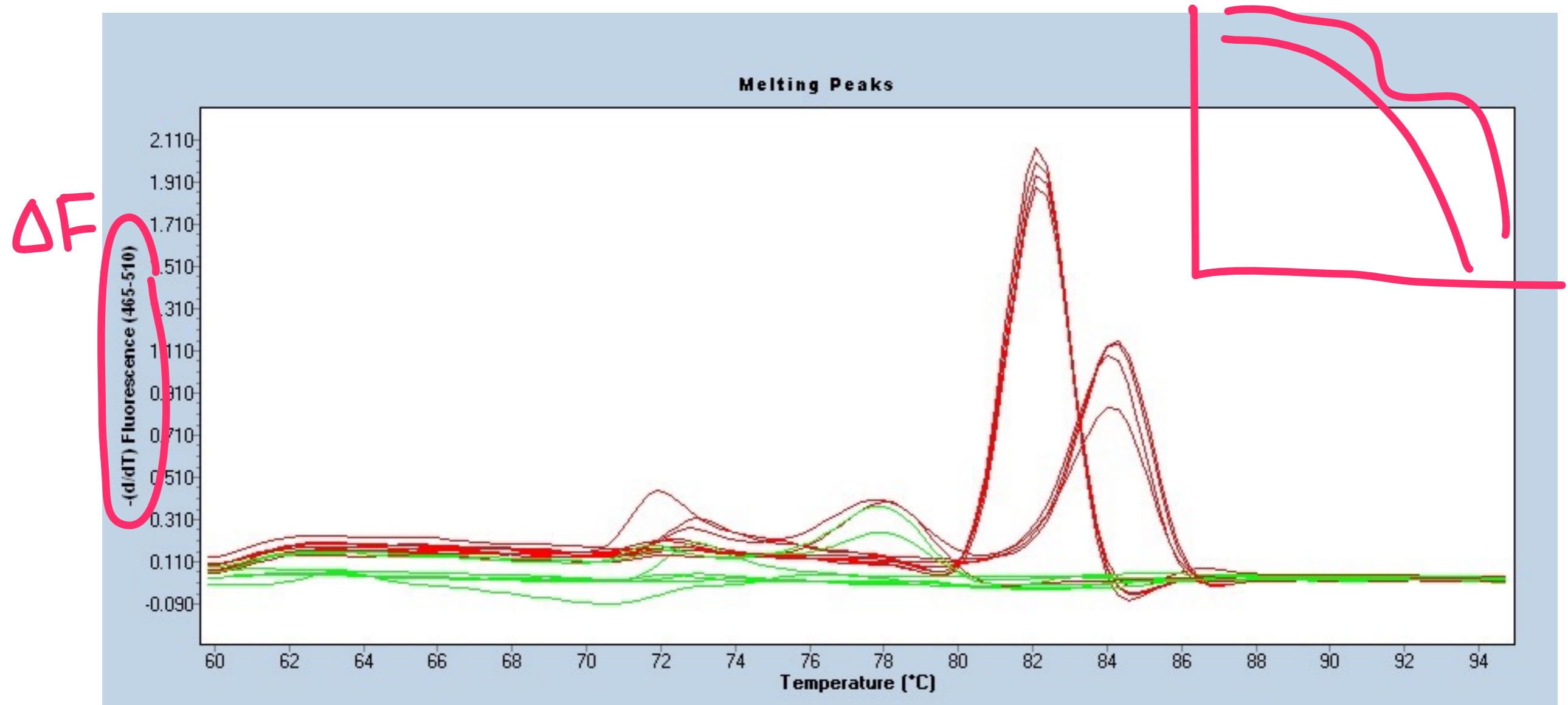


Cycle

$C_t$  - threshold

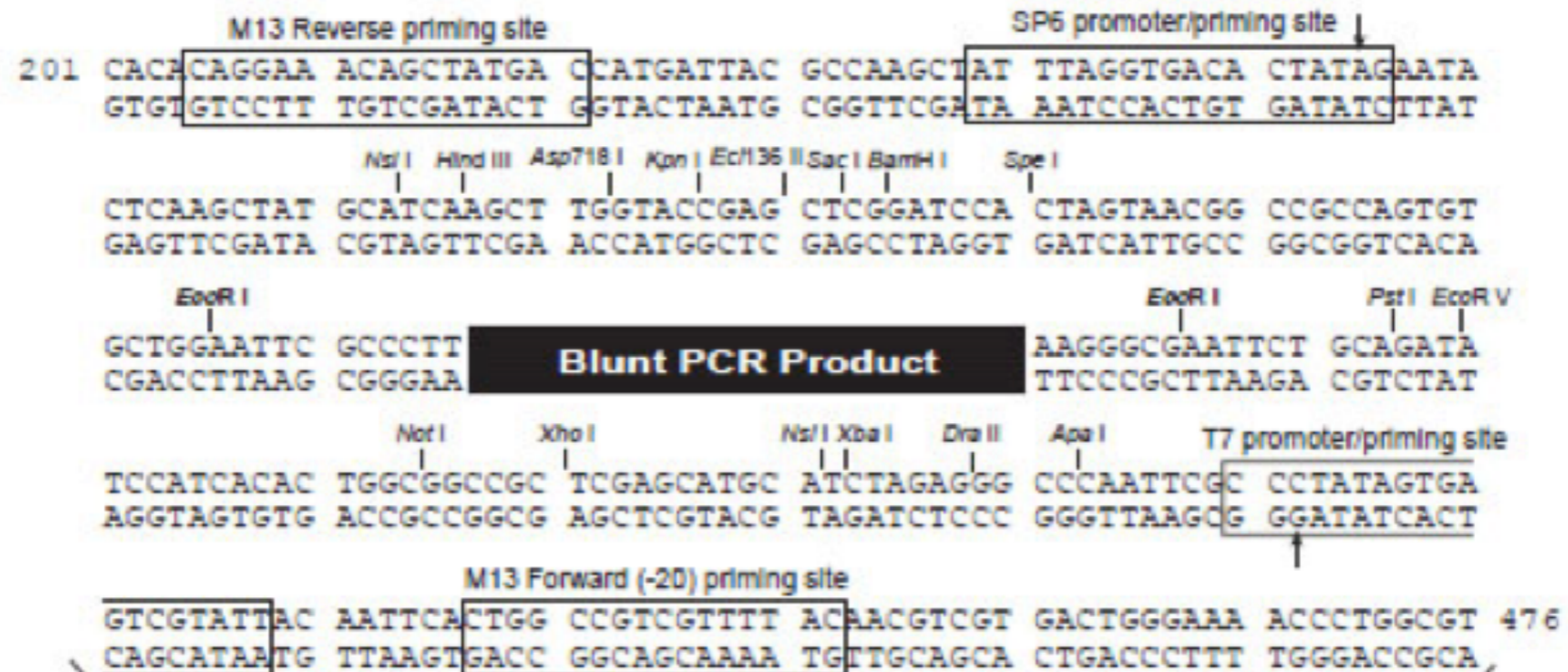
38+

# Comments on ALV experiment:



<http://www.idtdna.com/pages/support/technical-vault/decoded-archive/core-concepts/decoded/2014/01/20/interpreting-melt-curves-an-indicator-not-a-diagnosis>

# Insert Orientation



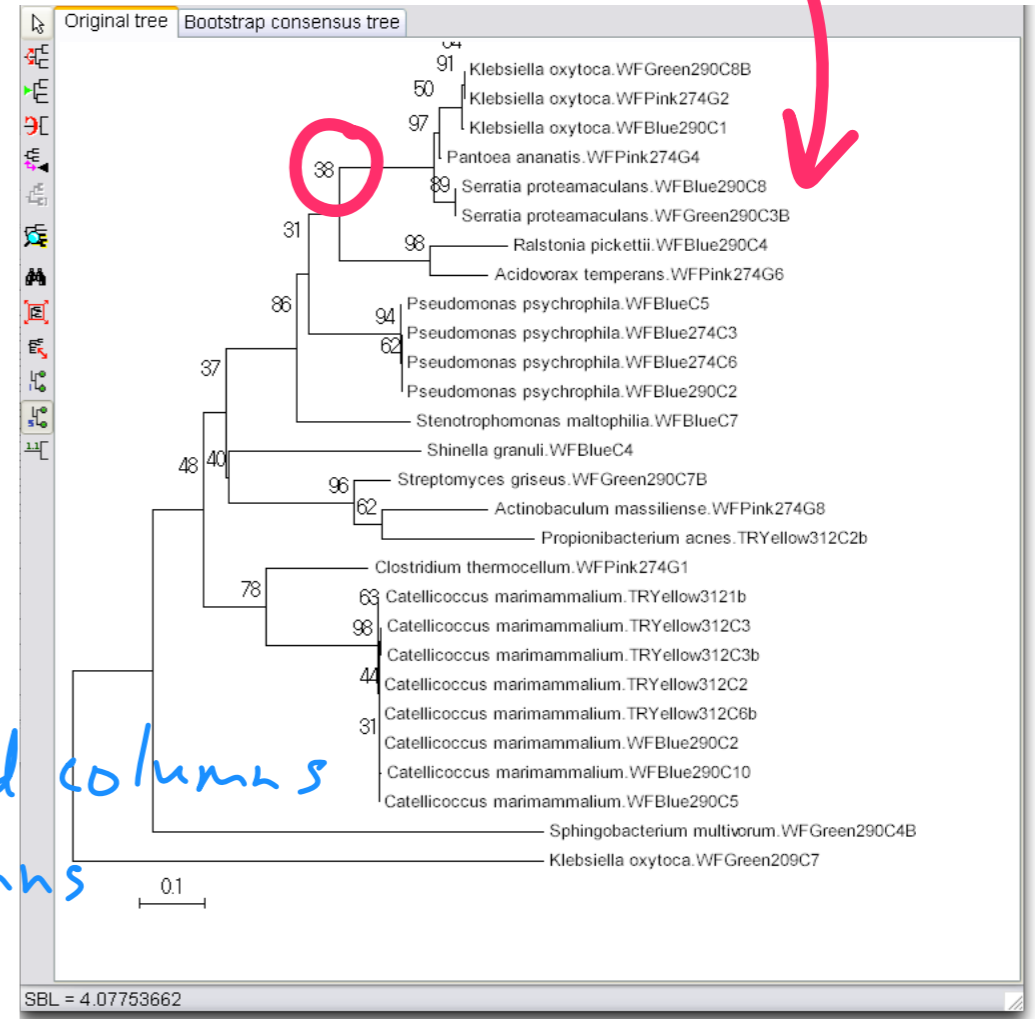


# Practical bootstrapping example (your tree):

.mas alignment file

Species/Abbrv	Group Name	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20					
1. Actinobaculum massiliense.WFPink274G8		A	A	T	A	C	G	T	T	C	T	C	G	G	G	C	T	T	G	T	A	C	A	C	A	C
2. Catellicoccus marimammalium.TRYellow3121b		A	A	T	A	C	G	T	T	C	C	C	G	G	G	C	T	T	G	T	A	C	A	C	A	C
3. Catellicoccus marimammalium.TRYellow312C2		A	A	T	A	C	G	T	T	C	C	C	G	G	G	C	T	T	G	T	A	C	A	C	A	C
4. Catellicoccus marimammalium.TRYellow312C3		A	A	T	A	C	G	T	T	C	C	C	G	G	G	C	T	T	G	T	A	C	A	C	A	C
5. Catellicoccus marimammalium.TRYellow312C3b		A	A	T	A	C	G	T	T	C	C	C	G	G	G	C	T	T	G	T	A	C	A	C	A	C
6. Catellicoccus marimammalium.TRYellow312C6b		A	A	T	A	C	G	T	T	C	C	C	G	G	G	C	T	T	G	T	A	C	A	C	A	C
7. Catellicoccus marimammalium.WFBlue290C10		A	A	T	A	C	G	T	T	C	C	C	G	G	G	C	T	T	G	T	A	C	A	C	A	C
8. Catellicoccus marimammalium.WFBlue290C2		A	A	T	A	C	G	T	T	C	C	C	G	G	G	C	T	T	G	T	A	C	A	C	A	C
9. Catellicoccus marimammalium.WFBlue290C5		A	A	T	A	C	G	T	T	C	C	C	G	G	G	C	T	T	G	T	A	C	A	C	A	C
10. Clostridium thermoCELLUM.WFPink274G1		A	A	T	A	C	G	T	T	C	C	C	G	G	G	C	T	T	G	T	A	C	A	C	A	C
11. Klebsiella oxytoca.WFBlue290C1		A	A	T	A	C	G	T	T	C	C	C	G	G	G	C	T	T	G	T	A	C	A	C	A	C
12. Klebsiella oxytoca.WFGreen209C7		A	A	T	A	C	G	T	T	C	C	C	G	G	G	C	T	T	G	T	A	C	A	C	A	C
13. Klebsiella oxytoca.WFGreen290C8B		A	A	T	A	C	G	T	T	C	C	C	G	G	G	C	T	T	G	T	A	C	A	C	A	C
14. Klebsiella oxytoca.WFPink274G2		A	A	T	A	C	G	T	T	C	C	C	G	G	G	C	T	T	G	T	A	C	A	C	A	C
15. Pantoea ananatis.WFPink274G4		A	A	T	A	C	G	T	T	C	C	C	G	G	G	C	T	T	G	T	A	C	A	C	A	C
16. Propionibacterium acnes.TRYellow312C2b		A	A	T	A	C	G	T	T	C	C	C	G	G	G	C	T	T	G	T	A	C	A	C	A	C

Tree #1 original seq  
Tree #2 5, 1, 5, 7, 12, 2, ...  
:  
} 500 MEGA  
Tree #n



In class I said the rows (species) were shuffled but it is actually the columns that are shuffled so;

Tree #1 contains 1 → 1400  
Tree #2 contains one or more shuffled columns  
Tree #3 contains other shuffled columns  
:

123456789 → 1400

# What's all the stuff on the Fast UniFrac site?

The screenshot shows the Galaxy web interface with the FastUniFrac tool selected. The tool options are listed below:

- Cluster samples** Uses the UniFrac metric to cluster the samples based on phylogenetic lineages they contain.
- Jackknife Sample Clusters** Performs statistical resampling and will allow you to see how confident you should be in the sample clustering results.
- PCoA** Uses the UniFrac metric to perform principal coordinates analysis on your samples, allowing you to see whether different types of samples are separated in different dimensions.
- P Test Significance** Tells you which pairs of samples are significantly different using the P Test.
- Sample counts** Tells you how many sequences are in each sample.
- Sample Distance Matrix** Shows you the UniFrac distances between each pair of samples.

#Sample	ID	Male	Female	Description
274	Male			gull from Carson Beach
312	Female			gull from Carson Beach
290	Female			gull from Carson Beach

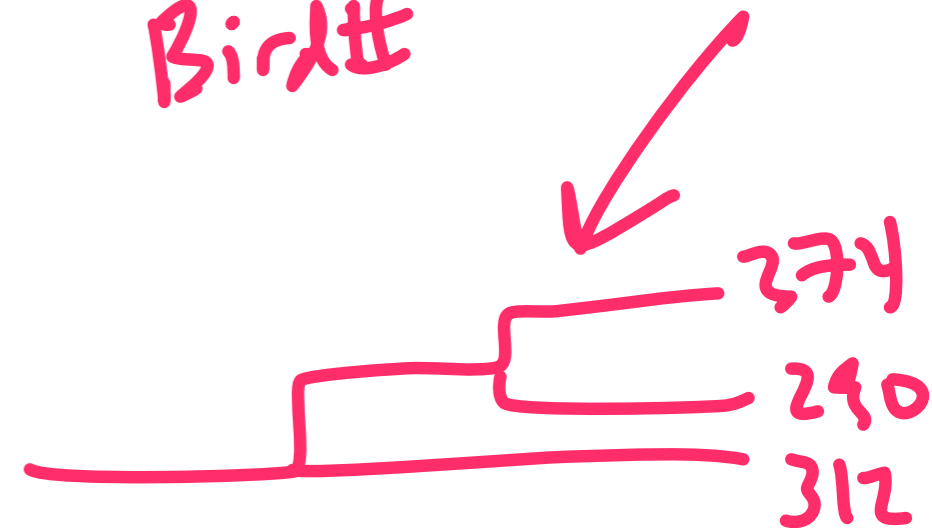
\* Input a tree (.nwk) w/ all of the class sequences

\* ID file Excel

⋮  
Seq

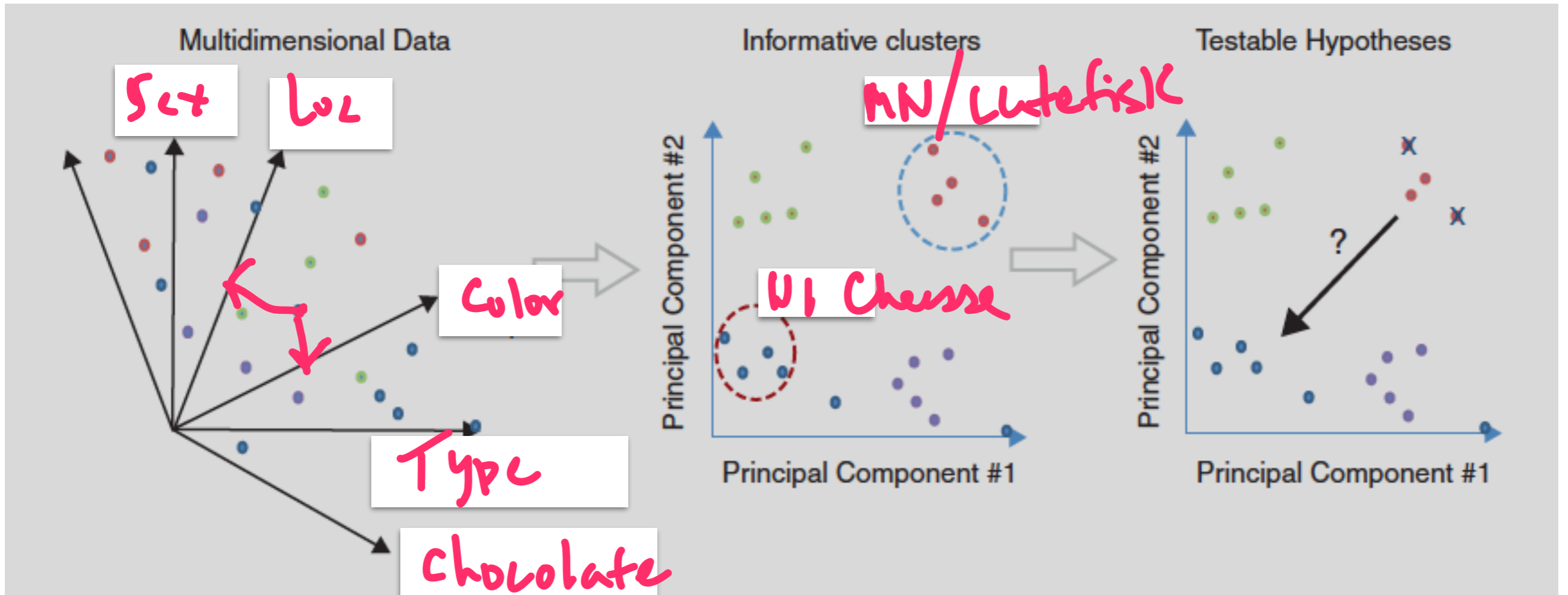
⋮  
Bird#

\* Mapping File





# (Hopefully) intuitive PCoA:



## Today in lab (practice analysis):

- Locate samples on Genewiz site that are good — use the spreadsheet to help you
- Analyze at least 9 sequences
- **Post your full length sequence text files** from ApE (those you use for your alignment in MEGA). **\*\*Make sure there is ONLY SEQUENCE in those files\*\***
- **Post your MEGA alignment file.** This is the .mas file (the multicolored spreadsheet-like window).
- *We will make sure that all sequences are posted as text files by Saturday at noon.*
- Do cool analysis + learn something no one knew before.