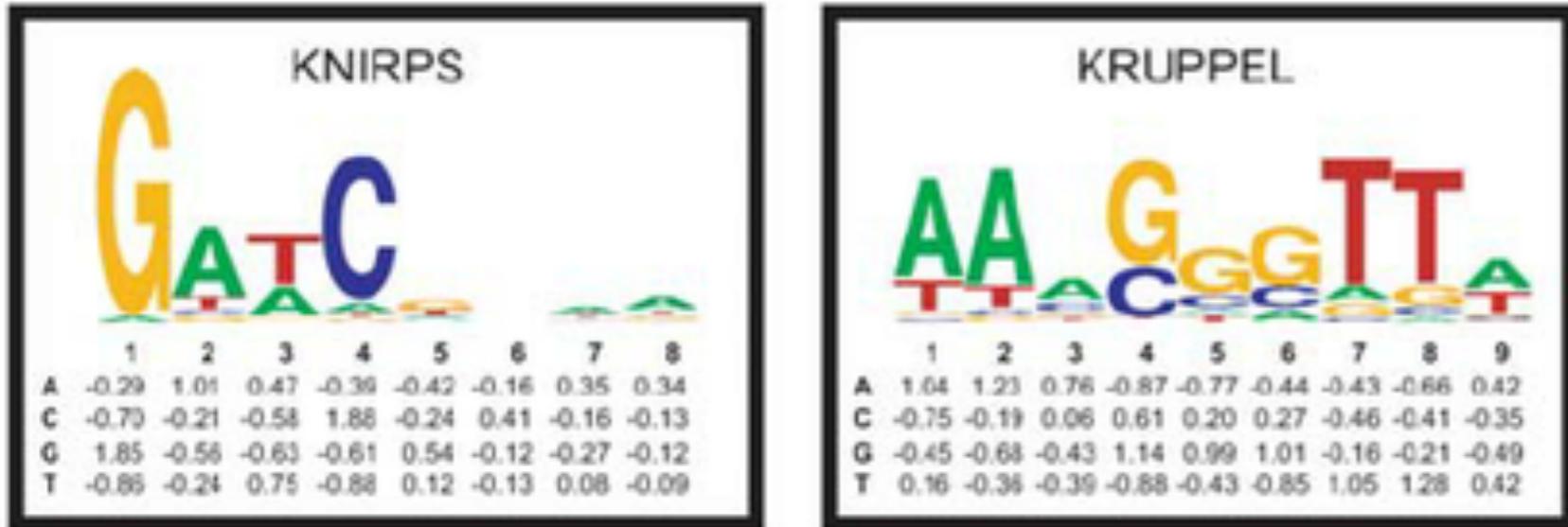
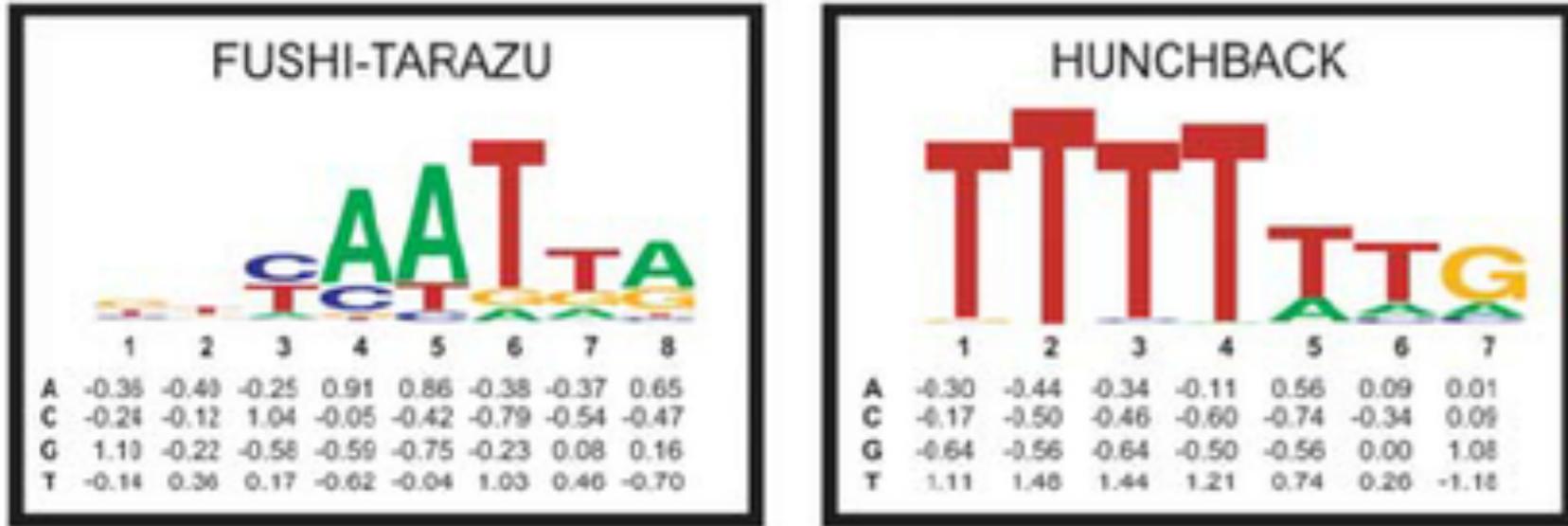
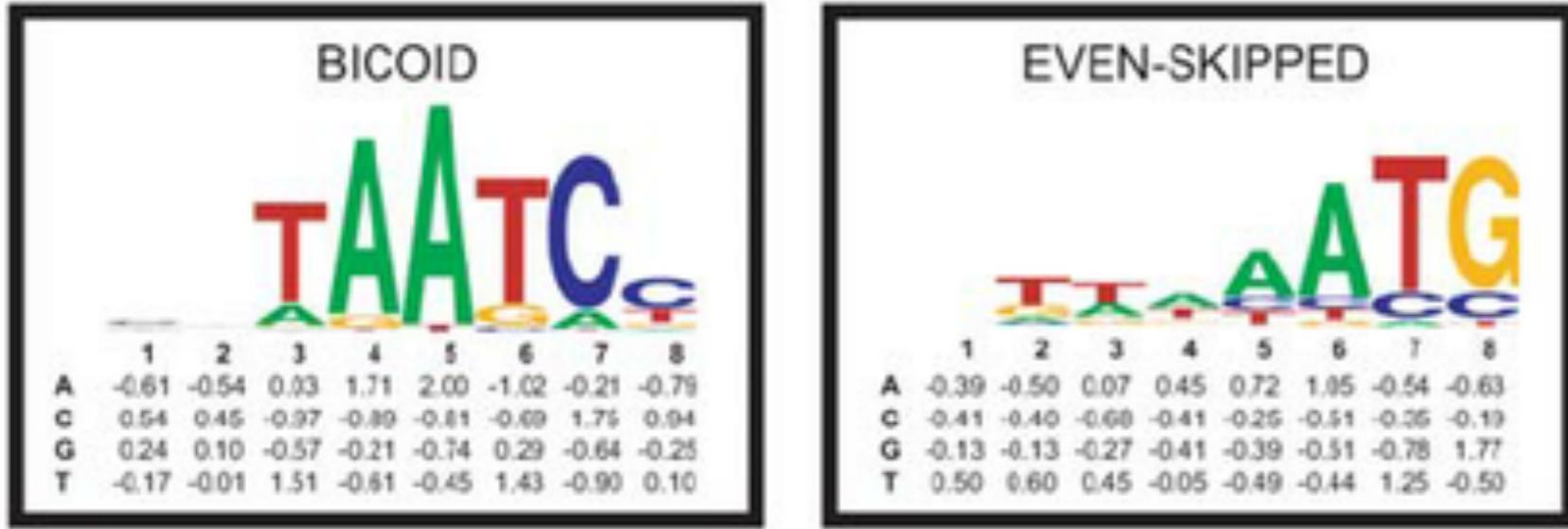
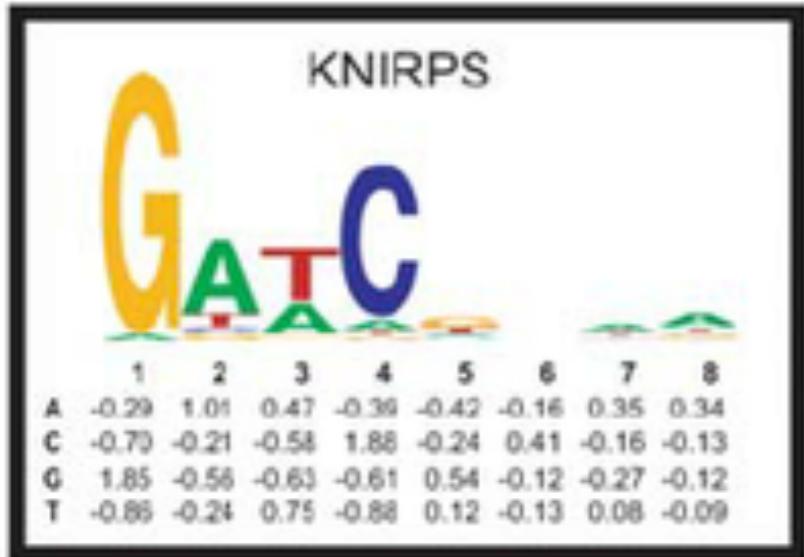
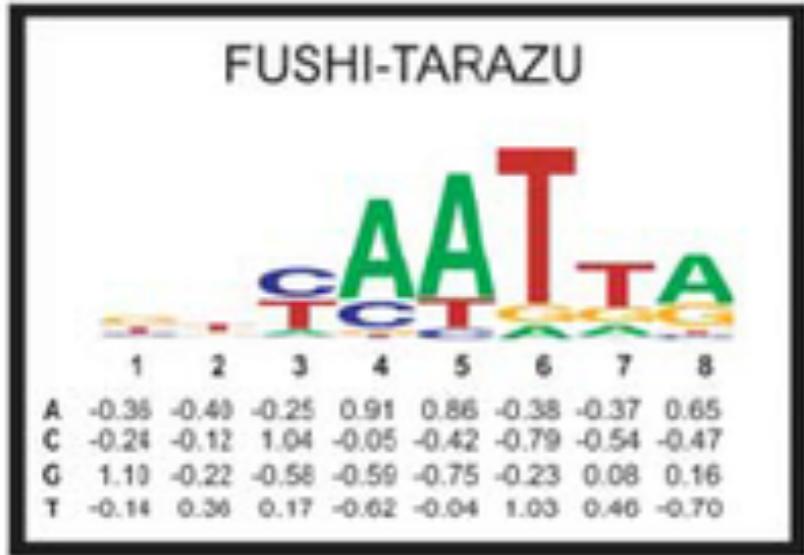
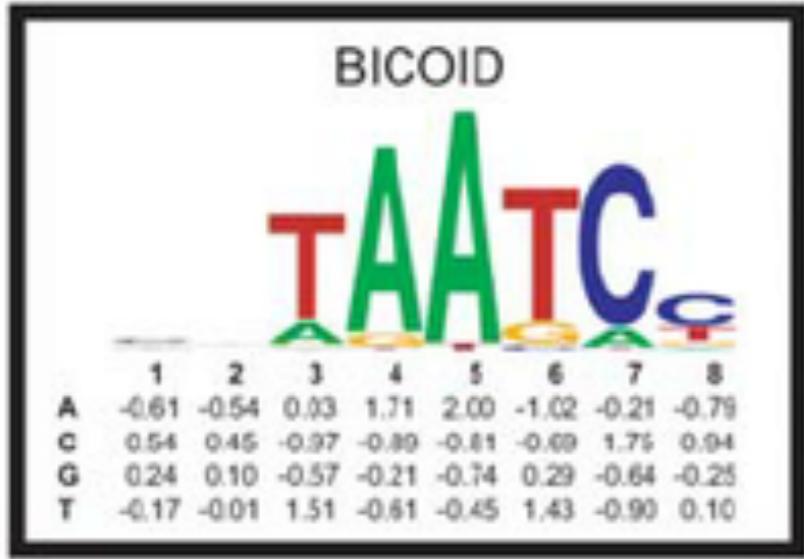


# **TFBS Meeting**

January 30, 2019

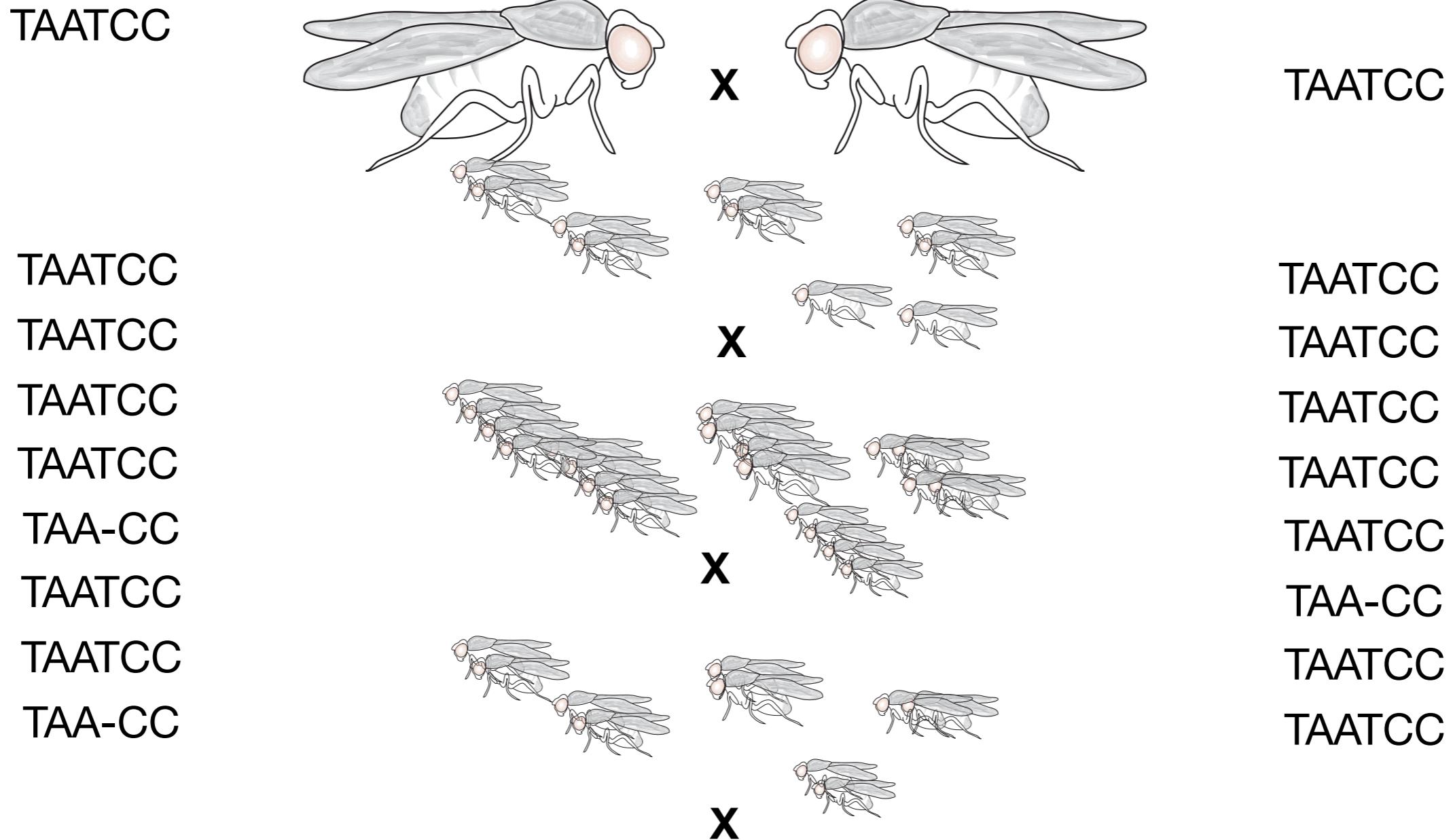
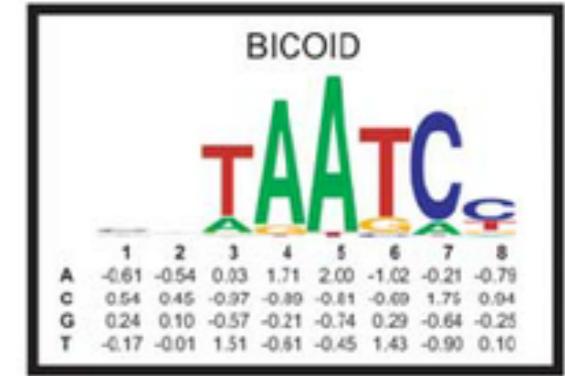
**Welcome Back!**



- \* **TFBS** are where important proteins (Transcription Factors) bind to DNA to affect gene regulation
- \* **TFBS** modulate in evolutionary time due to mutations.
- \* **TFBS loss:** They can modulate so much that they cannot bind anymore
- \* **TFBS gain:** DNA regions can become TFBS when mutations in a sequence change the sequence into a TFBS

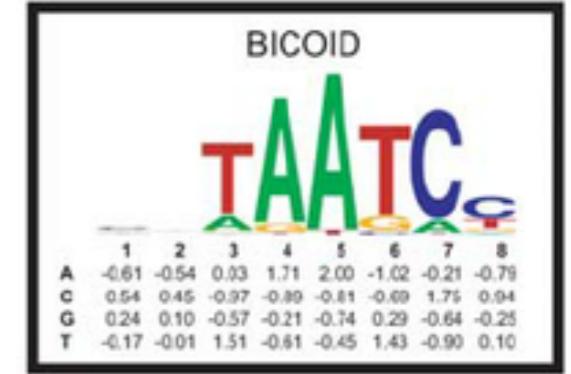
# Are there constraints or a pattern for on how a TFBS is lost or gained?

The perfect experiment:

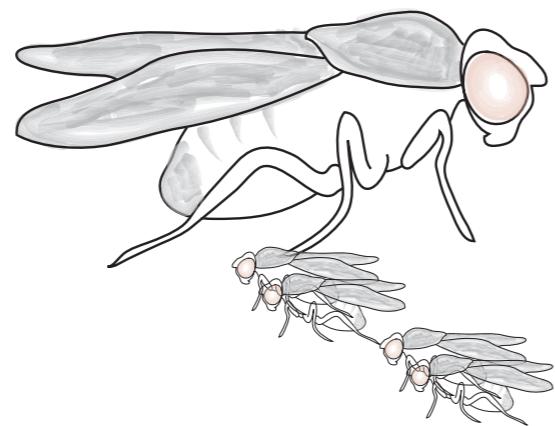


**Are there constraints or a pattern for  
on how a TFBS is lost or gained?**

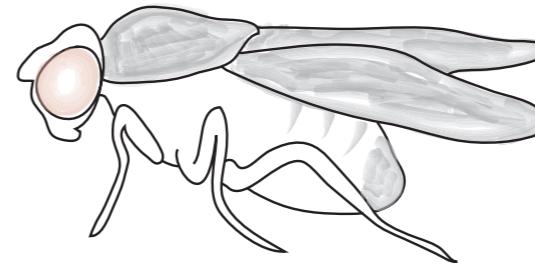
**The perfect experiment:**



TAATCC



X



TAATCC

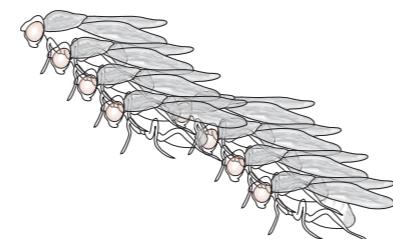
TAATCC

X



TAATCC

TAATCC



X



TAATCC

TAATCC



X



TAATCC

TAA-CC



X



TAATCC

TAATCC



X



TAA-CC

TAA-CC



X



TAATCC

TAATCC



X



TAATCC

**For millions of years**

## TFBS Loss?



**100 years**

TAATCC	TAATCC	TAATCC	TAATCC
TAATCC	TAATCC	TAATCC	TAATCC
TAATCC	TAATCC	TAATCC	TAATCC
TAATCC	TAATCC	TAATCC	TAATCC
TAATCC	TAATCC	TAATCC	TAATCC
TAATCC	TAATCC	TAATCC	TAATCC
TAATCC	TAATCC	TAATCC	TAATCC
TAATCC	TAATCC	TAATCC	TAATCC
TAATCC	TAATCC	TAATCC	TAATCC
TAATCC	TAATCC	TAATCC	TAATCC
TAAACC	TAAACC	TAAACC	TAAACC
TAAACC	TAAACC	TAAACC	TAAACC
TAAACC	TAAACC	TAAACC	TAAACC
TAAACC	TAAACC	TAAACC	TAAACC
TAAACC	TAAACC	TAAACC	TAAACC
TAAACG	TAAACG	TAAACG	AAAACC
TAAACG	TAAACG	TAAACG	AAAACC
TAAACG	TAAACG	TAAACG	AAAACC
GAAACG	GAAACG	GAAACG	AAAACC

**300 years**

**TAATCC**

100%

**TAAACC**

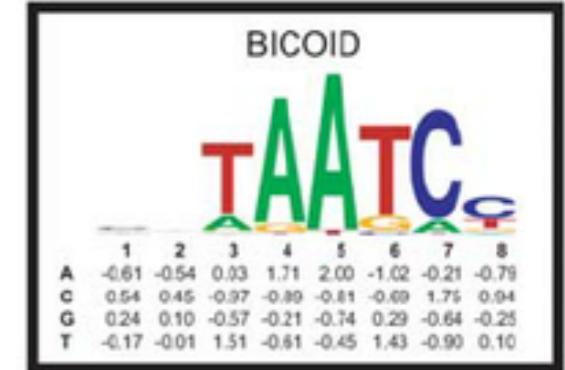
25%

75%

**AAAAACG** **TAAACG**

**800 years**

Step 1: T to A (pos 3)  
Step 2a: C to G (pos 6)  
Step 2b: T to A (pos 1)



Then we would have the “rules” for how Bicoid is lost in a genome!

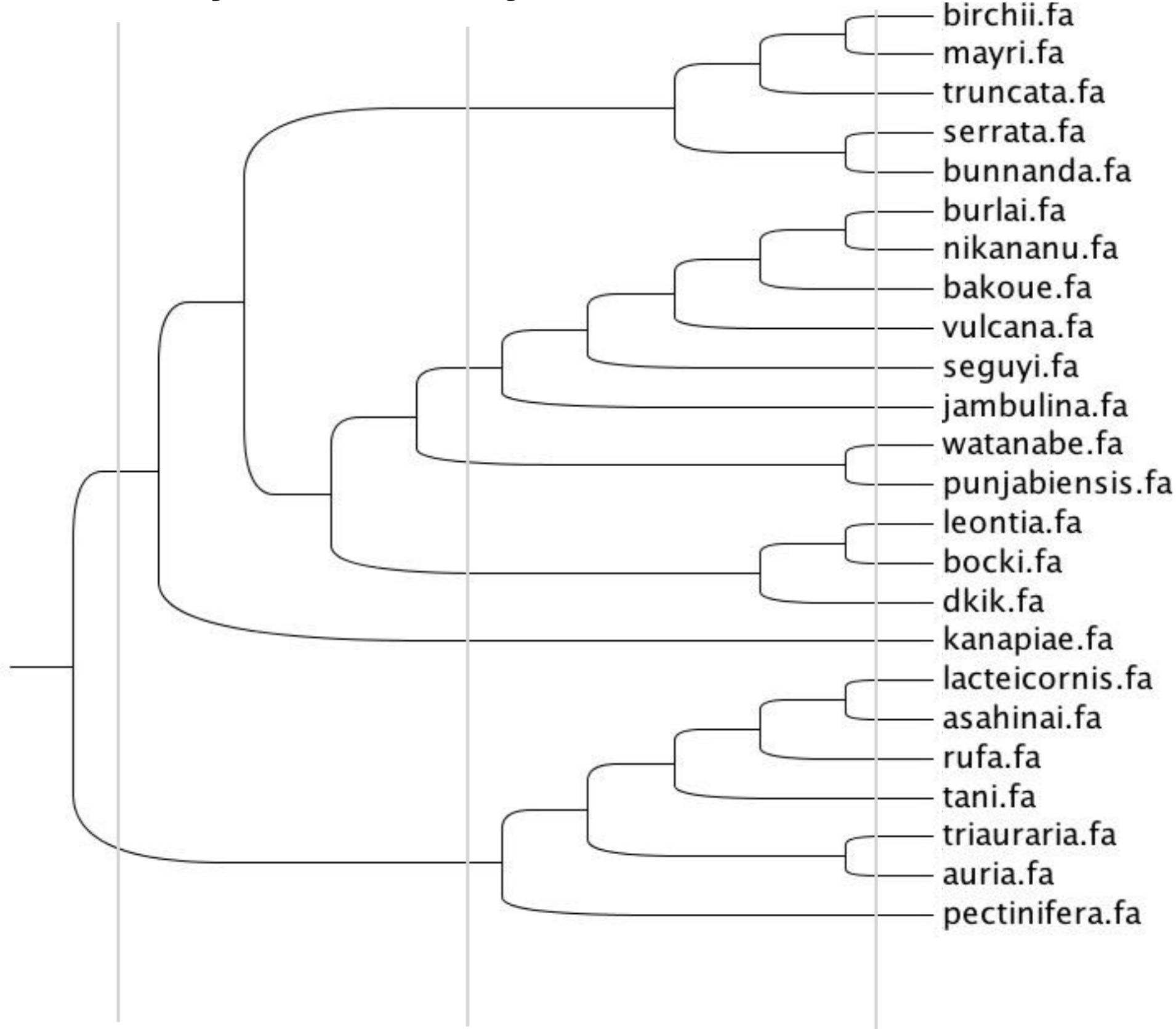
**But we can't do that.**

**41.3 million years**

**25 my**

**5 my**

**Now**



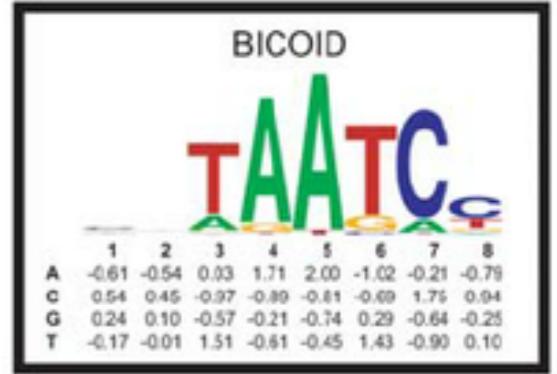
**All the Genomes!!**

## What do we need?

- ✓ The perfect evolutionary distance to sample
- ✓ All of the genomes
- ✓ Many non-coding regions that are known to regulate gene expression
- ✓ Automated way to find knownTFBS in a region
- ✓ Across evolutionary time (in alignment)
- ✓ Have a way to grab all the TFBS across all the species

Establish orthology

✓ We have this



**Species 1** GCTTT--TAATCCCTAGTGT--TT

**Species 2** GCCCCCTAA-TCCCTAGTGT--GT-

**Species 3** GCC--TAA--TCCCTAGTGT--GT

**Species 4** TAGGG----TAATCC---ATG--CC

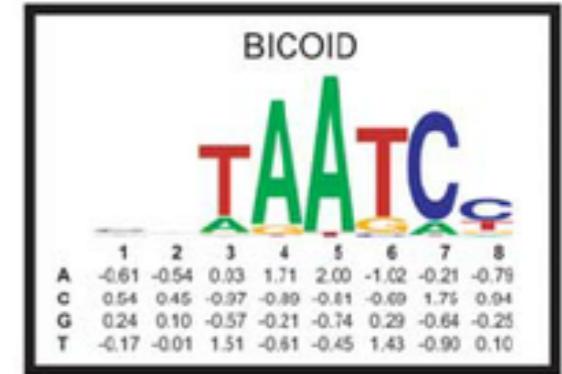
**Species 5** TAGGG----TAATCC---ATG--CT

**Species 6** TAGTCATTA-----ATCC-----

**Species 7** -----A-----

✓ We basically have this

A way to grab nearby Bicoid sites and the regions around the sites



**Species 1** GCTTT--**TAATCCCTAGTGT**--TT

**Species 2** GCCCCT**AA-TCCCTAGTGT**--GT-

**Species 3** GCC--**TAA**--TCCCTAGTGT--GT

**Species 4** TAGGG---**TAACCC**---ATG--CC

**Species 5** TAGGG---**AAACCC**---ATG--CT

**Species 6** TAGTCATT**A**-----ATCC-----

**Species 7** -----**A**-----



Bicoid with a high score

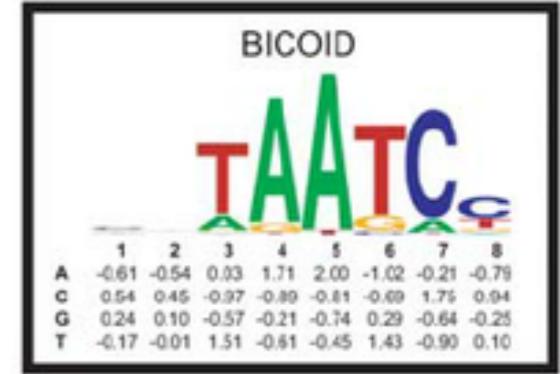


Bicoid with a low score



Not bicoid, but why

But what we want is more complex. We want to make sure that we are comparing the same Bicoid site between species (**orthology**)



**Species 1** GCTTT--**TAATCCCTAGTGT**--TT

**Species 2** GCCCCT**AA-TCCCTAGTGT**--GT-

**Species 3** GCC--**TAA**--TCCCTAGTGT--GT

**Species 4** TAGGG---**TAACCC**---ATG--CC

**Species 5** TAGGG---**AAACCC**---ATG--CT

**Species 6** TAGTCATT**A**-----ATCC-----

**Species 7** -----**A**-----



**Bicoid with a high score**

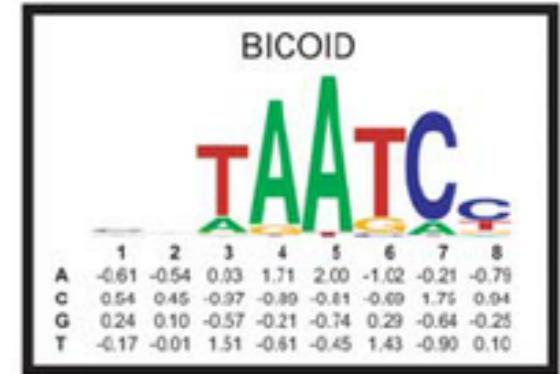


**Bicoid with a low score**



**Not bicoid, but why**

For example: Species 7, what is going on there?



**Species 1** GCTTT--TAATCCCTAGTGT--TT

**Species 2** GCCCCTAA-TCCCTAGTGT--GT-

**Species 3** GCC--TAA--TCCCTAGTGT--GT

**Species 4** TAGGG---TAACCC---ATG--CC

**Species 5** TAGGG---AAACCC---ATG--CT

**Species 6** TAGTCATTA-----ATCC-----

**Species 7** T... -----A----- ...ATCC



**Bicoid with a high score**

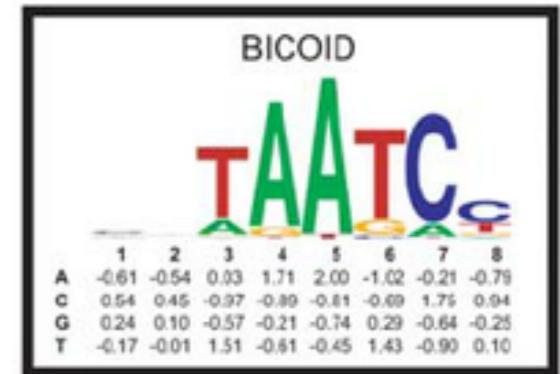


**Bicoid with a low score**



**Not bicoid, but why**

For example: Species 7, what is going on there?



**Species 1** GCTTT--**TAATCCCTAGTGT**--TT

**Species 2** GCCCCT**AA-TCCCTAGTGT**--GT-

**Species 3** GCC--**TAA**--TCCCTAGTGT--GT

**Species 4** TAGGG---**TAACCC**---ATG--CC

**Species 5** TAGGG---**AAACCC**---ATG--CT

**Species 6** TAGTCATT**A**-----ATCC-----

**Species 7** TTT -----A-----Seq ends ?



**Bicoid with a high score**

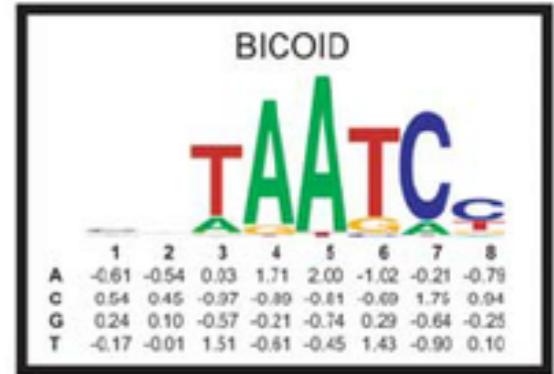


**Bicoid with a low score**



**Not bicoid, but why**

All we need to do is use one high score Bicoid position as an example



**Species 1** GCTTT--**TAATCCCTAGTGT**--TT

**Species 2** GCCCCTAA-TCCCTAGTGT--GT-

**Species 3** GCC--TAA--TCCCTAGTGT--GT

**Species 4** TAGGG---TAACCC---ATG--CC

**Species 5** TAGGG---AAACCC---ATG--CT

**Species 6** TAGTCATTA-----ATCC-----

**Species 7** -----A-----



**Bicoid with a high score**

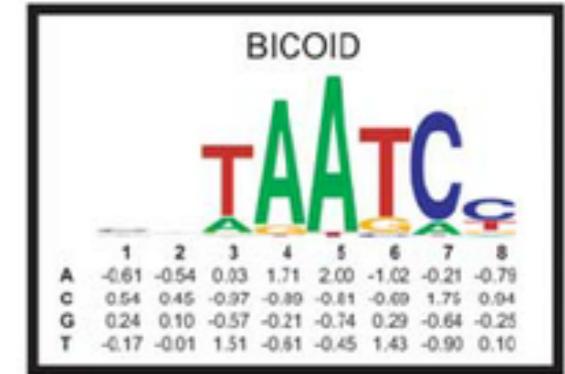


**Bicoid with a low score**



**Not bicoid, but why**

Grab everything at that exact position



**Species 1** GCTTT--TAATCCCTAGTGT--TT

**Species 2** GCCCCTAA-TCCCTAGTGT--GT-

**Species 3** GCC--TAA--TCCCTAGTGT--GT

**Species 4** TAGGG---TAACCC---ATG--CC

**Species 5** TAGGG---AAACCC---ATG--CT

**Species 6** TAGTCATTA-----ATCC-----

**Species 7** -----A-----



**Bicoid with a high score**

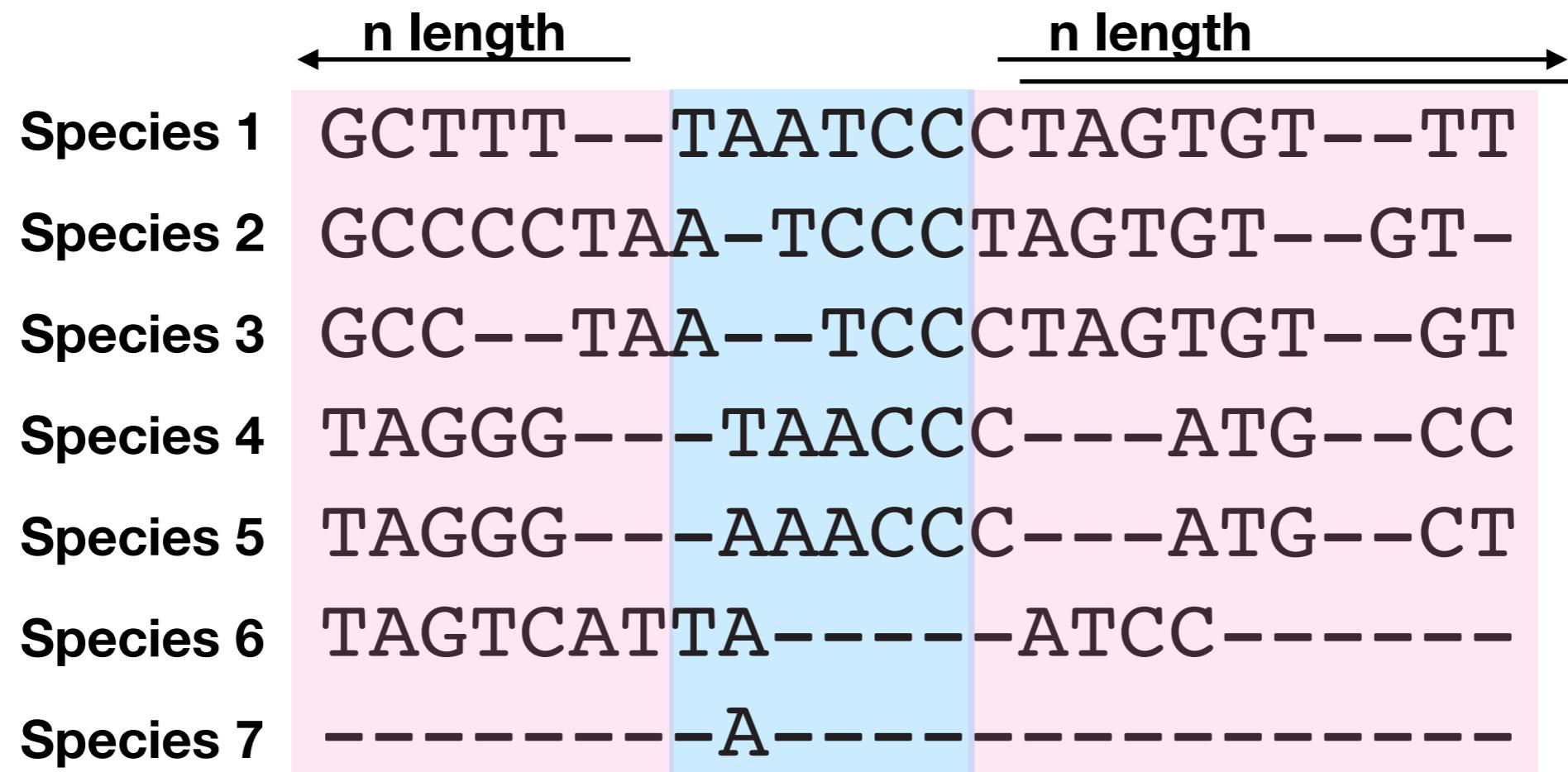
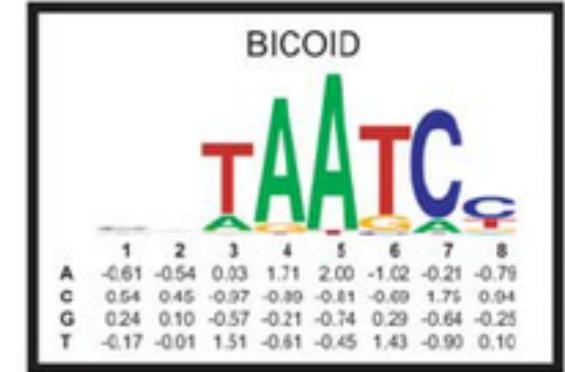


**Bicoid with a low score**



**Not bicoid, but why**

Grab up and down stream n-length (being an argument we can change)



Bicoid with a high score

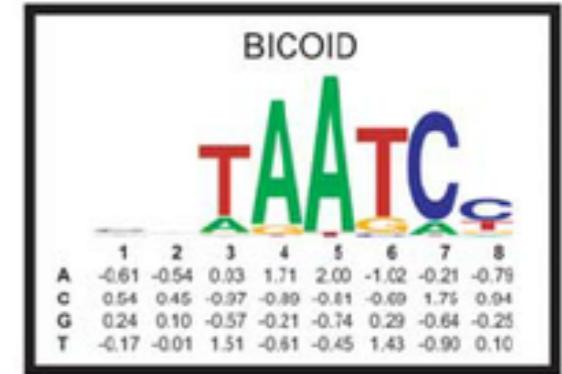


Bicoid with a low score



Not Bicoid, but why?

Establish Orthology.



**Species 1** GCTTT--TAATCCCTAGTGT--TT

**Species 2** GCCCCTAATCCCTAGTGT--GT-

**Species 3** GCC--TAATCCCTAGTGT--GT

**Species 4** TAGGG--TAACCG---ATG--CC

**Species 5** TAGGG--AAACCC---ATG--CT

**Species 6** GTCAT--TAATCC-----

**Species 7** -----A-----



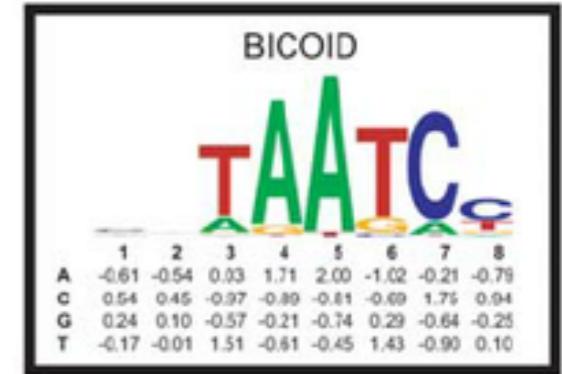
**Bicoid with a high score**



**Bicoid with a low score**



**Not bicoid, but why**



Identify what sort of mutations occur. Calculate if there is a pattern. Publish a really cool study!

<b>Species 1</b>	GCTTT-- <b>TAATCCCTAGTGT</b> --TT
<b>Species 2</b>	GCCCCT <b>TAATCCCTAGTGT</b> --GT-
<b>Species 3</b>	GCC-- <b>TAATCGCTAGTGT</b> --GT
<b>Species 4</b>	TAGGG-- <b>TAACCG</b> ---ATG--CC
<b>Species 5</b>	TAGGG-- <b>AAACCC</b> ---ATG--CT
<b>Species 6</b>	GTCAT-- <b>TAATCC</b> -----
<b>Species 7</b>	----- <b>A</b> -----



**Bicoid with a high score**



**Bicoid with a low score**



**Not bicoid, but why**