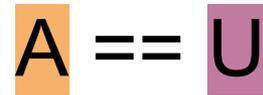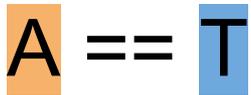# Lineage-dependent properties of 16S ribosomal RNA nucleotide composition
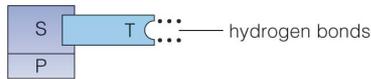
Ashwin Murali
Third Annual MIT PRIMES Conference,
May 19, 2013

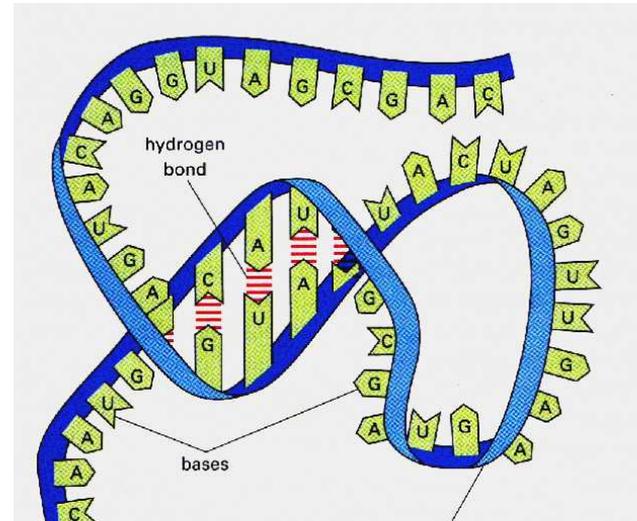# Nucleotides engage in base pairing in both DNA and RNA

A == T   G == C | A == U   G == C

G = guanine
C = cytosine
A = adenine
T = thymine

hydrogen bonds

S = deoxyribose sugar
P = phosphate group

© 2004 Wadsworth - Thomson Learning

hydrogen bond

bases

Secondary Strucuture of Helix 11 Loop

Helix 11

Helix 11 Loop

16S RNA

RNA component of the ribosome consists of single stranded RNA which is folded into a three-dimensional structure

Secondary Strucuture of Helix 11 Loop

Helix 11

Helix 11 Loop

16S RNA

25-30 nm
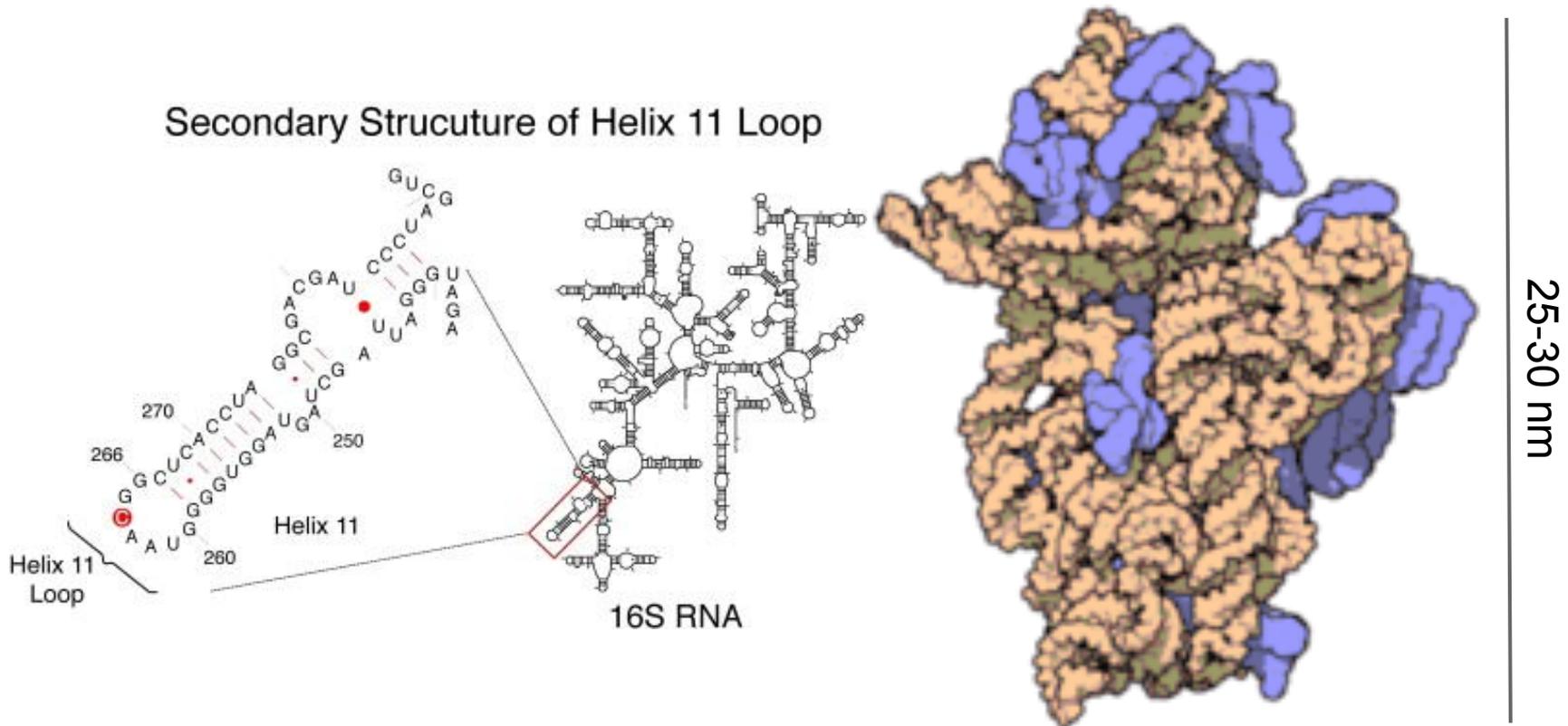
What is the ribosome?

- Responsible for protein synthesis

- Highly conserved through evolution

- Multi-component molecular assembly made of both RNA and protein

# Here, we study the nucleotide composition of the 16S subunit

- In prokaryotes (Bacteria and Archaea) the 16S length is roughly 1200 to 1400 bases

Why study 16S nucleotide statistics?

- 16S sequence is "phylogenetic fingerprint" for bacteria and archaea

- The Ribosomal Database Project (RDP) has a compilation of over 1.2 million bacteria/archaea 16S sequences available for study

- Study of nucleotide ratios offer structural insights
  - eg. Chargaff's Rules for A = T and G = C in DNA

- In our study, we were seeking to find any such interesting nucleotide distribution patterns

What did we expect?

- In general, we expect guanine and cytosine to be positively correlated as well as adenine and uracil
  - G - C and A - U pairs form stems in the secondary rRNA structure

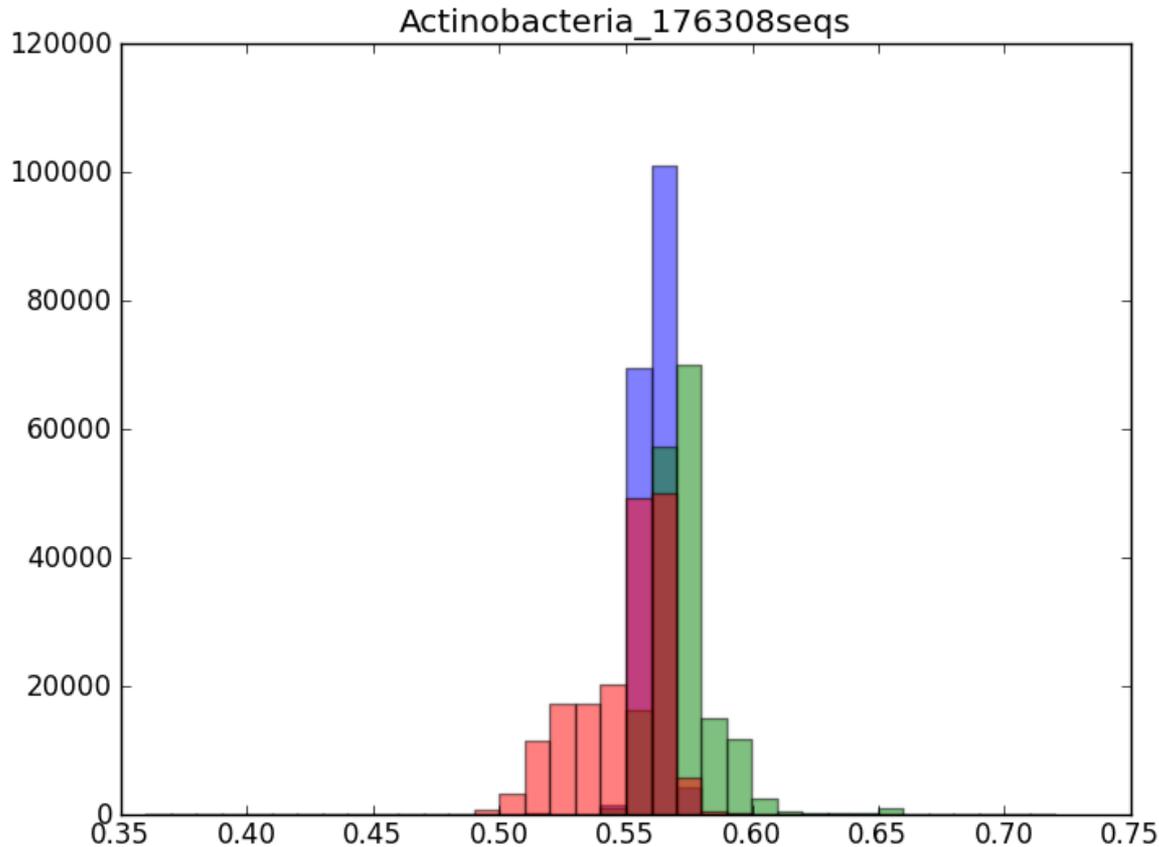Secondary Structure of Helix 11 Loop

Helix 11

Helix 11 Loop

16S RNA
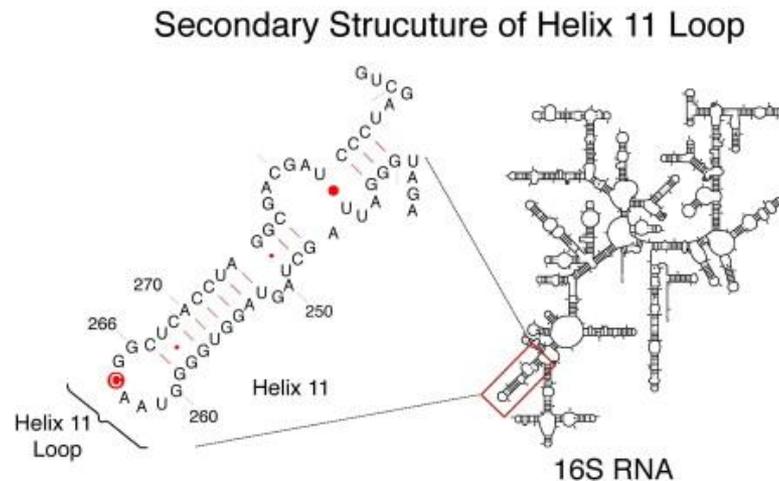
Proteobacteria
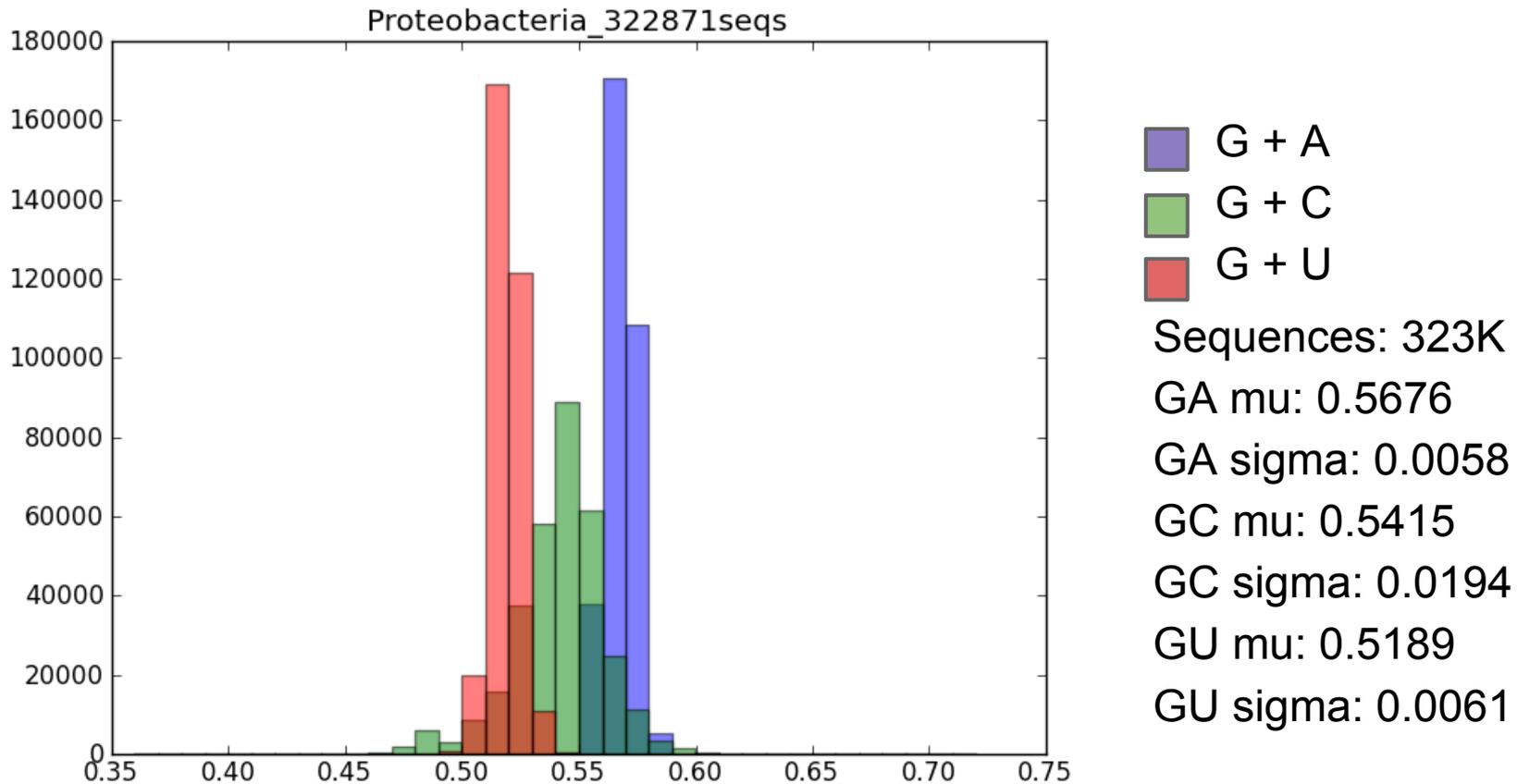
We find that purine and GU content is significantly more conserved than GC content in almost all major phyla.

Actinobacteria

Actinobacteria_176308seqs

Legend:
- G + A
- G + C
- G + U

Sequences: 176K
GA mu: 0.5604
GA sigma: 0.0041
GC mu: 0.5728
GC sigma: 0.013
GU mu: 0.5108
GU sigma: 0.019

We find Actinobacteria to be the only major phyla, that does not follow this overall trend; in Actinobacteria, GU content varies similarly to GC content.

Histograms indicate different nucleotide distributions in Actinobacteria

- Purine (Guanine + Adenine) content and GU content (Guanine + Uracil) appears conserved across most Bacteria and Archaea

- We find Actinobacteria to be an exception to this overall trend, with GU content variance greater than GC content.

# Correlation - Proteobacteria Clades
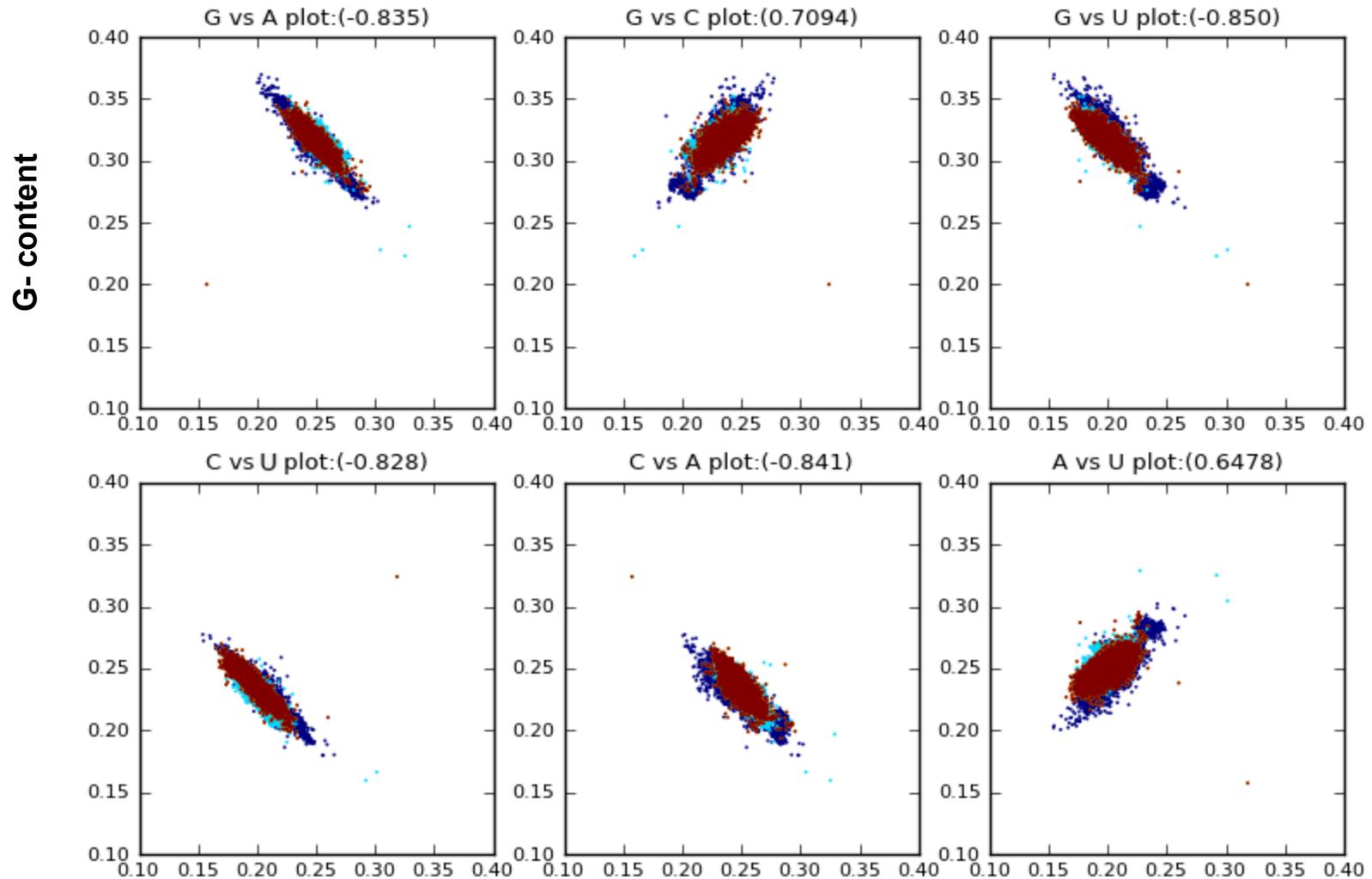
Correlation - Actinobacteria Clades
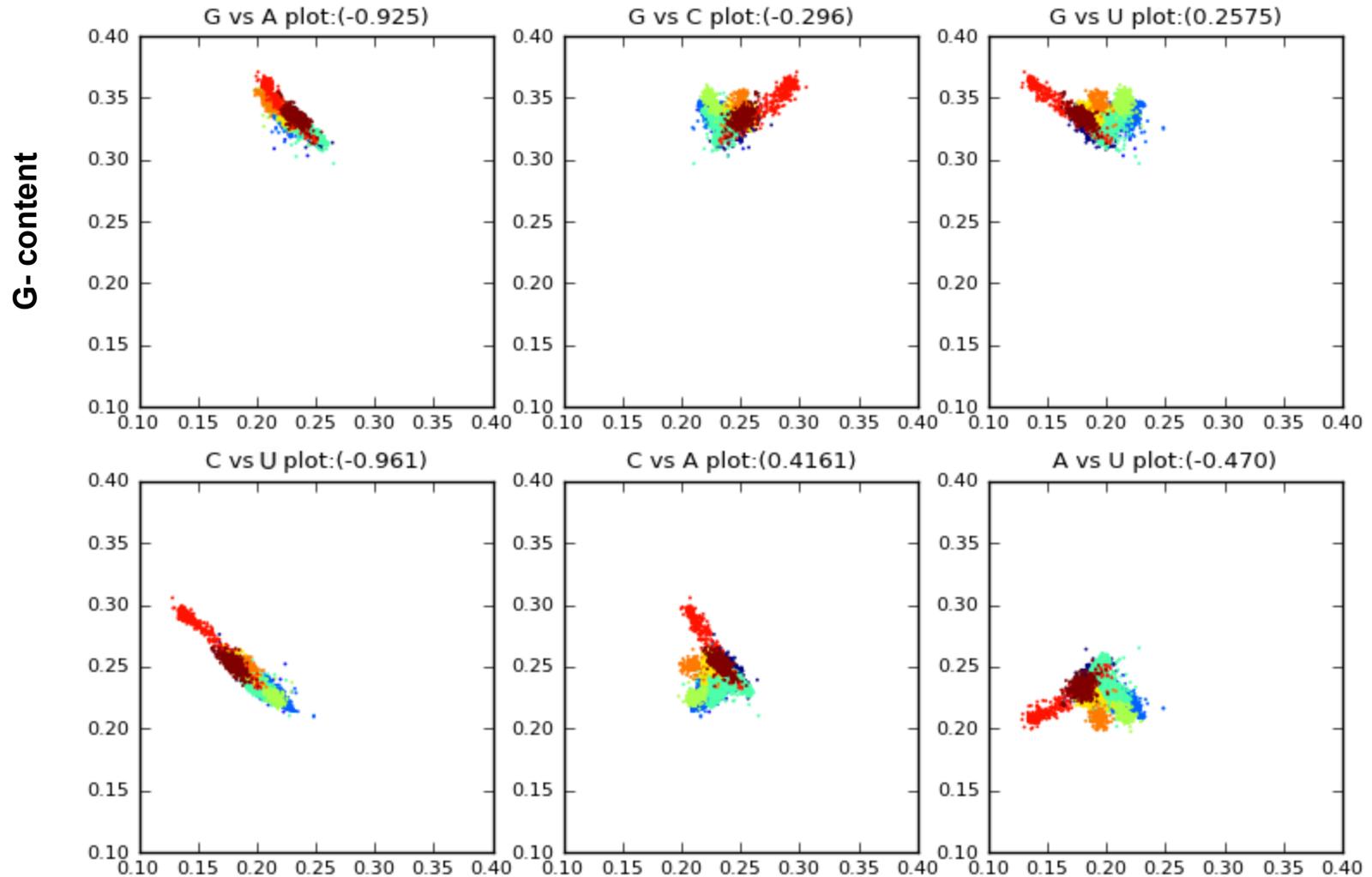
Scatter plots demonstrate that Actinobacteria follow different nucleotide content trends

- All other major bacterial and archaeal phyla show scatter plots similar to the Proteobacteria scatterplot, where we observe what we expected.

- Actinobacteria follow a completely different nucleotide correlation trend than other bacterial species

  - In Actinobacteria, guanine and uracil are the least correlated pair of nucleotides.

  - Actinobacteria are only phyla where G is positively correlated with U

Correlation - Actinobacteria Clades Magnified
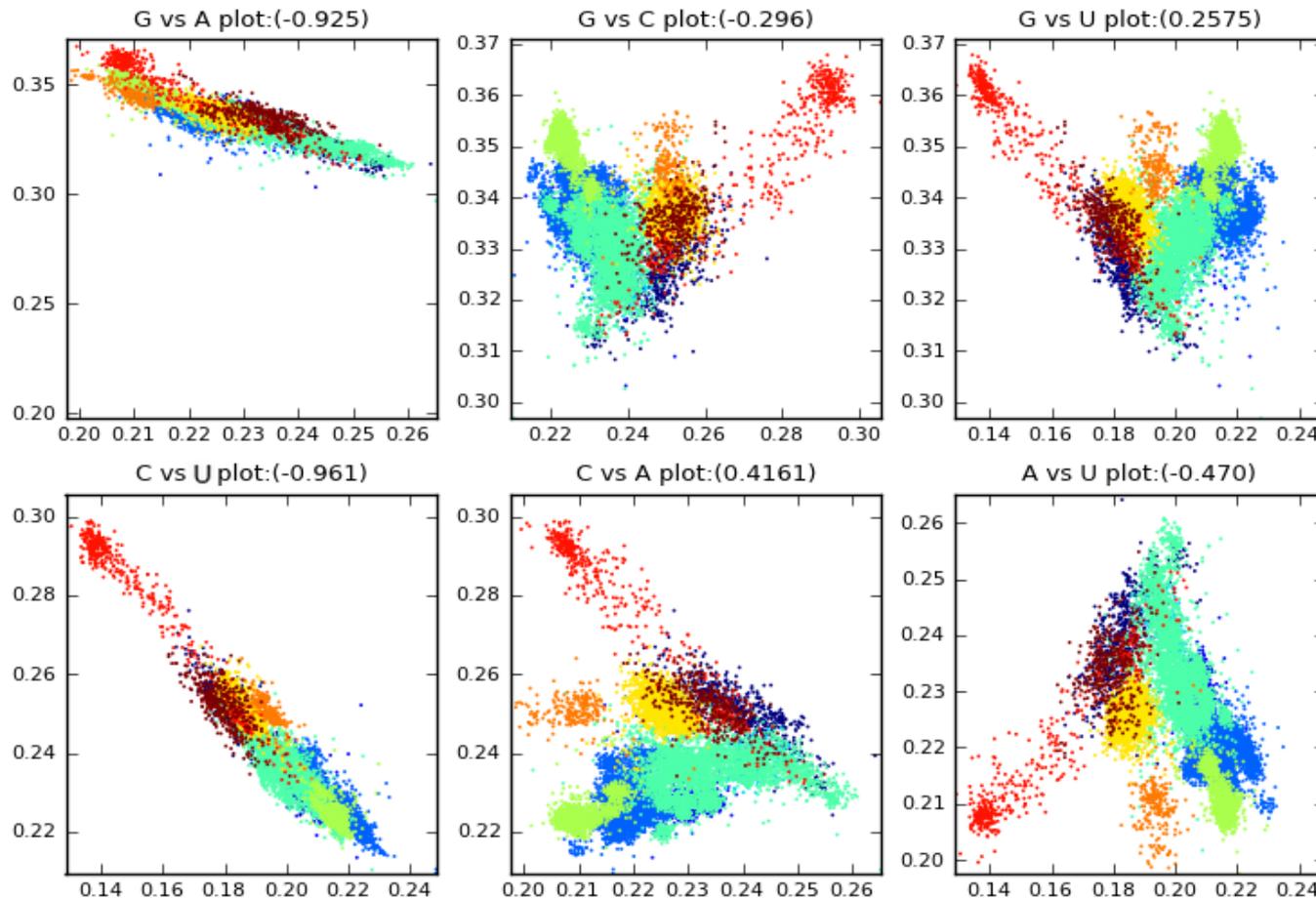
Actinobacteria Families:

- Acidimicrobidae
- Actinobacteridae
- Bifidobacteroales
- Coriobacteriaceae
- Nitriliruptoridae
- Rubrobacterineae
- Unclassified

Magnified scatter plots demonstrate that altered nucleotide content trend is observed in the majority of Actinobacteria clades

- We note that this lack of GU correlation is not only present in the larger Actinobacteria phyla but also manifests itself in many of the smaller subfamilies

- The only two clades that do not demonstrate this altered trend are Rubrobacterineae and Unclassified Actinobacteria

Can the known characteristics of Actinobacteria explain our results?

- Gram-positive

- Majority of phyla is aerobic

- They are usually decomposers that play a major role in almost all biological ecosystems.

- They produce naturally occuring antibiotics

Conclusions and Future work

- Known characteristics of actinobacteria do not explain trend

- Further investigation must be conducted to figure out the cause of this deviation from the overall trend of GC correlation.

  - hypothesis: additional GU correlation and decreased GC correlation may be due to increased G-U bonding at the expense of GC bonds due to RNA wobble.

- In addition, we would like to investigate if there are clades within other phyla with similar behavior to Actinobacteria.

- Comparison of variation with RNA stem+loop locations

Thanks to our mentors, Geoffrey Fudenberg, Maxim Imakaev, Professor Leonid Mirny and MIT PRIMES