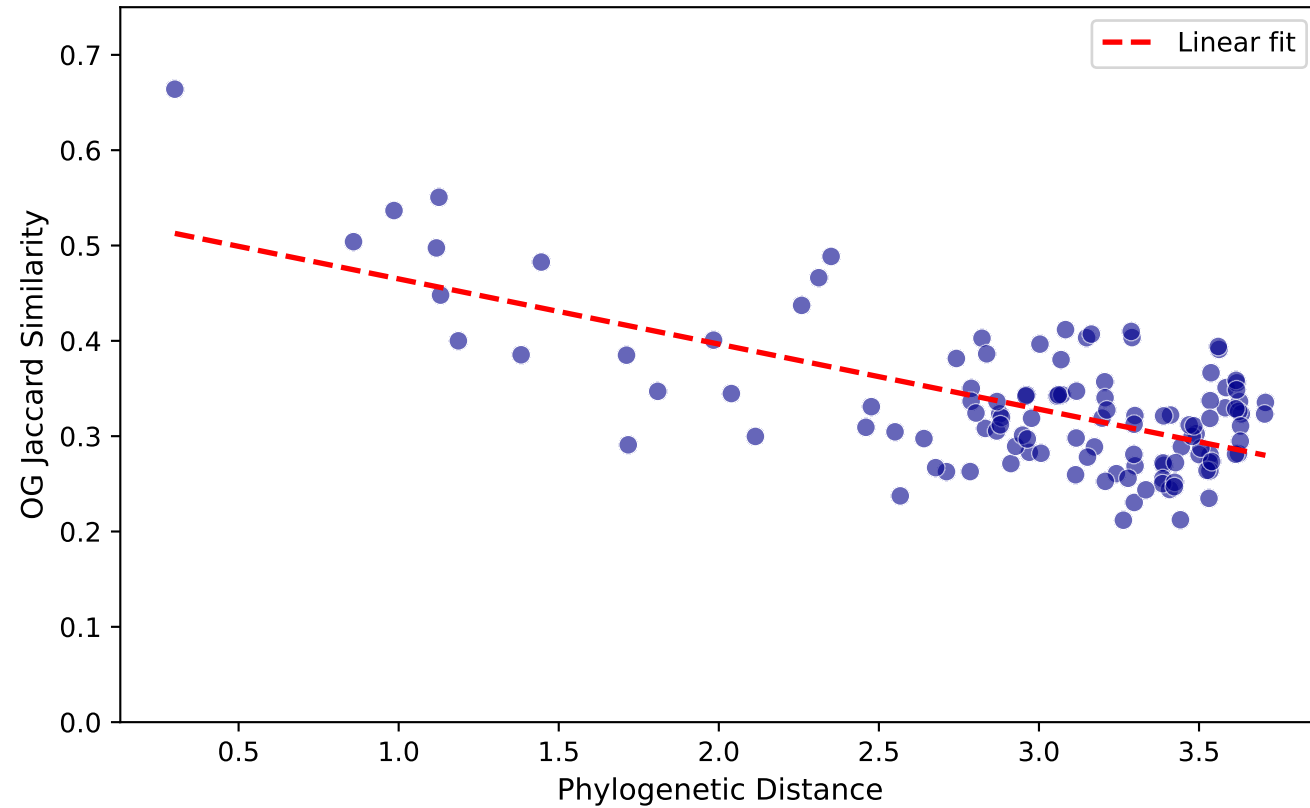
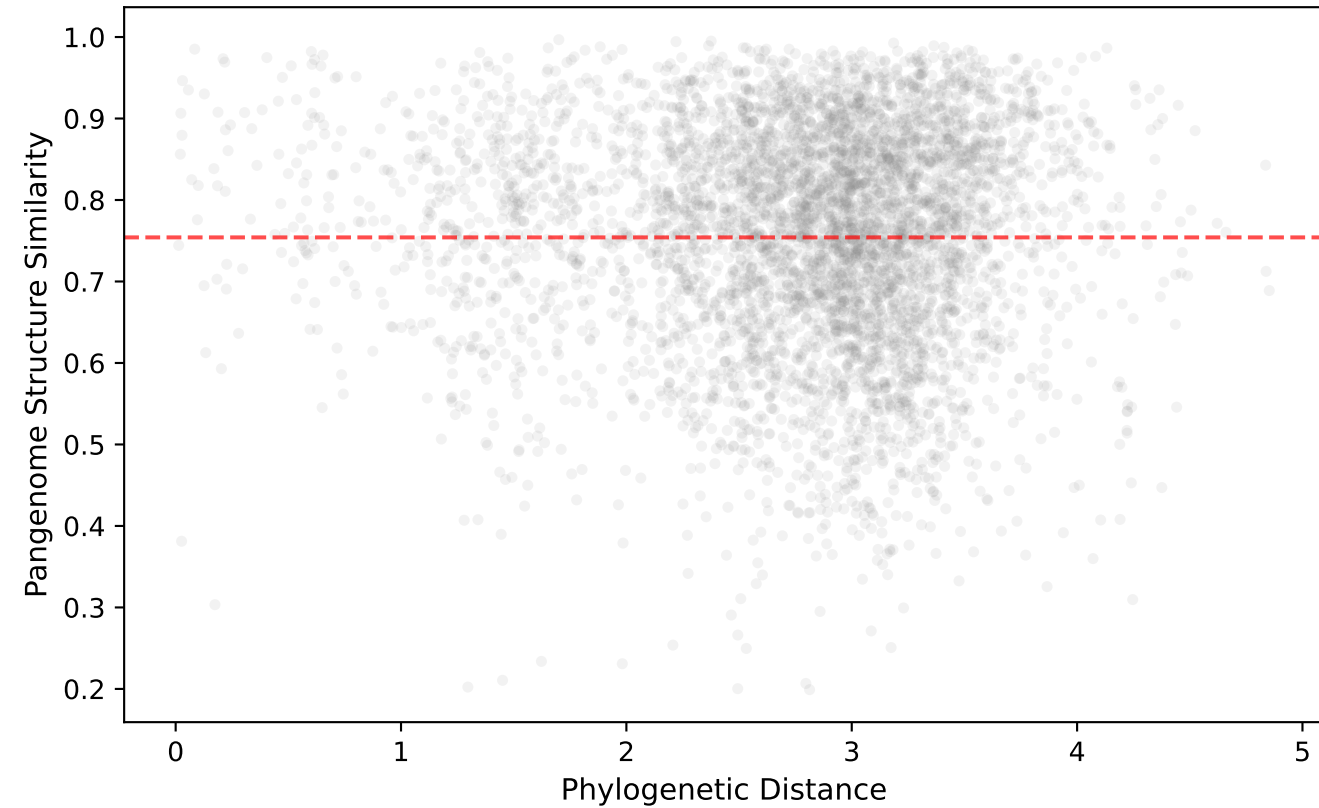


# Phylogeny vs Gene Content: Two Metrics Tell Different Stories

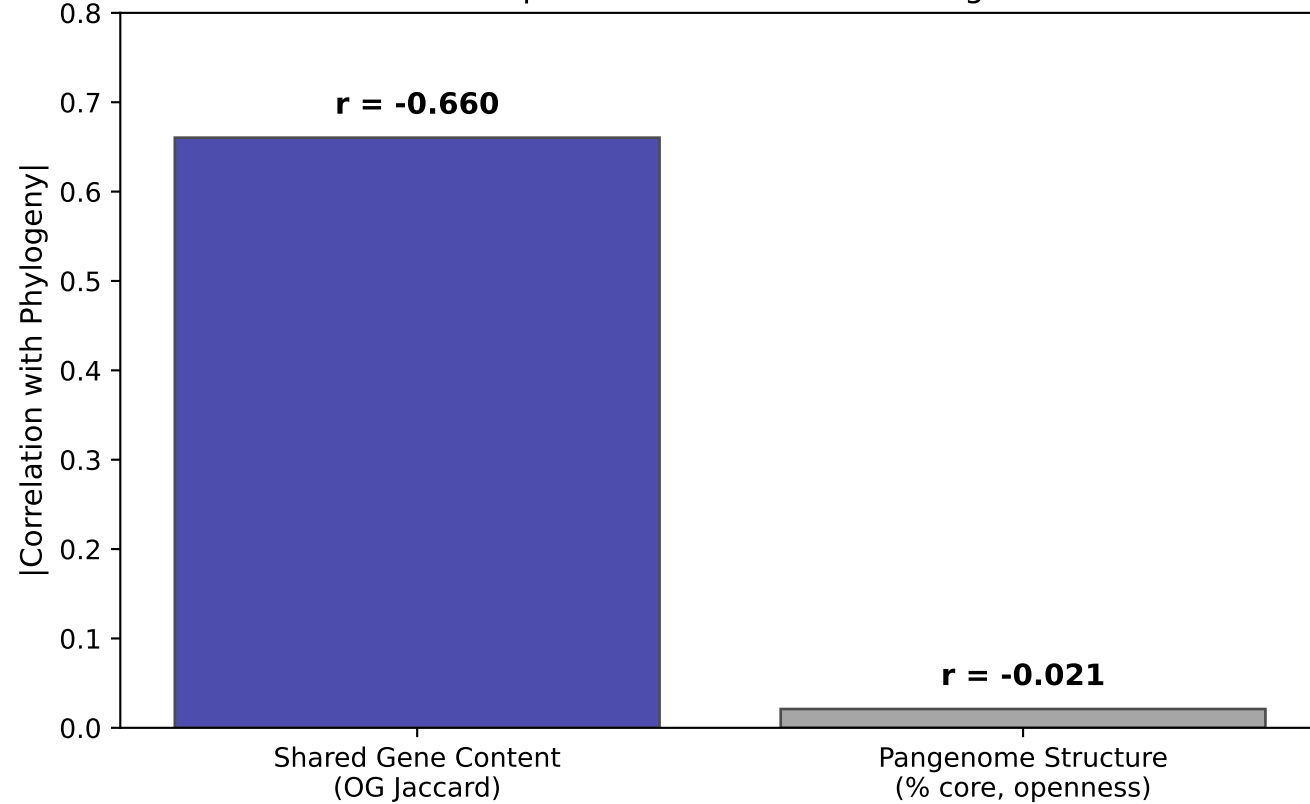
A. Shared Gene Content vs Phylogeny  
( $r = -0.660$ ,  $n = 120$  pairs)



B. Pangenome Structure vs Phylogeny  
( $r = -0.021$ ,  $n = 124,750$  pairs)



C. Comparison of Correlation Strengths



## KEY FINDINGS: Gene Content vs Phylogeny

### ACTUAL GENE CONTENT (OG Jaccard Similarity)

- Strong correlation with phylogeny:  $r = -0.66$
- Closely related species share more genes
- Same-genus pairs: ~66% gene overlap
- Cross-phylum pairs: ~30% gene overlap
- Gene content IS predicted by phylogeny

### PANGENOME STRUCTURE (% core, openness)

- No correlation with phylogeny:  $r \approx 0$
- Closely related species DON'T have similar structures
- Same-genus pairs show high variance
- Structure reflects ecological strategy, not ancestry

## INTERPRETATION

- WHAT genes you have = determined by ancestry
- HOW genes are organized = determined by ecology
- Two species in the same genus share many genes...
- ...but can have very different % core if they've adapted to different ecological niches
- Two distant species may converge on similar pangenome structures if they occupy similar niches