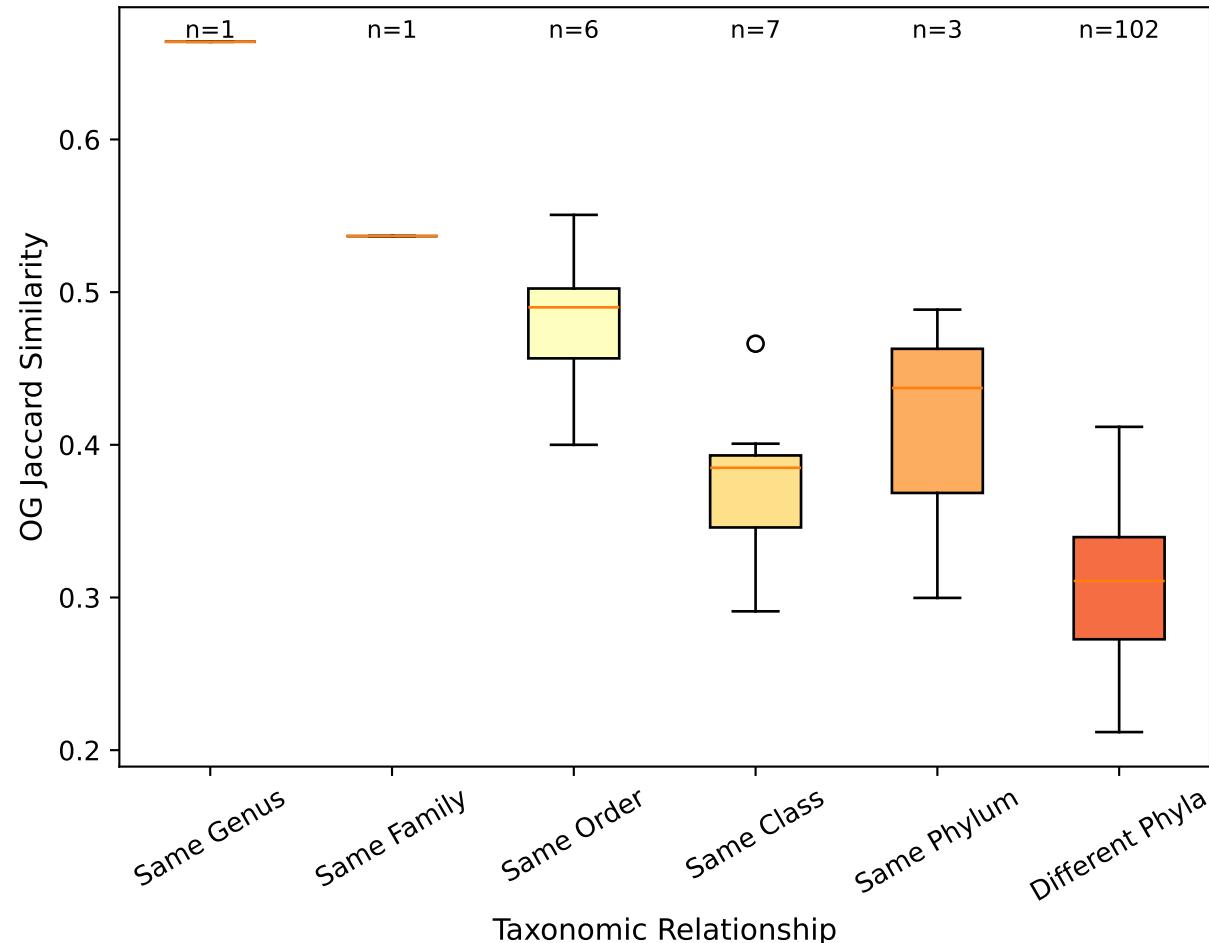


Gene Content Similarity by Taxonomic Relationship

A. Gene Content Similarity Decreases with Taxonomic Distance



GENE CONTENT SIMILARITY SUMMARY

Relationship	Mean Jaccard	Std	N pairs	Mean Shared OGs
Same Genus	0.664	nan	1	832
Same Family	0.537	nan	1	768
Same Order	0.480	0.052	6	820
Same Class	0.374	0.055	7	735
Same Phylum	0.408	0.098	3	718
Different Phyla	0.310	0.047	102	551

INTERPRETATION:

- Same-genus species share ~66% of their core genes
- Same-phylum species share ~35-40% of core genes
- Cross-phylum species still share ~30% (universal genes)

The ~30% baseline represents:

- Core cellular machinery (ribosomes, transcription)
- Basic metabolic enzymes (glycolysis, TCA cycle)
- Essential housekeeping genes