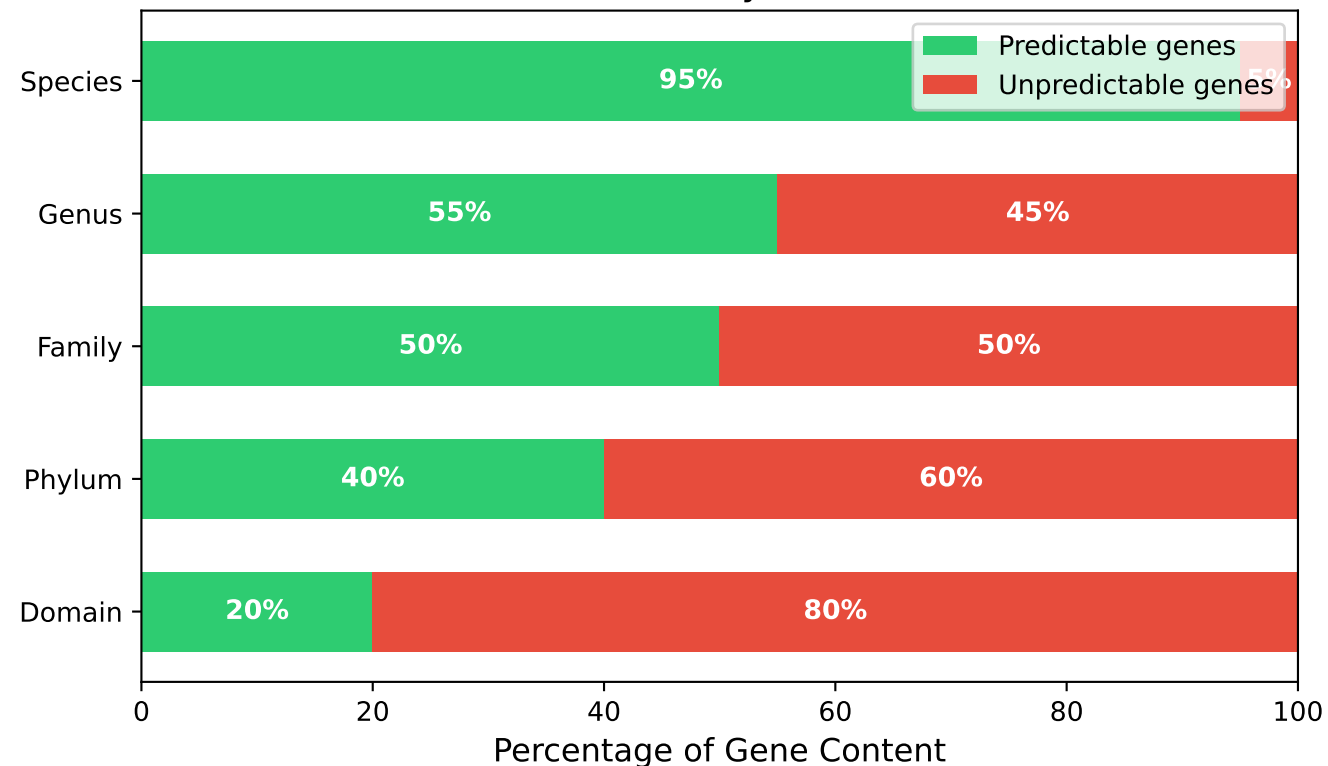


Functional Predictability from Taxonomic ID



What Taxonomy Actually Tells You About Function

Domain level:

- Universal functions only (ribosomes, transcription)
- Almost nothing about specific metabolism

Phylum level:

- Cell architecture (Gram +/-)
- Basic energy metabolism type
- Nothing about specific functions

Family level:

- General metabolic strategy
- Major pathway presence
- Nothing about substrates or niche

Genus level:

- Core metabolic pathways
- ~50% of gene functions
- Nothing about ecological strategy

Species level:

- Nearly complete functional prediction
- Core genome known (~95%)
- Only strain variation unknown

KEY FINDING:

Pangenome openness (% core) does NOT correlate with phylogeny ($r = -0.001$)

- Genome architecture reflects ECOLOGY
- Not evolutionary history
- Taxonomy \neq Ecological strategy