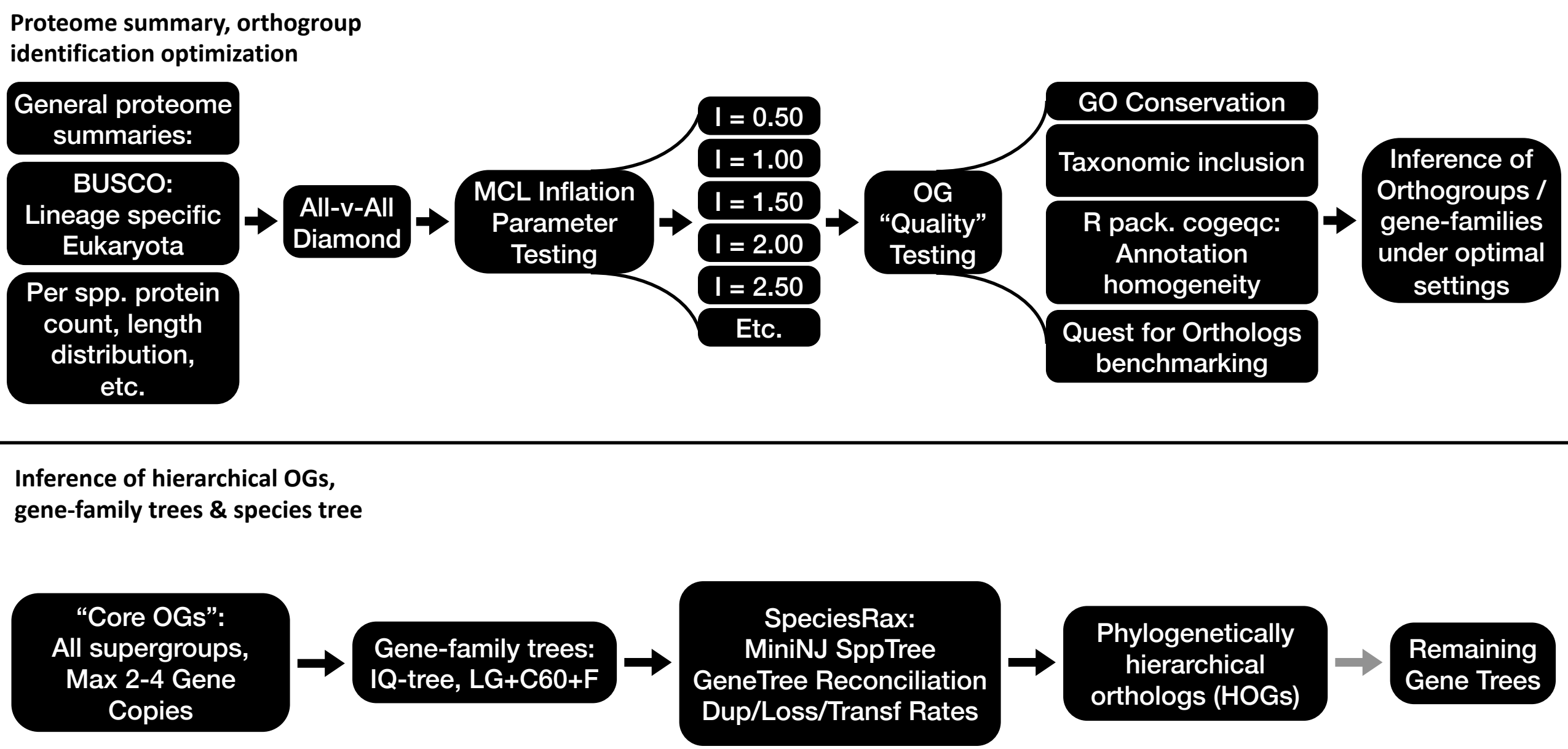
Proteome summary, orthogroup identification optimization
General proteome summaries:
BUSCO: Lineage specific Eukaryota
Per spp. protein count, length distribution, etc.
All-v-All Diamond
MCL Inflation Parameter Testing
I = 0.50
I = 1.00
I = 1.50
I = 2.00
I = 2.50
Etc.
OG "Quality" Testing
GO Conservation
Taxonomic inclusion
R pack. cogeqc: Annotation homogeneity
Quest for Orthologs benchmarking
Inference of Orthogroups / gene-families under optimal settings
Inference of hierarchical OGs, gene-family trees & species tree
"Core OGs": All supergroups, Max 2-4 Gene Copies
Gene-family trees: IQ-tree, LG+C60+F
SpeciesRax: MiniNJ SppTree GeneTree Reconciliation Dup/Loss/Transf Rates
Phylogenetically hierarchical orthologs (HOGs)
Remaining Gene Trees