# Filters

cond\_species\_not\_symbiont = ~df\_taxonomy['species'].str.contains('\s(?:endosymbiont|ectosymbiont|symbiont|symbiotic|epibiont)')

cond\_species\_others = ~df\_taxonomy['species'].str.contains('\s(?:clone|environmental sample)')

cond\_species\_start = df\_taxonomy['species'].str.contains('^(?:Candidatus[\s]){0,1}[\[]{0,1}[A-Z][a-z]+[\]]{0,1}[\s][a-z\-]+')

cond\_species\_invalid = ~df\_taxonomy['species'].str.contains('\s[\(]nom\.\s(?:inval|nud|illeg)\.[\)]')

cond\_filter = cond\_species\_not\_symbiont & cond\_species\_others & cond\_species\_start & cond\_species\_invalid

df\_taxonomy = df\_taxonomy[cond\_filter]