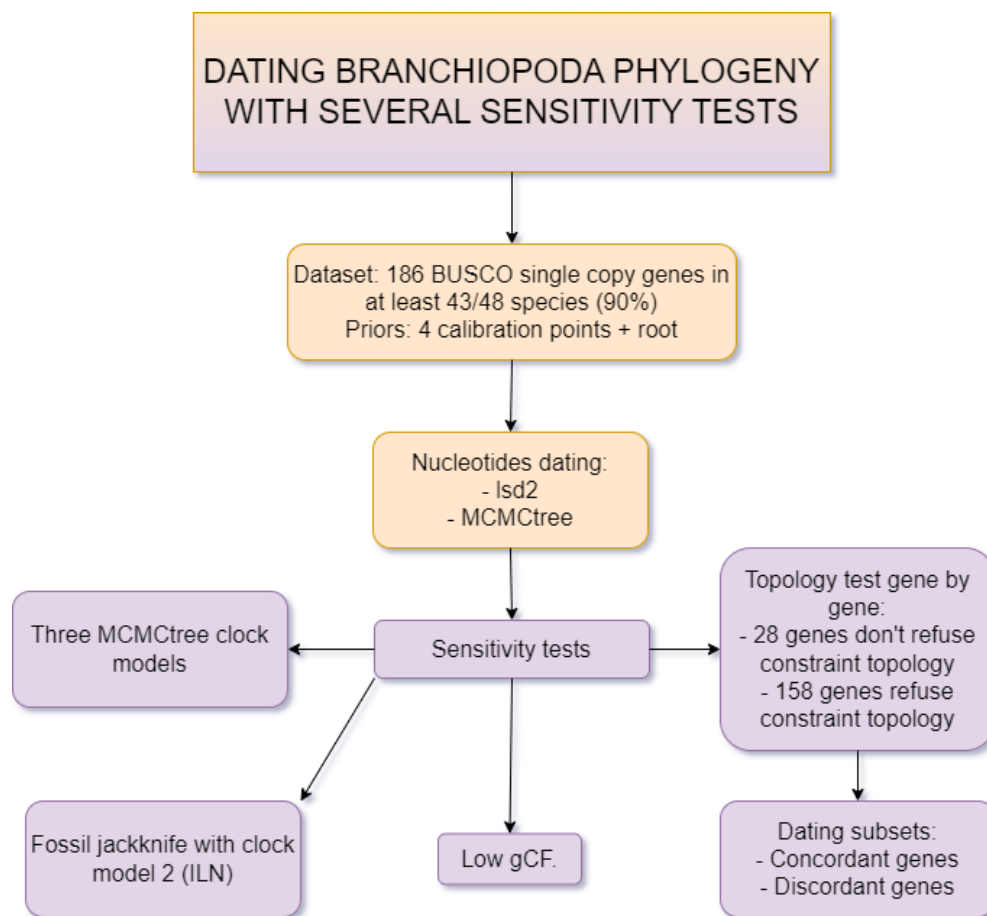


# Branchiopoda Dating

Sumup

Niccolò Righetti

January 2024



## 0.1 Fatto in tesi

### 0.1.1 Filogenesi iqtree nucleotidi

Dataset: 186 geni BUSCO single copy in almeno 43/48 specie (90%).

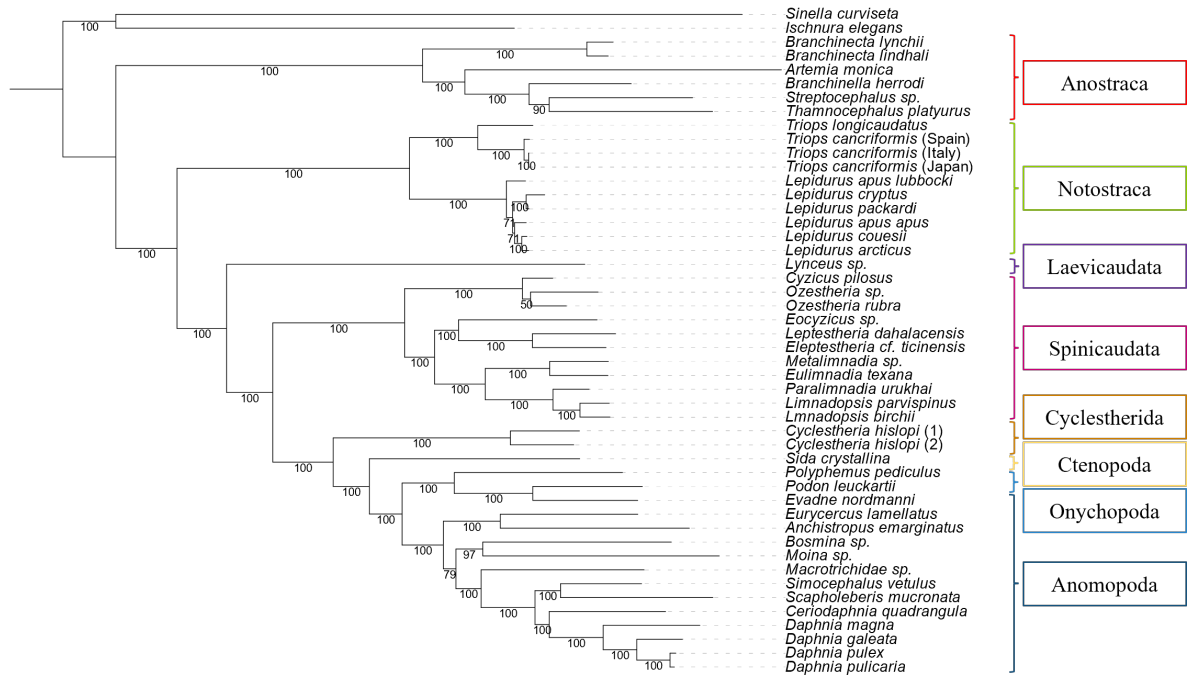


Figure 1: Maximum likelihood phylogenetic tree built on the nucleotide alignment of the 186 BUSCO genes present in at least 90% of the species (43/48) using IQ-TREE. Numbers at branches represent ultrafast bootstrap proportion.

### 0.1.2 Dating sui nucleotidi con lsd2 e i 3 clock model di mcmctree

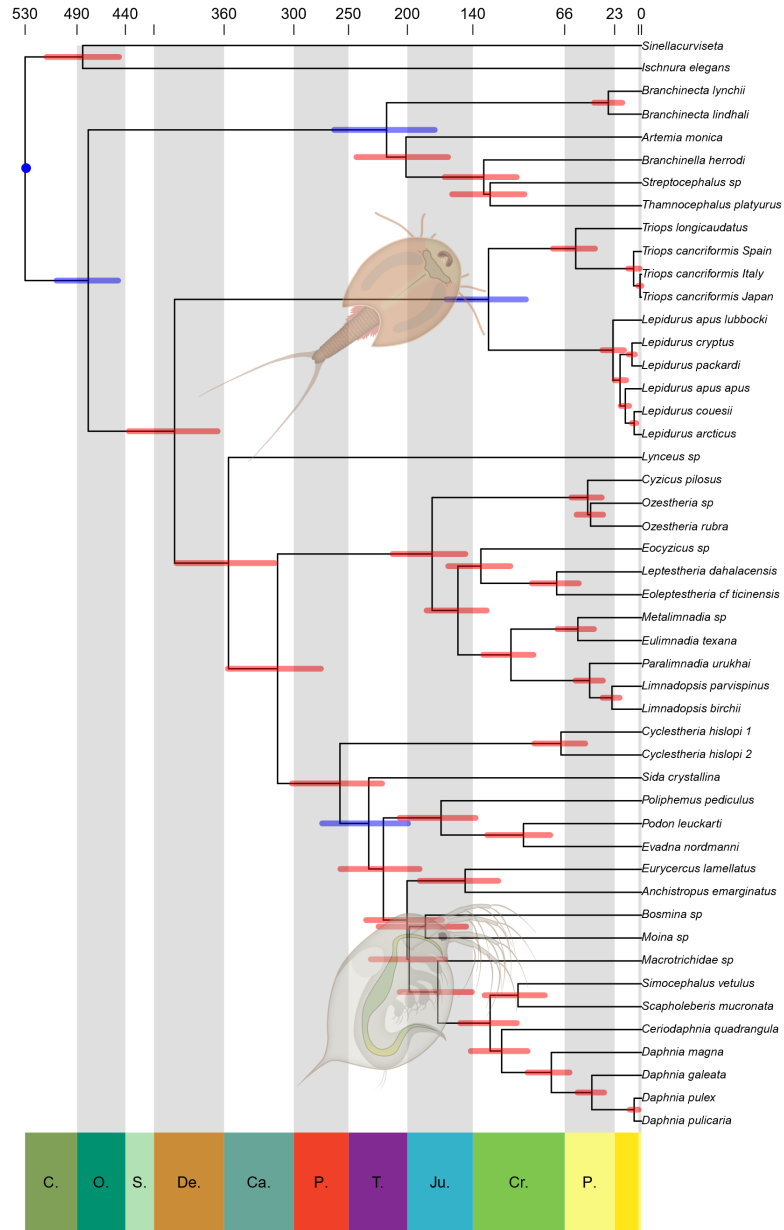


Figure 2: Tree dated with lsd2: blue bars are the prior calibrations. The two silhouettes represent the most studied groups of branchiopods, on top Notostraca and at the bottom Cladocera.

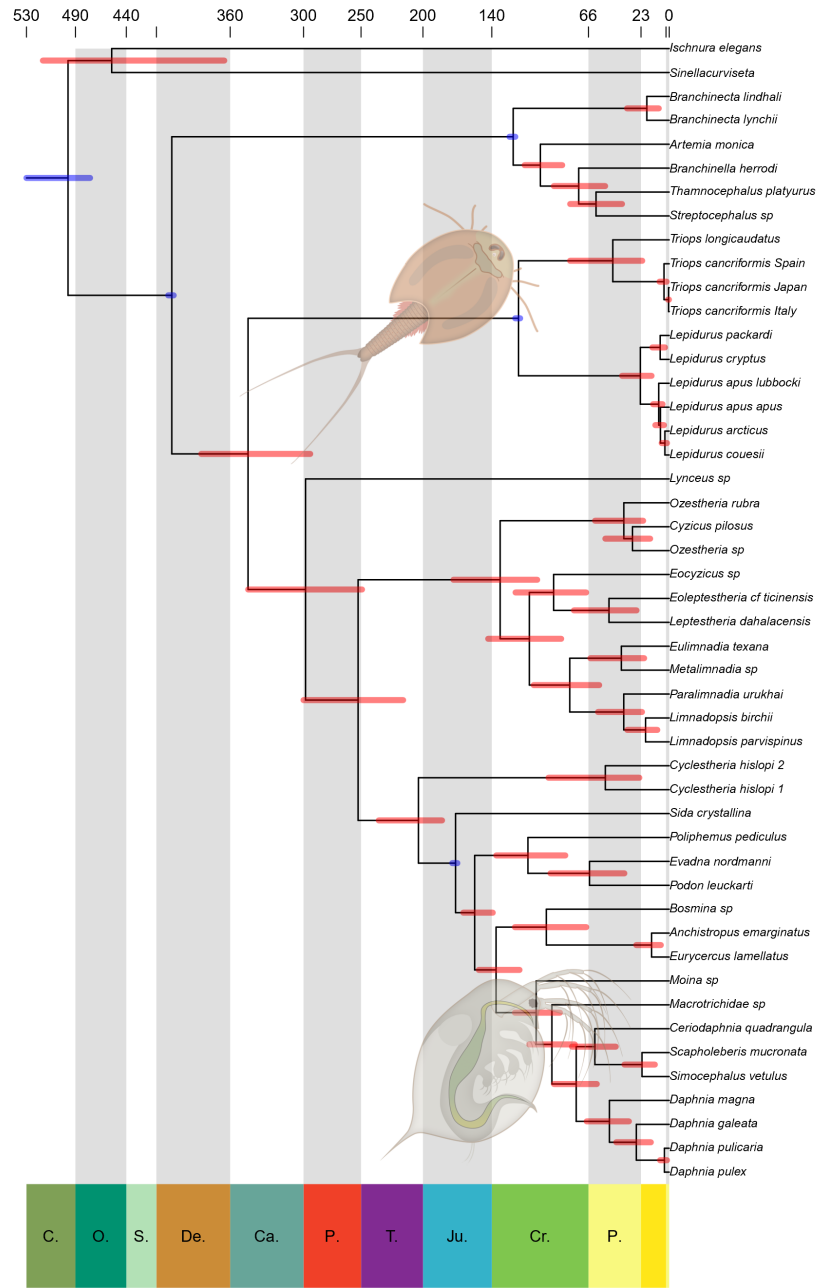


Figure 3: Tree dated with MCMCtree with correlated rates (GBM) Clock Model: blue bars are the prior calibrations. The two silhouettes represent the most studied groups of branchiopods, on top Notostraca and at the bottom Cladocera.

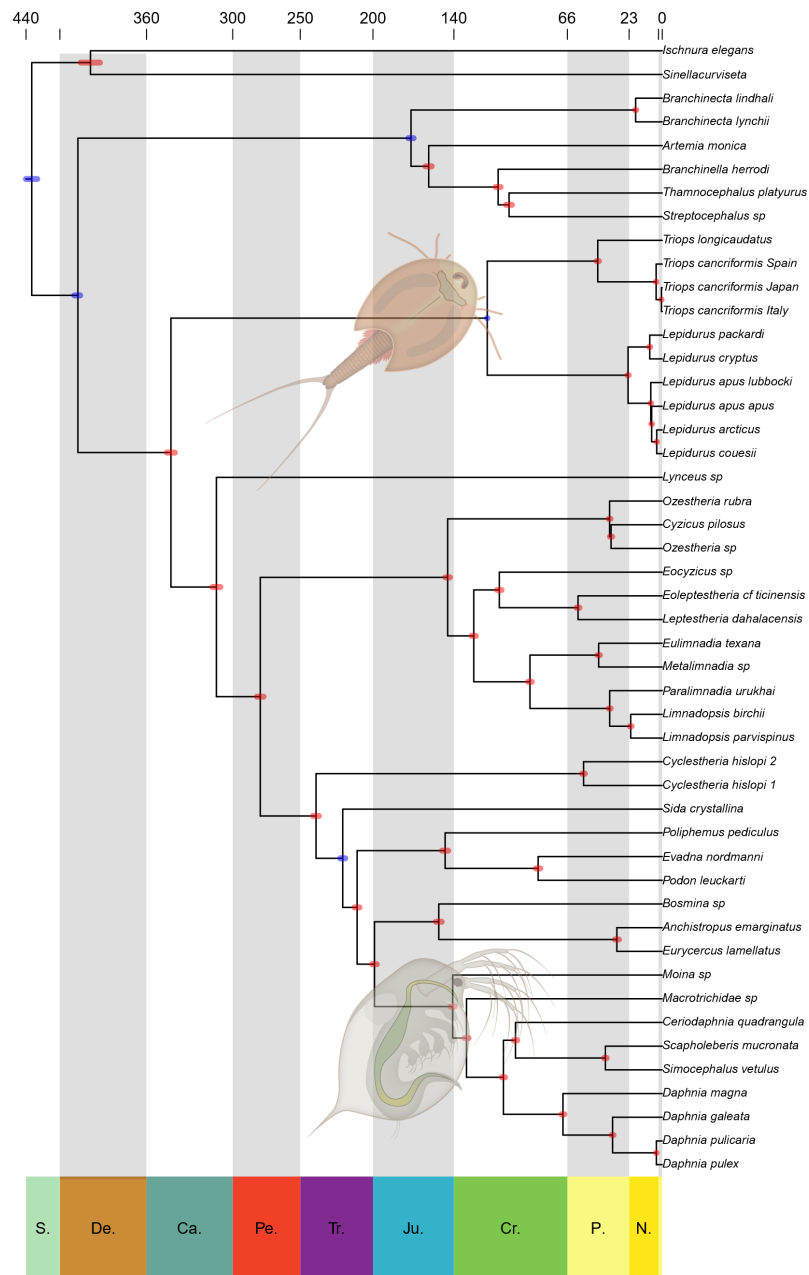


Figure 4: Tree dated with MCMCtree with global (strict) Clock Model: blue bars are the prior calibrations. The two silhouettes represent the most studied groups of branchiopods, on top Notostraca and at the bottom Cladocera.

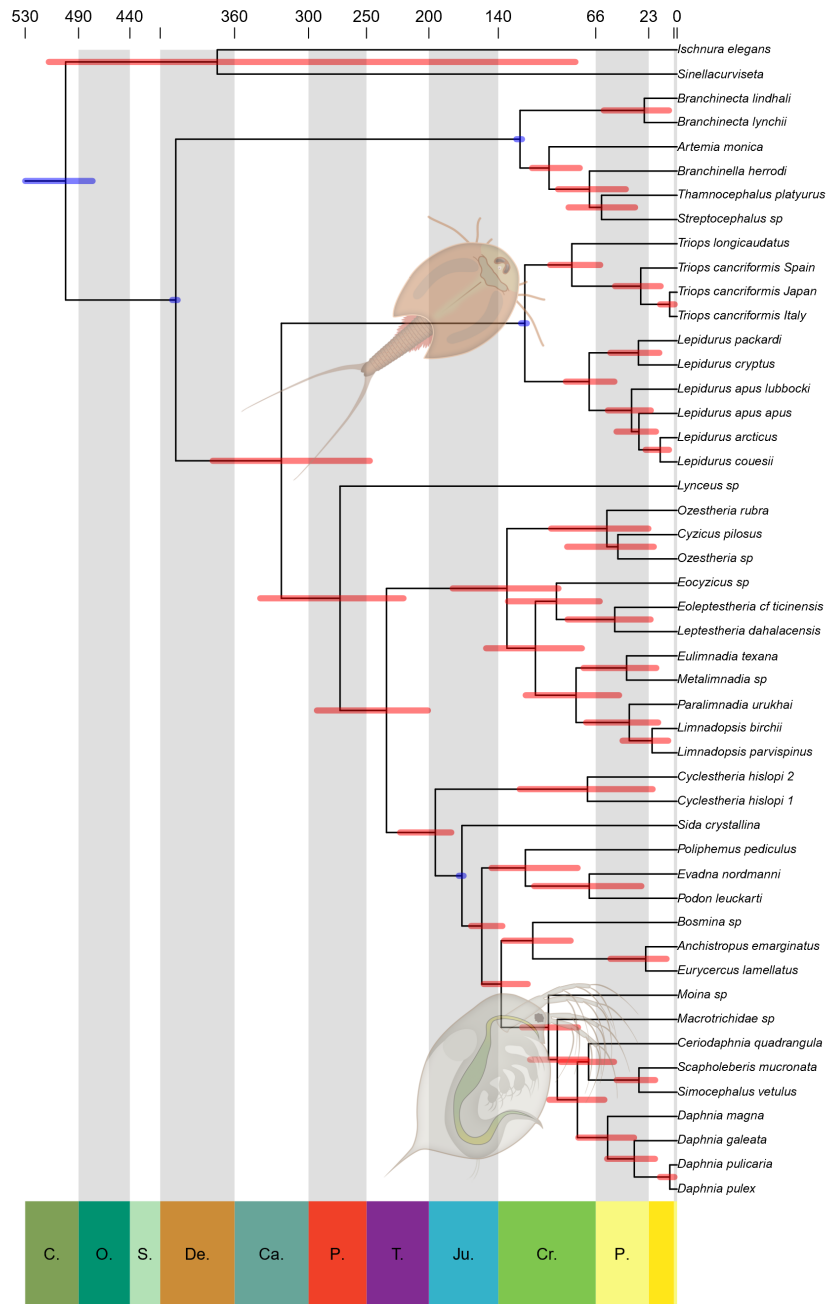


Figure 5: Tree dated with MCMCtree with ILN Clock Model: blue bars are the prior calibrations. The two silhouettes represent the most studied groups of branchiopods, on top Notostraca and at the bottom Cladocera.



Figure 6: Comparison of the dating of the major clades of Branchiopoda with 4 different methods (i.e., lsd2 and MCMCtree with GBM, strict, and ILN clock models) with confidence intervals.

## 0.2 Fatto dopo la laurea

### 0.2.1 Jackknife fossili

Jackknife dei fossili con i 186 geni BUSCO single copy e clock model 2, ovvero l'independent rates.

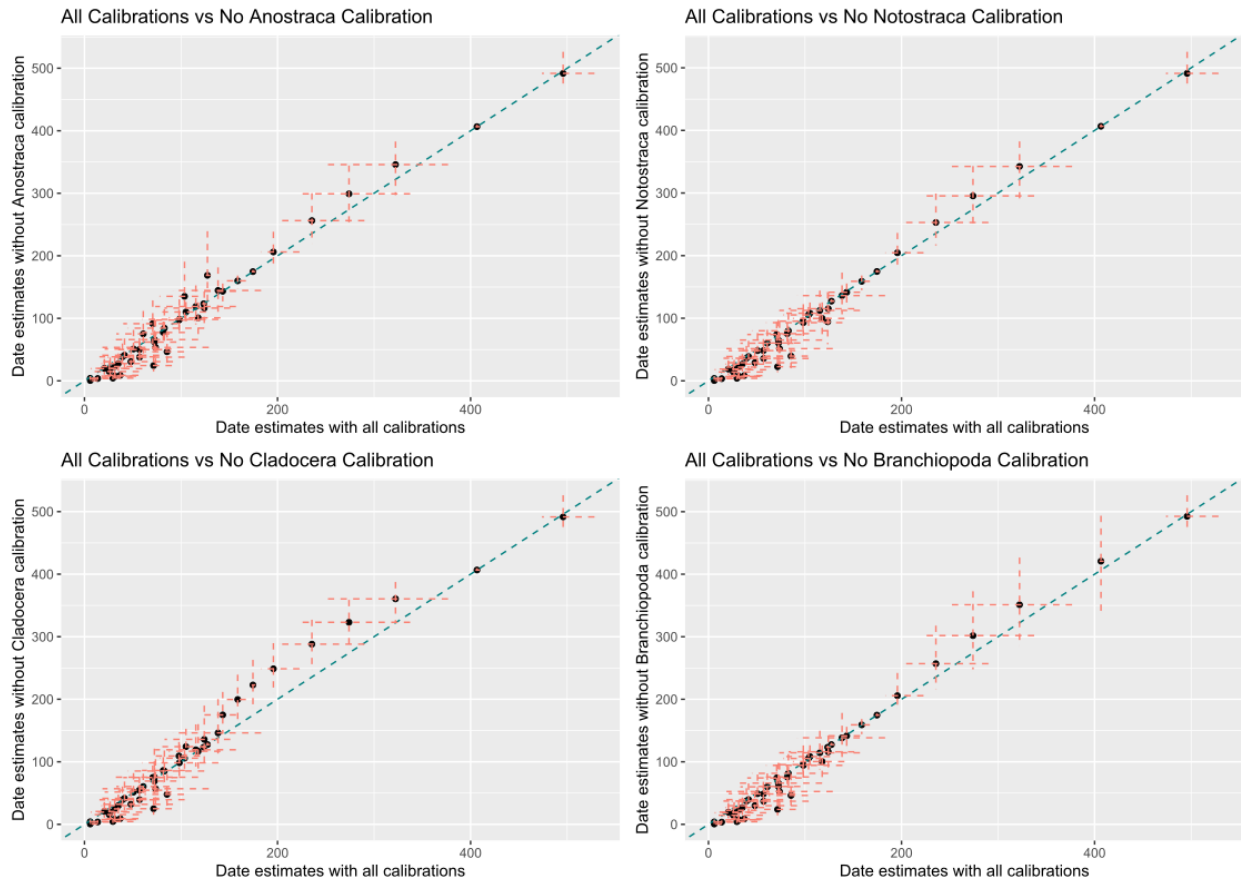


Figure 7: Dating of all nodes removing each time one node



## 0.2.2 gCF bassi

gCF molto bassi a prescindere che siano nodi recenti o antichi:

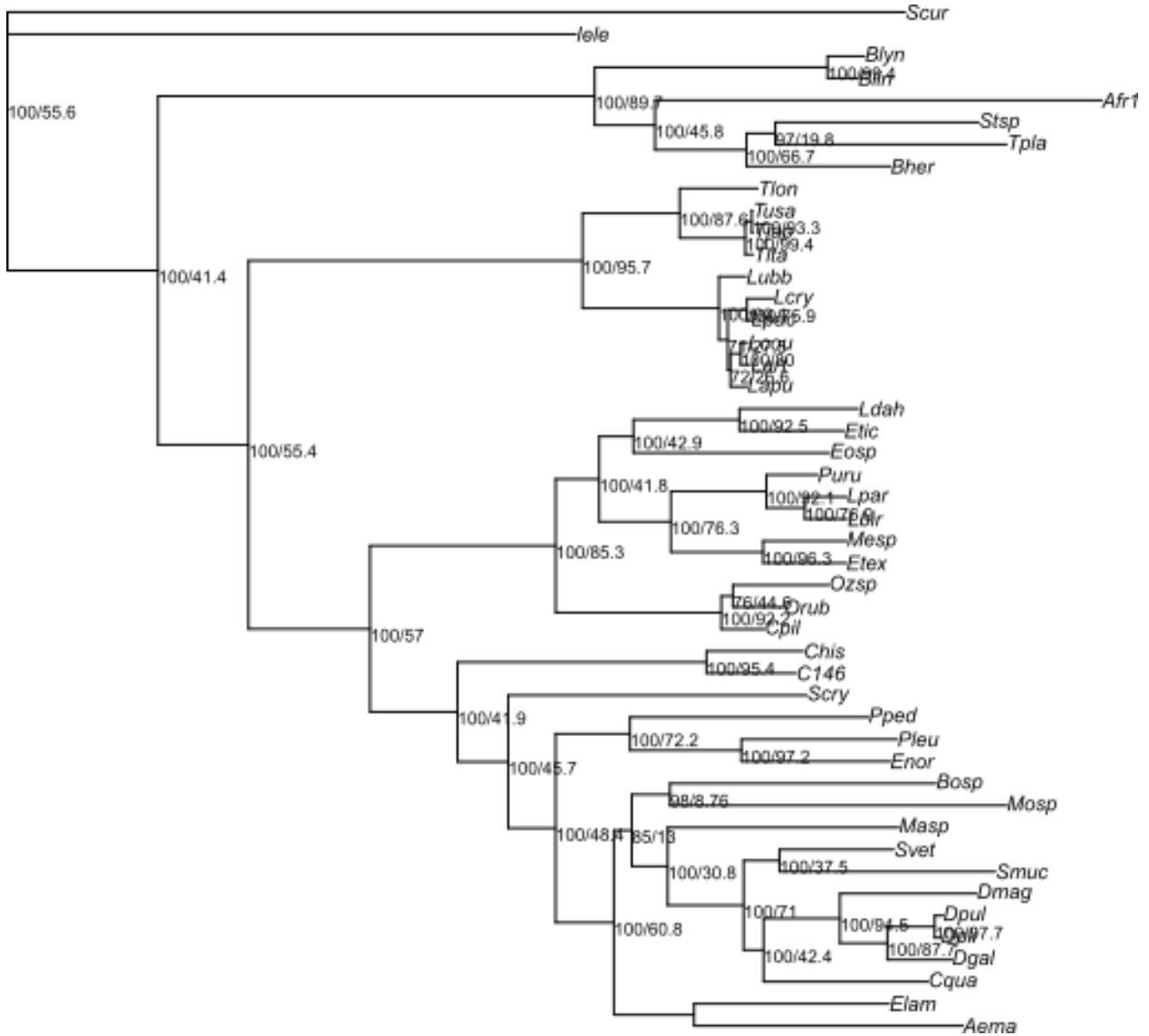


Figure 8: gCF tree with all species, each branch label is BS/gCF

[illegible]

Figure 9: gCF tree without *Lynceus* sp., each branch label is BS/gCF

Questo si può vedere meglio dal seguente plot dei gCF coi due dataset.

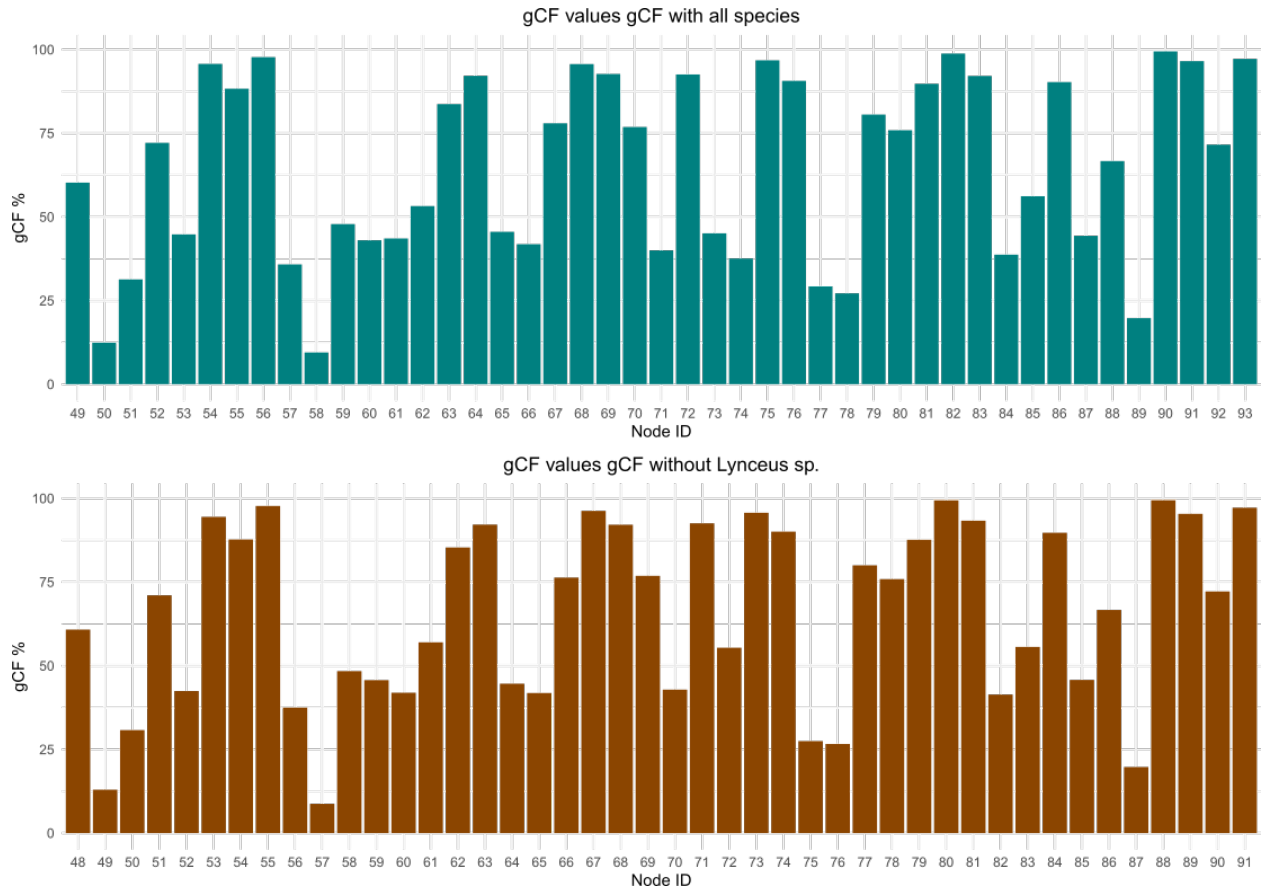


Figure 10: gCF distributions with the two datasets

Non avendo ricevuto risposte certe riguardo il funzionamento dei gCF in caso di assenza di specie negli allineamenti dei singoli geni, ho provato a fare una filogenesi con gli allineamenti contenenti tutte le specie. Purtroppo sono solo 3, di conseguenza i gCF vengono per ogni nodo 0%-33%-66%-100%, non risolvendo il dubbio.

### 0.2.3 Topology test

Visti i dubbi sui gCF e sulle topologie discordi, ho fatto due gene tree per ogni gene, uno senza constraint e uno con constraint la topologia generale. Ho poi fatto dei topology test a coppie gene per gene per vedere quali geni rifiutassero la topologia constraint.

Risultato:

- 28 geni non rifiutano la topologia (detti per comodità "concordant genes")
- 159 geni rifiutano la topologia (detti "discordant genes")

Ho quindi fatto due allineamenti concatenati e rifatto il dating. Come si vede dal grafico seguente, non si evidenziano particolari differenze con il dating con tutti i geni insieme.

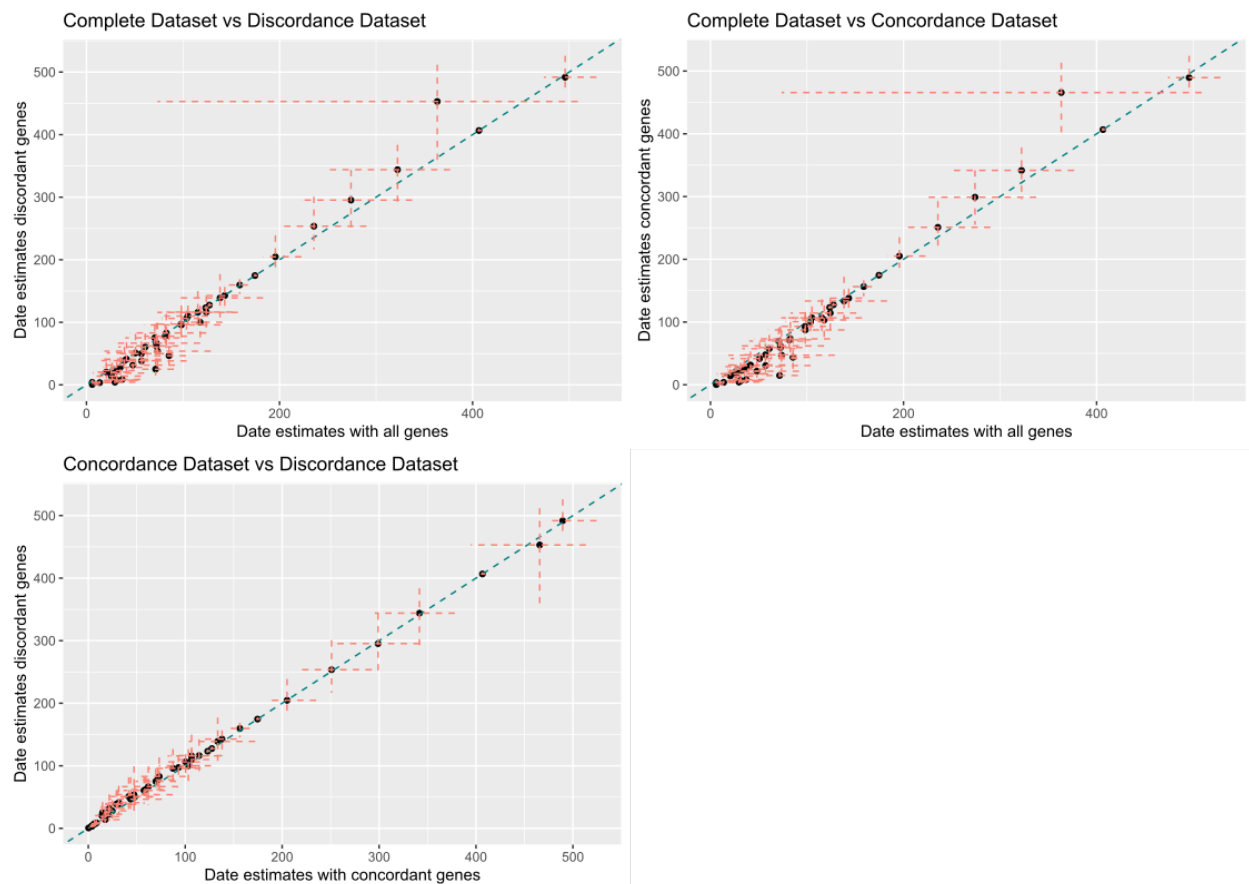


Figure 11: Comparison of the date estimates for each node with the three datasets: all genes, concordant genes, discordant genes.

Visto quanto osservato con Lynceus, ho deciso di rifare il processo togliendolo dagli allineamenti. Questo ha aumentato notevolmente il numero di concordant genes.

- 181 concordant genes
- 6 discordant genes

Purtroppo però ancora non ho ancora potuto vedere se/come cambia il dating con questo subset.