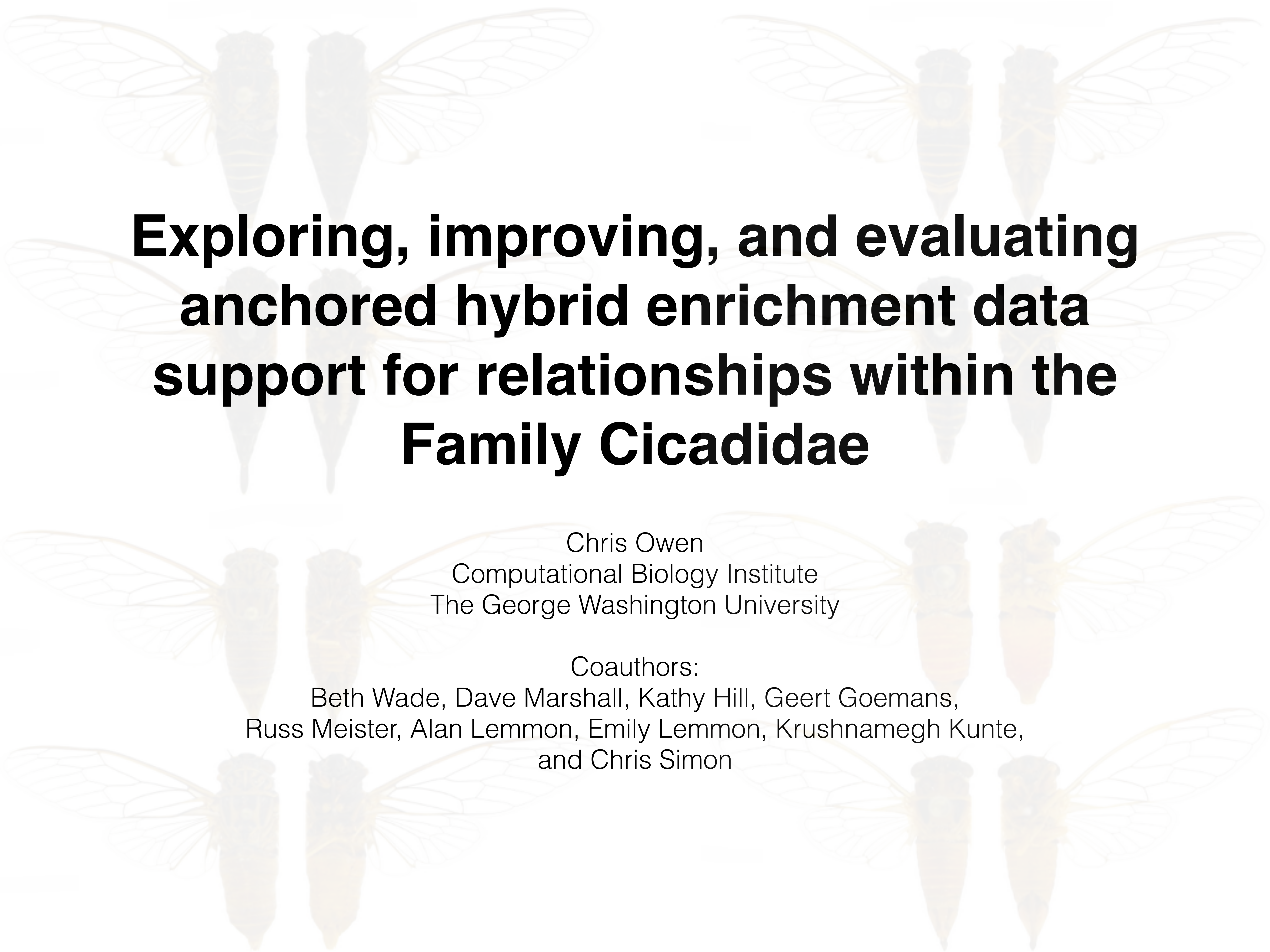


Exploring, improving, and evaluating anchored hybrid enrichment data support for relationships within the Family Cicadidae

Chris Owen
Computational Biology Institute
The George Washington University

Coauthors:
Beth Wade, Dave Marshall, Kathy Hill, Geert Goemans,
Russ Meister, Alan Lemmon, Emily Lemmon, Krushnamegh Kunte,
and Chris Simon

Outline

1. Lessons learned from transcriptome phylogenomic datasets

2. Dealing with the unknown: what should the phylogeny look like and what are these sequences?

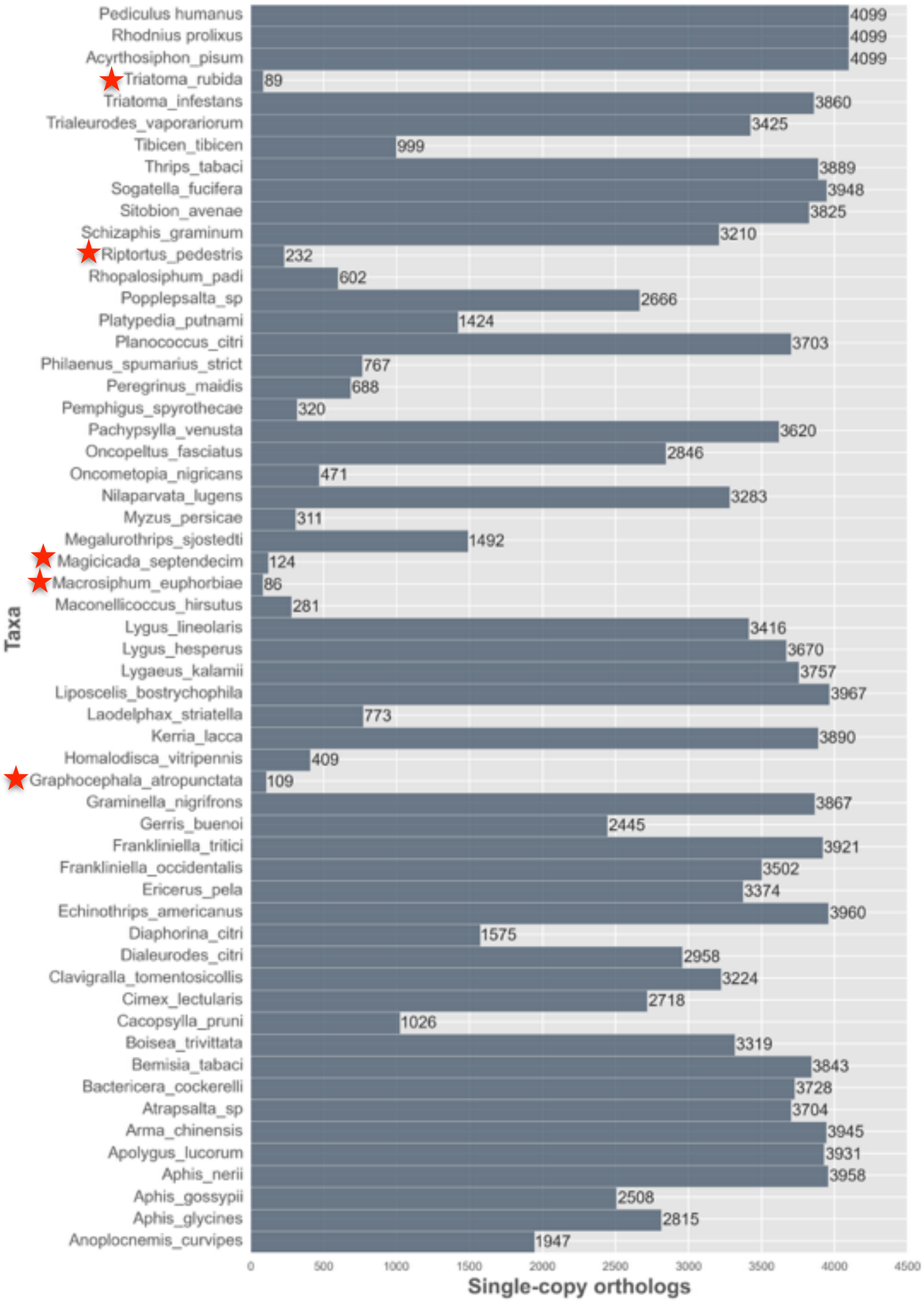
3. Model violation and gene filtering

4. Gene tree / species tree estimation

5. Exploring tree space

Lessons learned from Hemiptera transcriptome phylogenomic datasets

Taxon & Ortholog Sampling



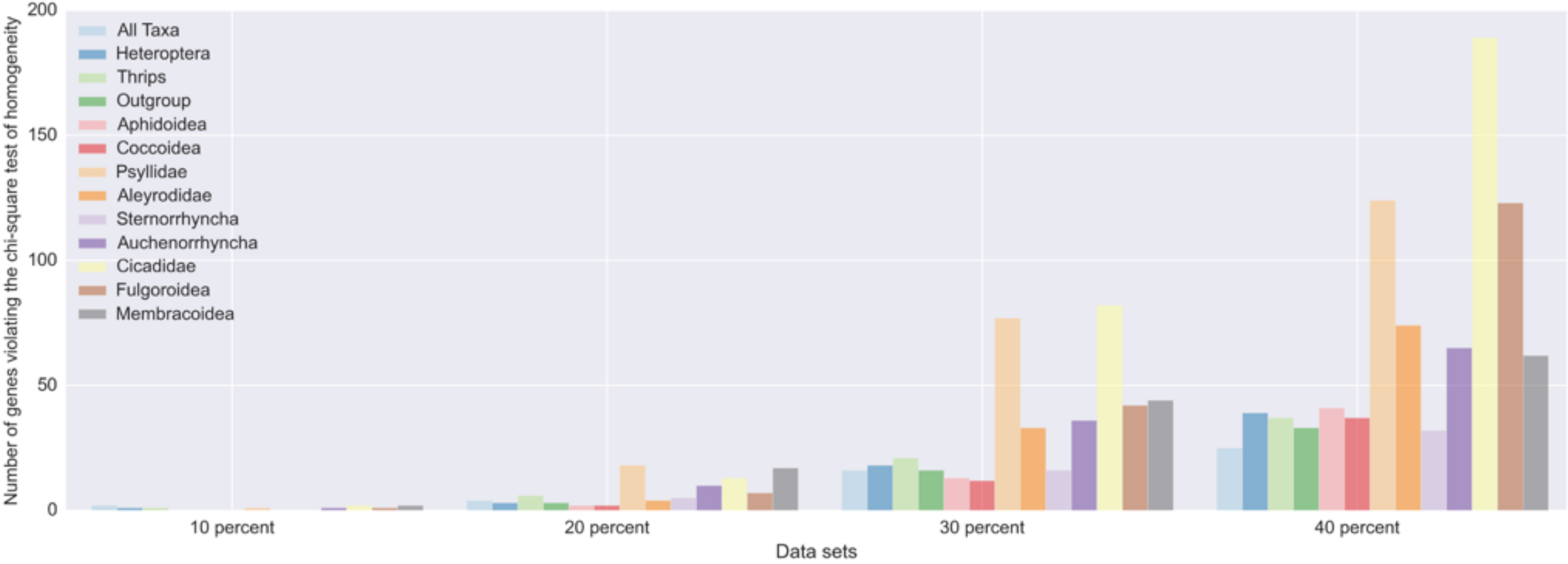
Outgroups
- Thysanoptera: 5 species
- Phthiraptera: 1 species
- Psocoptera: 1 species

Auchenorrhyncha
- Cicadoidea: 5 species
- Cercopoidea: 1 species
- Membracoidea: 4 species
- Fulgoroidea: 4 species

Heteroptera
- Pentatomoidea: 1 species
- Coreoidea: 4 species
- Lygaeoidea: 2 species
- Cimicoidea: 1 species
- Miroidea: 3 species
- Reduvioidea: 3 species
- Gerroidea: 1 species

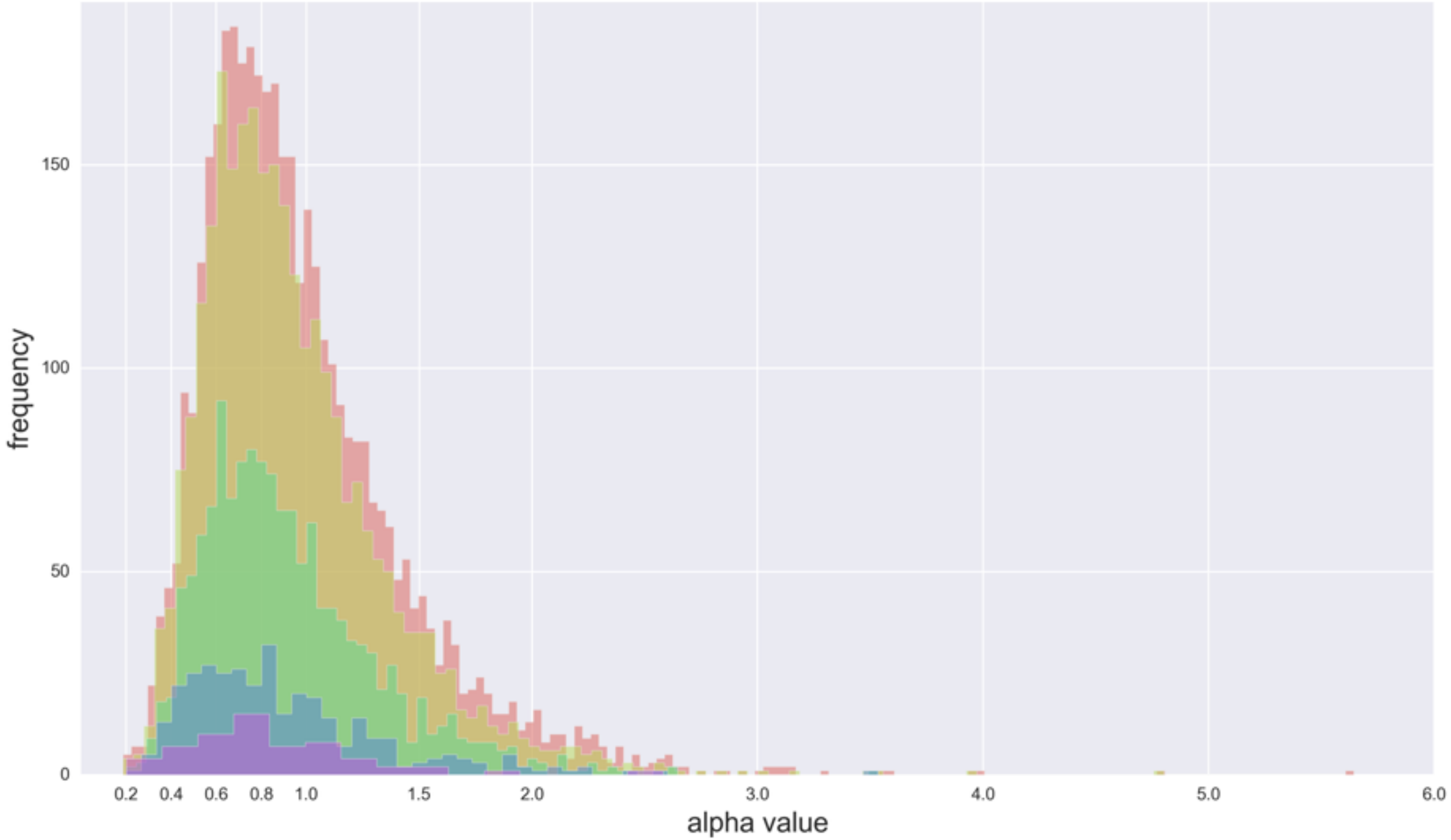
Sternorrhyncha
- Coccoidea: 4 species
- Aphidoidea: 10 species
- Psylloidea: 4 species
- Aleyrodidae: 3 species

Choosing the best genes: compositional bias



- Chi-square test among taxa and higher taxa ($p<0.5$)
- Suffers from type 2 error (Foster 2004)

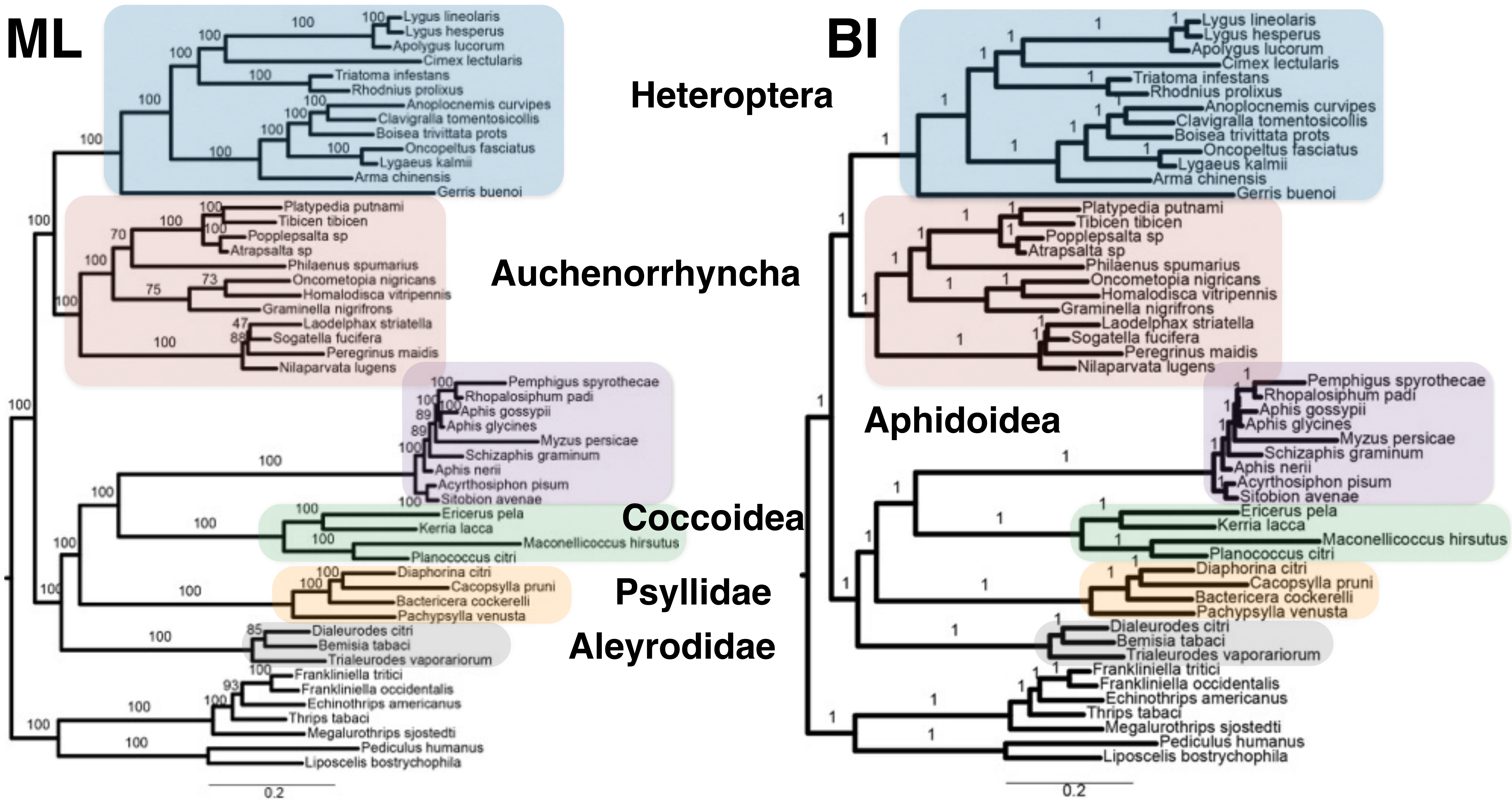
Choosing the best genes: among site rate variation



- removed genes alpha value < 1 (Yang 1994)

Hemiptera phylogeny: most curated dataset

Alignment: 463 genes; 106,740 AA; 52 taxa

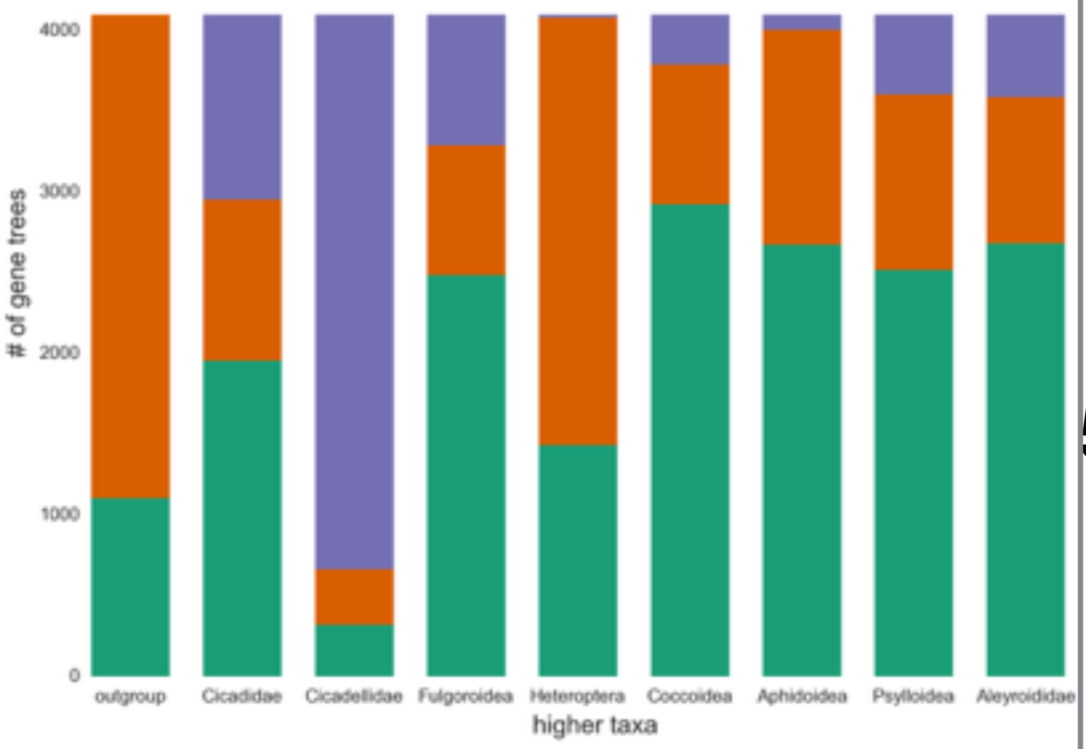


Comparison of monophyly among datasets

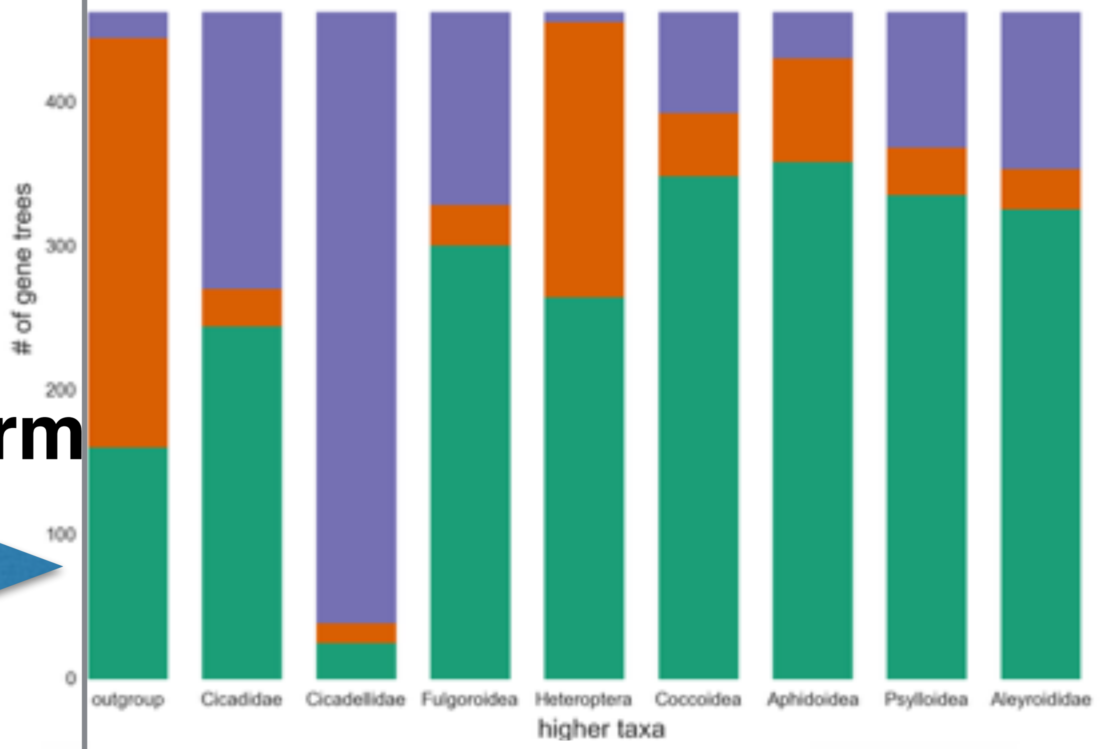
🟩 monophyletic 🟧 non-monophyletic 🟪 single sp.

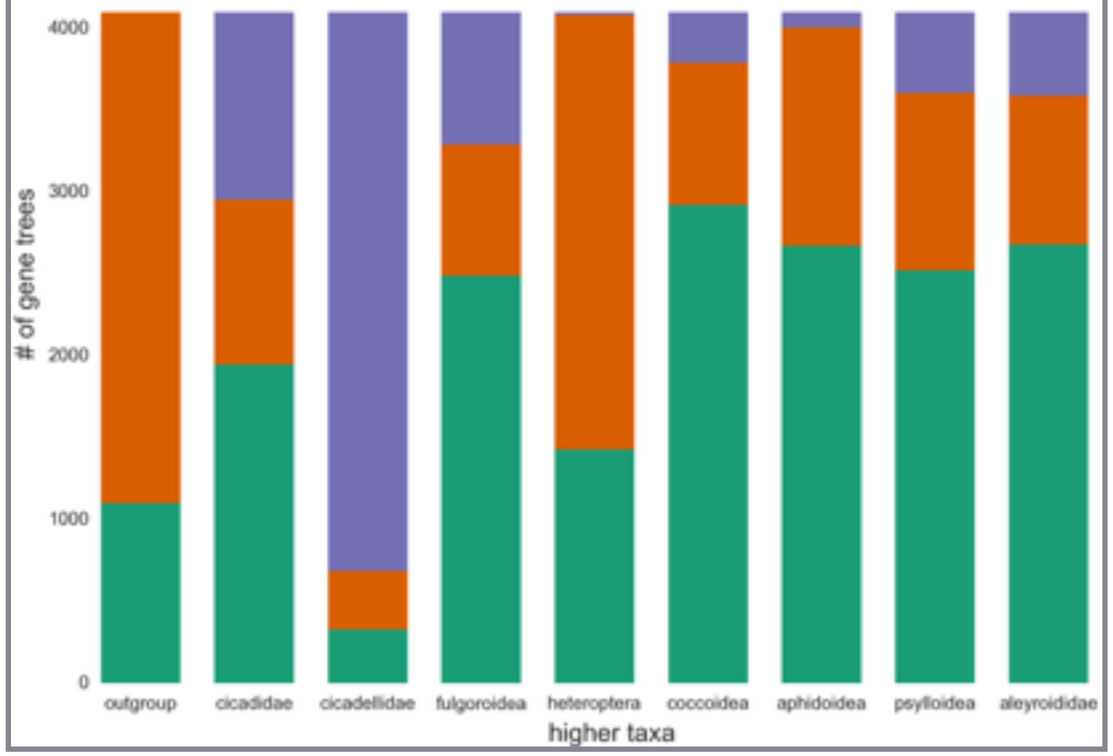
All genes

Curated dataset

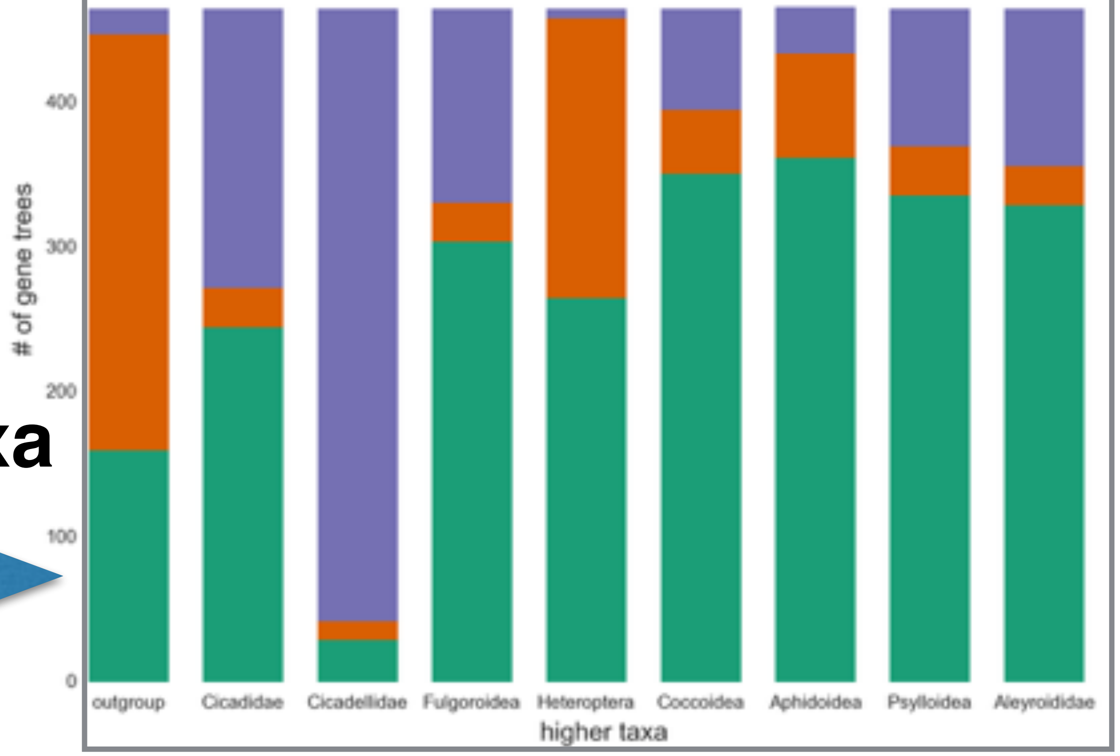


5 taxa rm





All taxa

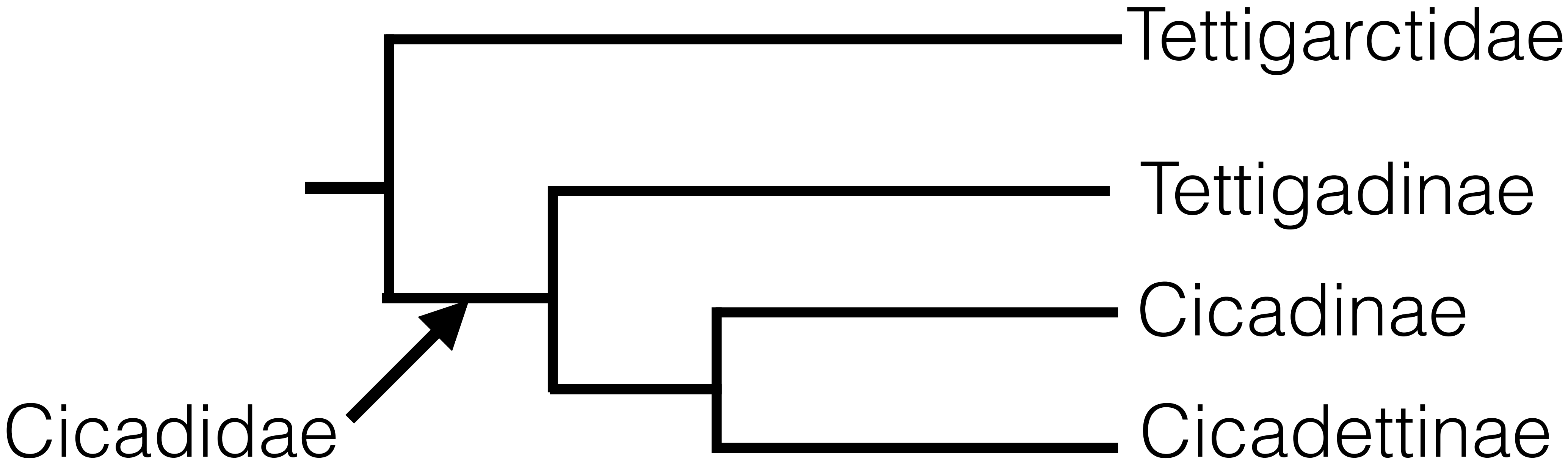


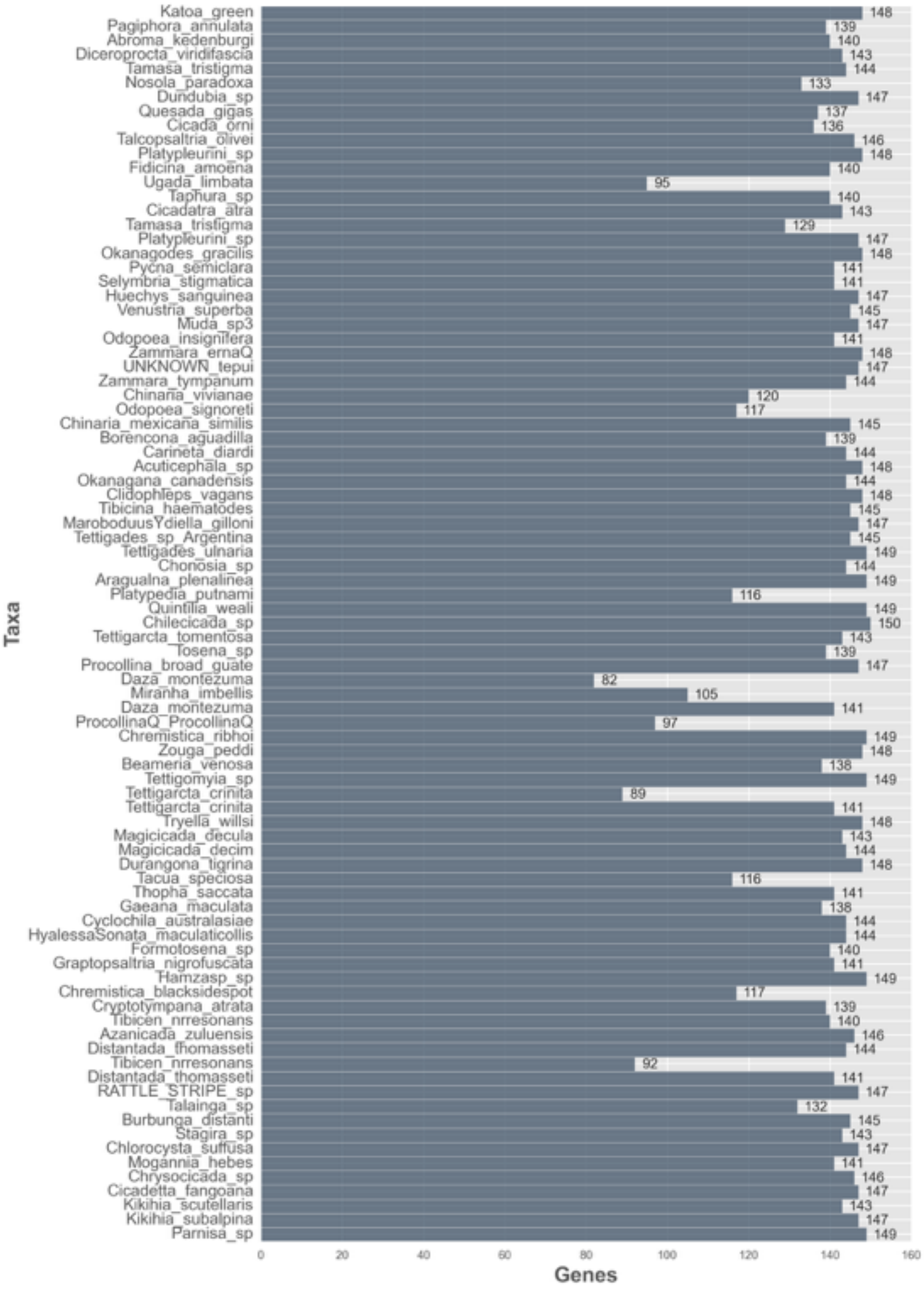
Cicadidae hybrid capture phylogenomics

Dealing with the unknown: what should the phylogeny look like?

- No molecular hypotheses of subfamily and tribal relationships

- Moulds (2005) proposed subfamily relationships and Australian tribal relationships (117 morphological characters)





Gene matrix occupancy

- 150 loci; 87 taxa; ~50k bp
- 2 taxa < 90 loci
- 2 taxa 90 < loci < 100
- 1 taxon with 150 loci

Dealing with the unknown: what are these loci?

Pipeline

1. Blast cds transcripts to loci or map transcriptome short reads to loci using bowtie2
2. Use MACSE (Ranwez et al. 2011) for first alignment
3. Use Muscle for refinement (-*refine*) to clean it up

Results

147 loci are a combination of coding and non-coding

3 loci are non-coding UTR’s

Sequence partitioning, modeling, and phylogeny estimation

1. Each gene partitioned by coding PartitionFinder and non-coding

2. Models estimated in PartitionFinder

3. Gblocks / no Gblocks

4. Phylogenies:

    - Partitioned gene trees: Garli

    - Concatenated matrices (including k-means cluster modeling (Frandsen et al. 2015)): RAxML

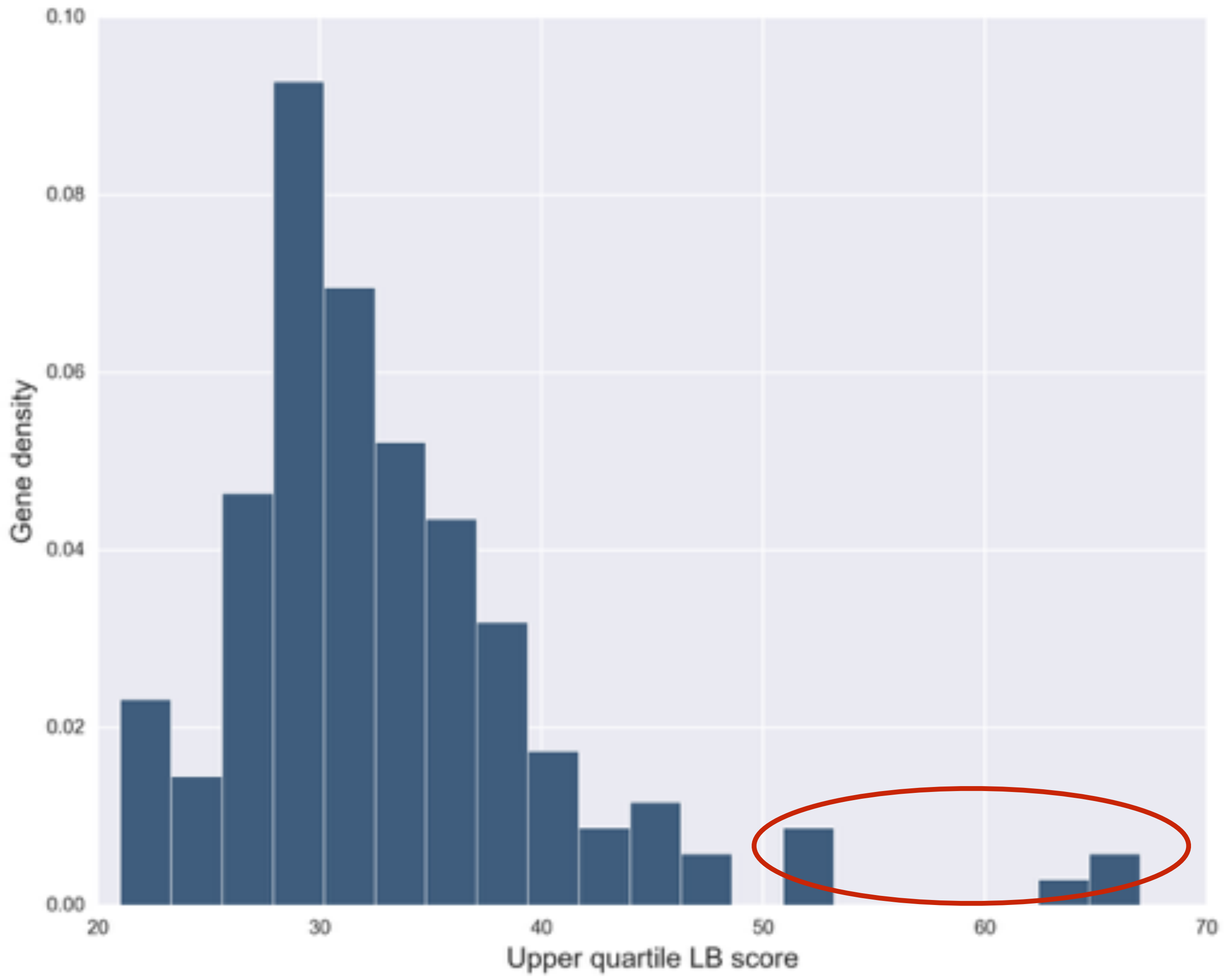
    - Concatenated partitioned matrices: RAxML

    - Gene tree / species tree methods: Astral

Gene filtering: base composition

- Chi-square test in BaCoCa (Kuck and Struck 2014)
- Cicadidae hybrid capture dataset: No significant differences at $p = 0.05$ -> $p = 0.20$

Gene tree filtering: removing loci with long branches



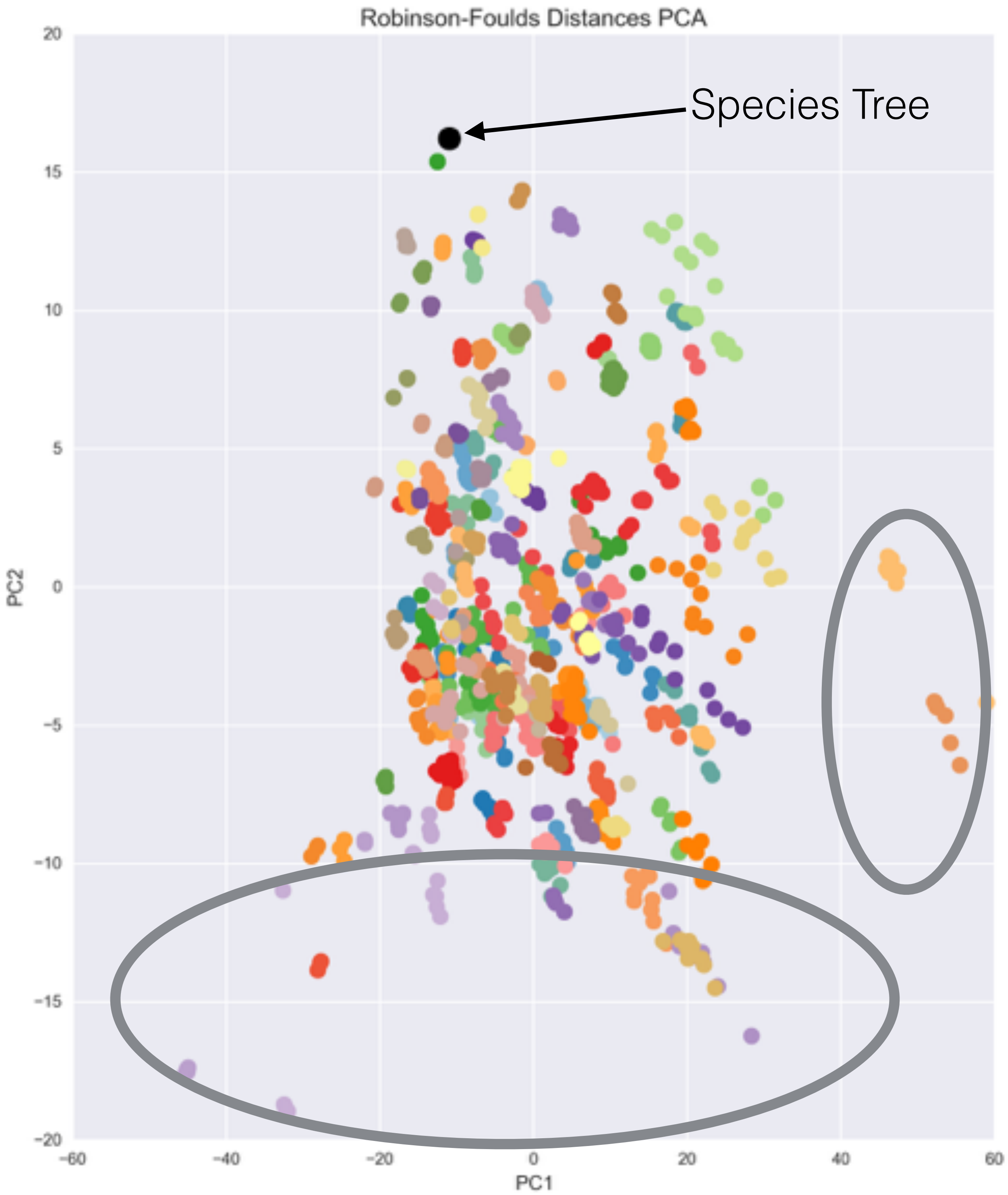
Long Branch Score

$$LB_i = \left(\frac{\overline{PD_i}}{\overline{PD_a}} - 1\right) * 100$$

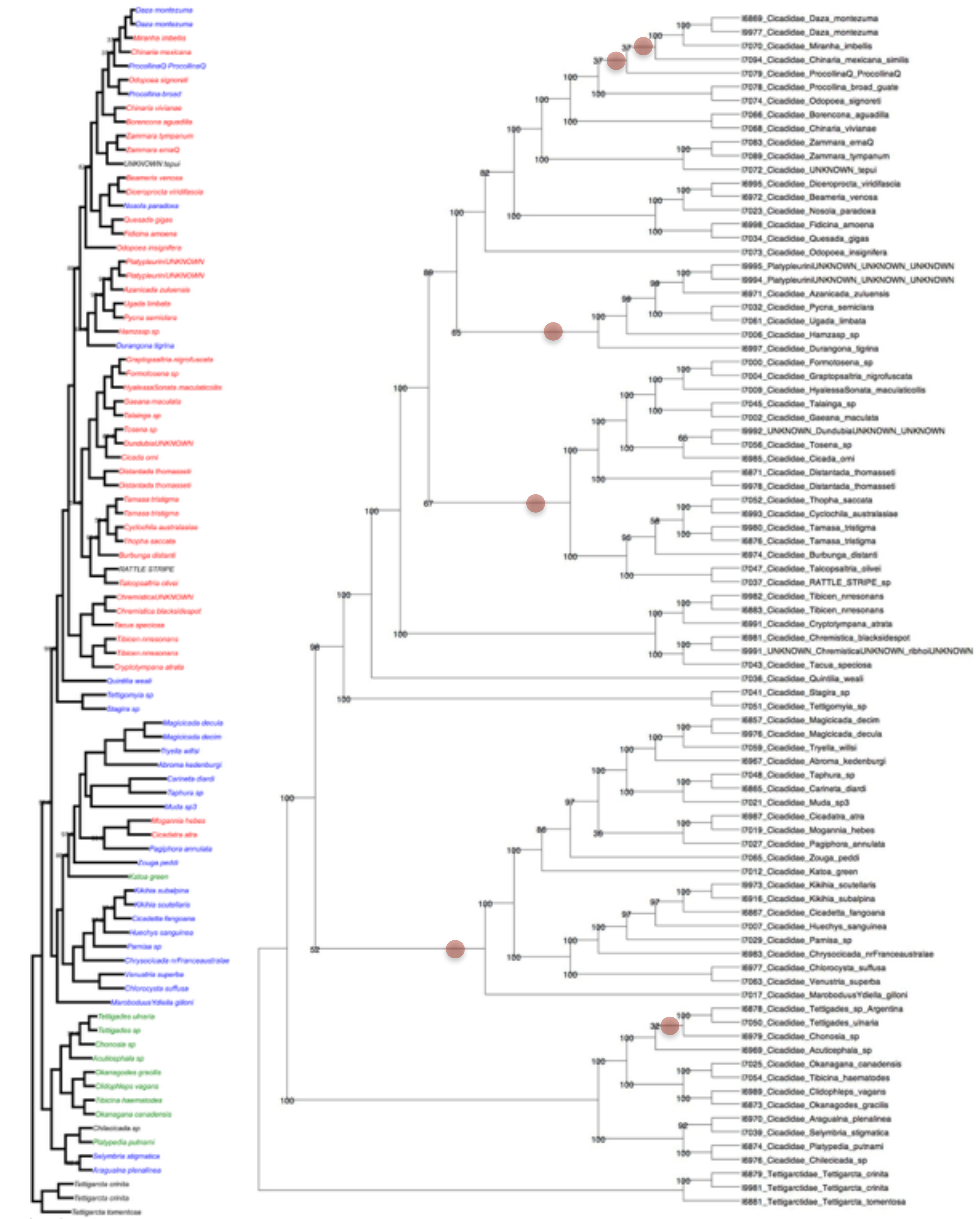
- mean pairwise patristic distance (PD) of a taxon to all other taxa in the tree relative to the avg. pairwise PD over all taxa

- LB score upper quartile of each partition in TreSpEx (Struck 2014)
- unrooted gene trees
- Brinkman & Philippe 2008; Bergsten 2005

Gene filtering: Robinson-Foulds distances



- Weighted Robinson-Foulds distances
    - shared bipartitions and sum of square differences in branch lengths
- 10 best trees from each gene tree search
- Species tree from RAxML partitioned concatenated search
- Removed 12 genes


Daza montezuma
Daza montezuma
Miranha imbellis
Chinaria mexicana
ProcollinaQ ProcollinaQ
Odopoea signoreti
Procollina broad
Chinaria vivianae
Borencona aguadilla
Zammara tympanum
Zammara ernaQ
UNKNOWN tepui
Beameria venosa
Diceroprocta viridifascia
Nosola paradoxa
Quesada gigas
Fidicina amoena
Odopoea insignifera
PlatypleuriniUNKNOWN
PlatypleuriniUNKNOWN
Azanicada zuluensis
Ugada limbata
Pycna semiclara
Hamzasp sp
Durangona tigrina
Graptopsaltria nigrofuscata
Formotosena sp
HyalessaSonata maculaticollis
Gaeana maculata
Talainga sp
Tosena sp
DundubiaUNKNOWN
Cicada orni
Distantada thomasseti
Distantada thomasseti
Tamasa tristigma
Tamasa tristigma
Cyclochila australasiae
Thopha saccata
Burbunga distanti
RATTLE STRIPE
Talcopsaltria olivei
ChremisticaUNKNOWN
Chremistica blacksidespot
Tacua speciosa
Tibicen nmesonans
Tibicen nmesonans
Cryptotympana atrata
Quintilia weali
Tettigomyia sp
Stagira sp
Magicicada decula
Magicicada decim
Tryella willsi
Abroma kedenburgi
Carineta diardi
Taphura sp
Muda sp3
Mogannia hebes
Cicadatra atra
Pagiphora annulata
Zouga peddi
Katoa green
Kikihia subalpina
Kikihia scutellaris
Cicadetta fangoana
Huechys sanguinea
Pamisa sp
Chrysocicada nrFranceaustralae
Venustria superba
Chlorocysta suffusa
MaroboduusYdiella gilloni
Tettigades ulnaria
Tettigades sp
Chonosia sp
Acuticephala sp
Okanagodes gracilis
Clidophleps vagans
Tibicina haematodes
Okanagana canadensis
Chilecicada sp
Platypedia putnami
Selymbria stigmatica
Aragualna plenalinea
Tettigarcta crinita
Tettigarcta crinita
Tettigarcta tomentosa
I6869_Cicadidae_Daza_montezuma
I9977_Cicadidae_Daza_montezuma
I7070_Cicadidae_Miranha_imbellis
I7094_Cicadidae_Chinaria_mexicana_similis
I7079_Cicadidae_ProcollinaQ_ProcollinaQ
I7078_Cicadidae_Procollina_broad_guate
I7074_Cicadidae_Odopoea_signoreti
I7066_Cicadidae_Borencona_aguadilla
I7068_Cicadidae_Chinaria_vivianae
I7083_Cicadidae_Zammara_ernaQ
I7089_Cicadidae_Zammara_tympanum
I7072_Cicadidae_UNKNOWN_tepui
I6995_Cicadidae_Diceroprocta_viridifascia
I6972_Cicadidae_Beameria_venosa
I7023_Cicadidae_Nosola_paradoxa
I6998_Cicadidae_Fidicina_amoena
I7034_Cicadidae_Quesada_gigas
I7073_Cicadidae_Odopoea_insignifera
I9995_PlatypleuriniUNKNOWN_UNKNOWN_UNKNOWN
I9994_PlatypleuriniUNKNOWN_UNKNOWN_UNKNOWN
I6971_Cicadidae_Azanicada_zuluensis
I7032_Cicadidae_Pycna_semiclara
I7061_Cicadidae_Ugada_limbata
I7006_Cicadidae_Hamzasp_sp
I6997_Cicadidae_Durangona_tigrina
I7000_Cicadidae_Formotosena_sp
I7004_Cicadidae_Graptopsaltria_nigrofuscata
I7009_Cicadidae_HyalessaSonata_maculaticollis
I7045_Cicadidae_Talainga_sp
I7002_Cicadidae_Gaeana_maculata
I9992_UNKNOWN_DundubiaUNKNOWN_UNKNOWN
I7056_Cicadidae_Tosena_sp
I6985_Cicadidae_Cicada_orni
I6871_Cicadidae_Distantada_thomasseti
I9978_Cicadidae_Distantada_thomasseti
I7052_Cicadidae_Thopha_saccata
I6993_Cicadidae_Cyclochila_australasiae
I9980_Cicadidae_Tamasa_tristigma
I6876_Cicadidae_Tamasa_tristigma
I6974_Cicadidae_Burbunga_distanti
I7047_Cicadidae_Talcopsaltria_olivei
I7037_Cicadidae_RATTLE_STRIPE_sp
I9982_Cicadidae_Tibicen_nmesonans
I6883_Cicadidae_Tibicen_nmesonans
I6991_Cicadidae_Cryptotympana_atrata
I6981_Cicadidae_Chremistica_blacksidespot
I9991_UNKNOWN_ChremisticaUNKNOWN_ribhoiUNKNOWN
I7043_Cicadidae_Tacua_speciosa
I7036_Cicadidae_Quintilia_weali
I7041_Cicadidae_Stagira_sp
I7051_Cicadidae_Tettigomyia_sp
I6857_Cicadidae_Magicicada_decim
I9976_Cicadidae_Magicicada_decula
I7059_Cicadidae_Tryella_willsi
I6967_Cicadidae_Abroma_kedenburgi
I7048_Cicadidae_Taphura_sp
I6865_Cicadidae_Carineta_diardi
I7021_Cicadidae_Muda_sp3
I6987_Cicadidae_Cicadatra_atra
I7019_Cicadidae_Mogannia_hebes
I7027_Cicadidae_Pagiphora_annulata
I7065_Cicadidae_Zouga_peddi
I7012_Cicadidae_Katoa_green
I9973_Cicadidae_Kikihia_scutellaris
I6916_Cicadidae_Kikihia_subalpina
I6867_Cicadidae_Cicadetta_fangoana
I7007_Cicadidae_Huechys_sanguinea
I7029_Cicadidae_Pamisa_sp
I6963_Cicadidae_Chrysocicada_nrFranceaustralae
I6977_Cicadidae_Chlorocysta_suffusa
I7063_Cicadidae_Venustria_superba
I7017_Cicadidae_MaroboduusYdiella_gilloni
I6878_Cicadidae_Tettigades_sp_Argentina
I7050_Cicadidae_Tettigades_ulnaria
I6979_Cicadidae_Chonosia_sp
I6969_Cicadidae_Acuticephala_sp
I7025_Cicadidae_Okanagana_canadensis
I7054_Cicadidae_Tibicina_haematodes
I6989_Cicadidae_Clidophleps_vagans
I6873_Cicadidae_Okanagodes_gracilis
I6970_Cicadidae_Aragualna_plenalinea
I7039_Cicadidae_Selymbria_stigmatica
I6874_Cicadidae_Platypedia_putnami
I6976_Cicadidae_Chilecicada_sp
I6879_Tettigarctidae_Tettigarcta_crinita
I9981_Tettigarctidae_Tettigarcta_crinita
I6881_Tettigarctidae_Tettigarcta_tomentosa

Concatenated Tree

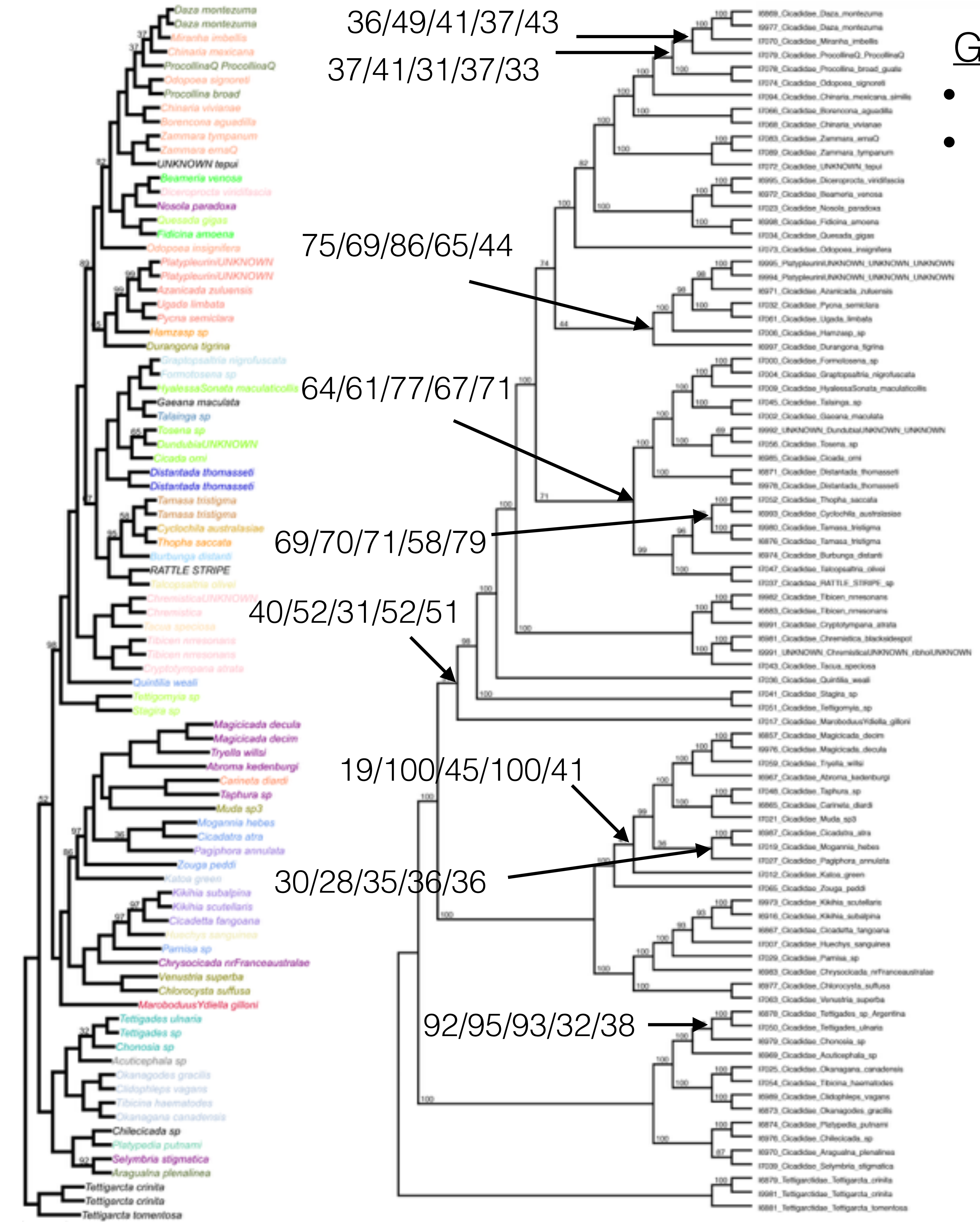
- 6 branches not supported by bootstrap support
- 2 unsupported branches are sister to larger clades
- RAxML gene partitioned

🟥 Cicadinae

🟦 Cicadettinae

🟩 Tettigadinae

● BS < 67



Gene filtering and alternative models

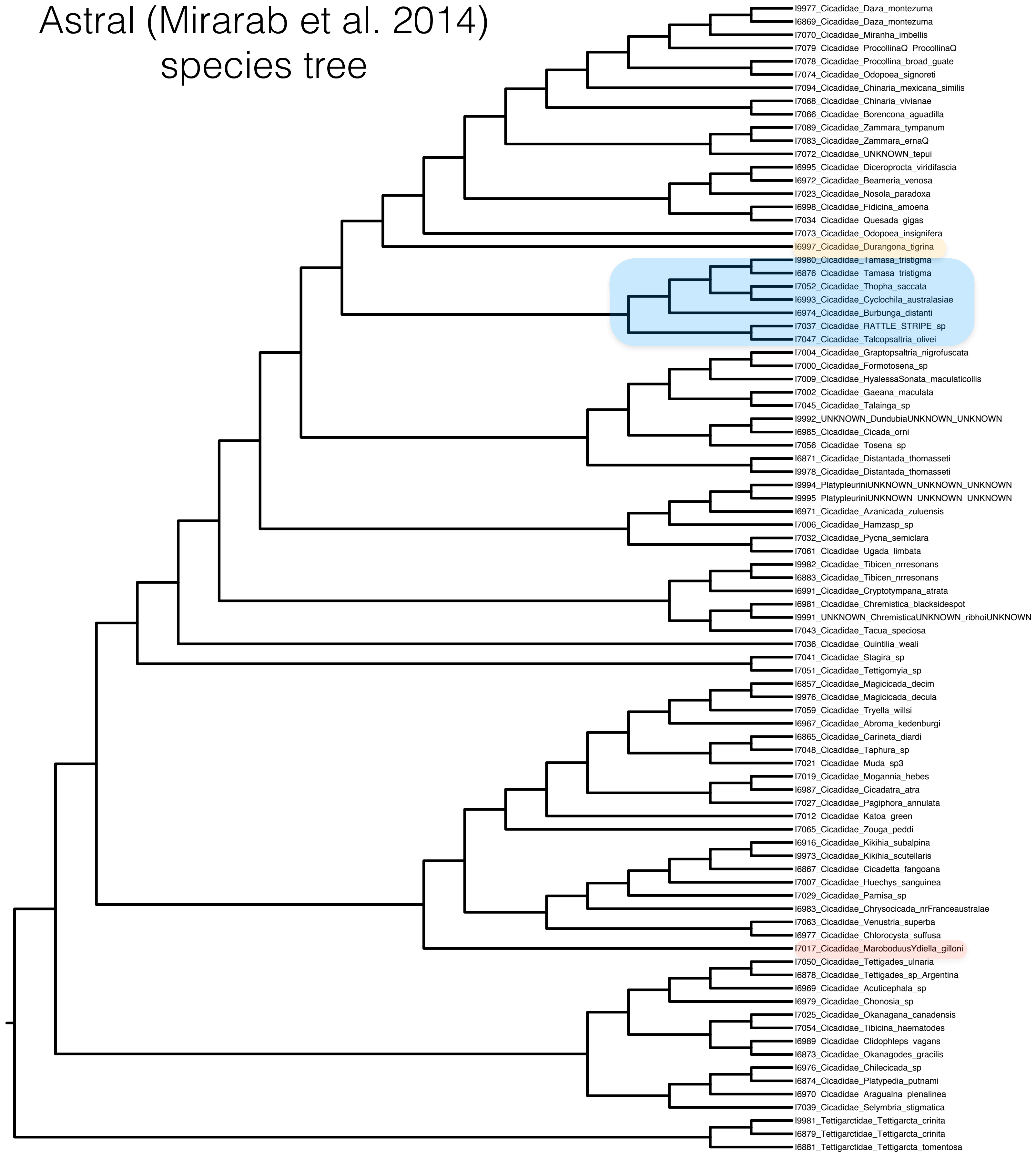
- all produce the same concat tree
- six unsupported branches still unsupported to varying degrees in addition to other branches

Bootstrap support values =
no filtering, one model /
Gblocks, partitioned/
Gblocks, K-means/
RF 12 loci rm, gene partitioned/
RF 12 loci rm, k-means partitioned

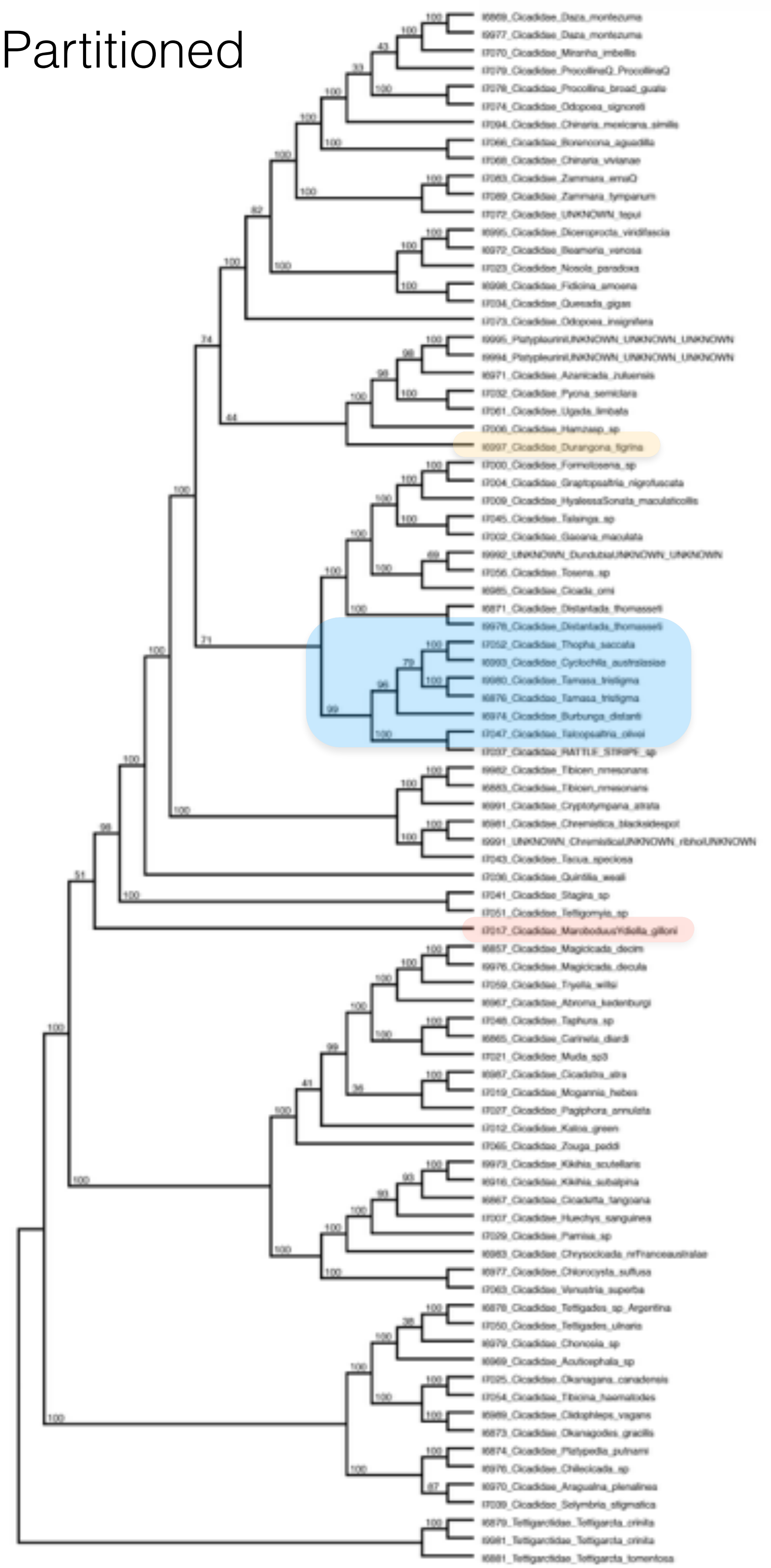
unique colors = tribes

Gene tree / species tree analysis

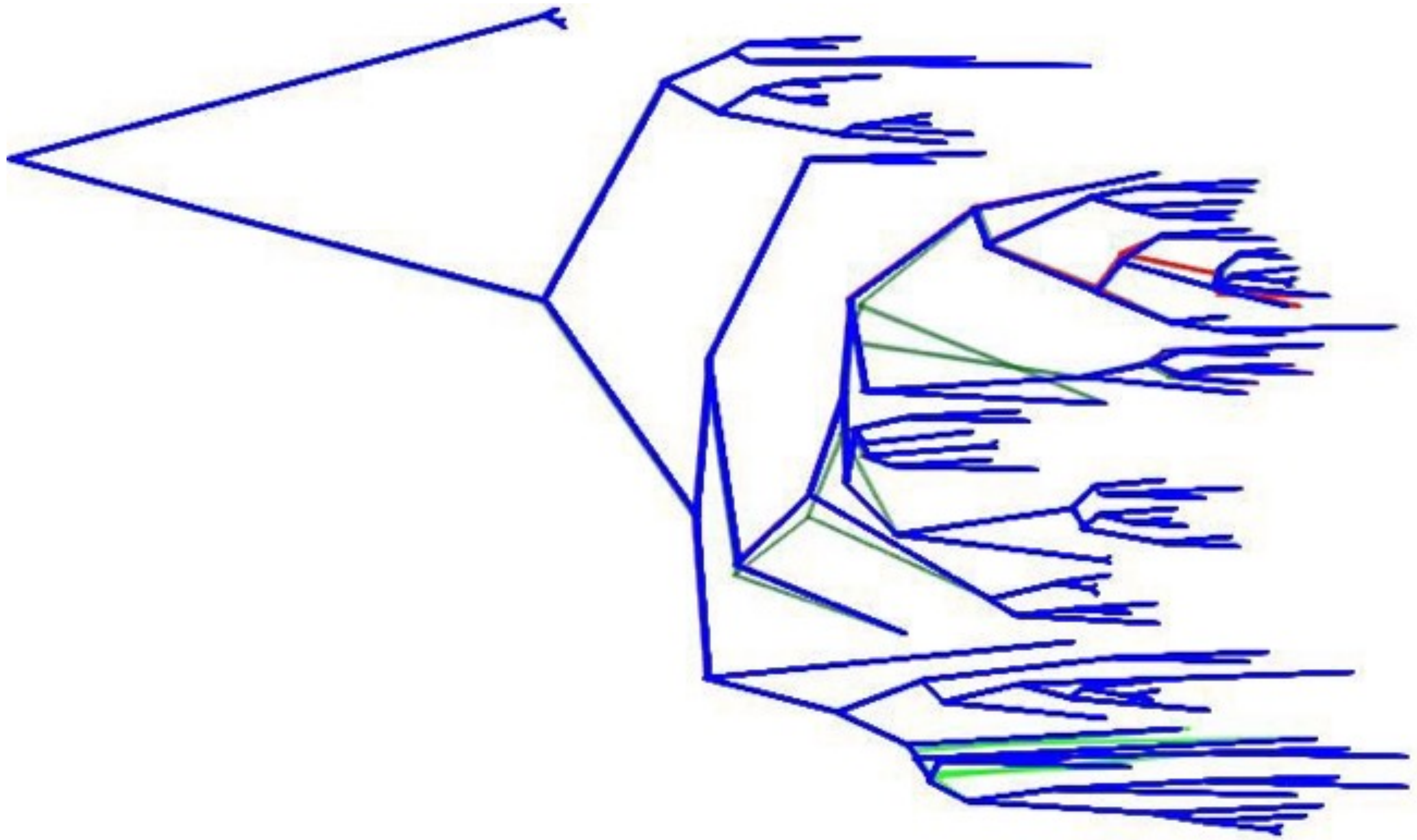
Astral (Mirarab et al. 2014) species tree


I9977_Cicadidae_Daza_montezuma
I6869_Cicadidae_Daza_montezuma
I7070_Cicadidae_Miranha_imbellis
I7079_Cicadidae_ProcollinaQ_ProcollinaQ
I7078_Cicadidae_Procollina_broad_guate
I7074_Cicadidae_Odopoea_signoreti
I7094_Cicadidae_Chinaria_mexicana_similis
I7068_Cicadidae_Chinaria_vivianae
I7066_Cicadidae_Borencona_aguadilla
I7089_Cicadidae_Zammara_tympanum
I7083_Cicadidae_Zammara_ernaQ
I7072_Cicadidae_UNKNOWN_tepui
I6995_Cicadidae_Diceroprocta_viridifascia
I6972_Cicadidae_Beameria_venosa
I7023_Cicadidae_Nosola_paradoxa
I6998_Cicadidae_Fidicina_amoena
I7034_Cicadidae_Quesada_gigas
I7073_Cicadidae_Odopoea_insignifera
I6997_Cicadidae_Durangona_tigrina
I9980_Cicadidae_Tamasa_tristigma
I6876_Cicadidae_Tamasa_tristigma
I7052_Cicadidae_Thopha_saccata
I6993_Cicadidae_Cyclochila_australasiae
I6974_Cicadidae_Burbunga_distanti
I7037_Cicadidae_RATTLE_STRIPE_sp
I7047_Cicadidae_Talcopsaltria_olivei
I7004_Cicadidae_Graptopsaltria_nigrofuscata
I7000_Cicadidae_Formotosena_sp
I7009_Cicadidae_HyalessaSonata_maculaticollis
I7002_Cicadidae_Gaeana_maculata
I7045_Cicadidae_Talainga_sp
I9992_UNKNOWN_DundubiaUNKNOWN_UNKNOWN
I6985_Cicadidae_Cicada_orni
I7056_Cicadidae_Tosena_sp
I6871_Cicadidae_Distantada_thomasseti
I9978_Cicadidae_Distantada_thomasseti
I9994_PlatypleuriniUNKNOWN_UNKNOWN_UNKNOWN
I9995_PlatypleuriniUNKNOWN_UNKNOWN_UNKNOWN
I6971_Cicadidae_Azanicada_zuluensis
I7006_Cicadidae_Hamzasp_sp
I7032_Cicadidae_Pycna_semiclara
I7061_Cicadidae_Ugada_limbata
I9982_Cicadidae_Tibicen_nrresonans
I6883_Cicadidae_Tibicen_nrresonans
I6991_Cicadidae_Cryptotympana_atrata
I6981_Cicadidae_Chremistica_blacksidespot
I9991_UNKNOWN_ChremisticaUNKNOWN_ribhoiUNKNOWN
I7043_Cicadidae_Tacua_speciosa
I7036_Cicadidae_Quintilia_weali
I7041_Cicadidae_Stagira_sp
I7051_Cicadidae_Tettigomyia_sp
I6857_Cicadidae_Magicicada_decim
I9976_Cicadidae_Magicicada_decula
I7059_Cicadidae_Tryella_willsi
I6967_Cicadidae_Abroma_kedenburgi
I6865_Cicadidae_Carineta_diardi
I7048_Cicadidae_Taphura_sp
I7021_Cicadidae_Muda_sp3
I7019_Cicadidae_Mogannia_hebes
I6987_Cicadidae_Cicadatra_atra
I7027_Cicadidae_Pagiphora_annulata
I7012_Cicadidae_Katoa_green
I7065_Cicadidae_Zouga_peddi
I6916_Cicadidae_Kikihia_subalpina
I9973_Cicadidae_Kikihia_scutellaris
I6867_Cicadidae_Cicadetta_fangoana
I7007_Cicadidae_Huechys_sanguinea
I7029_Cicadidae_Parnisa_sp
I6983_Cicadidae_Chrysocicada_nrFranceaustralae
I7063_Cicadidae_Venustria_superba
I6977_Cicadidae_Chlorocysta_suffusa
I7017_Cicadidae_MaroboduusYdiella_gilloni
I7050_Cicadidae_Tettigades_ulnaria
I6878_Cicadidae_Tettigades_sp_Argentina
I6969_Cicadidae_Acuticephala_sp
I6979_Cicadidae_Chonosia_sp
I7025_Cicadidae_Okanagana_canadensis
I7054_Cicadidae_Tibicina_haematodes
I6989_Cicadidae_Clidophleps_vagans
I6873_Cicadidae_Okanagodes_gracilis
I6976_Cicadidae_Chilecicada_sp
I6874_Cicadidae_Platypedia_putnami
I6970_Cicadidae_Aragualna_plenalinea
I7039_Cicadidae_Selymbria_stigmatica
I9981_Tettigarctidae_Tettigarcta_crinita
I6879_Tettigarctidae_Tettigarcta_crinita
I6881_Tettigarctidae_Tettigarcta_tomentosa

RAxML Partitioned


I6869_Cicadidae_Daza_montezuma
I9977_Cicadidae_Daza_montezuma
I7070_Cicadidae_Miranha_imbellis
I7079_Cicadidae_ProcollinaQ_ProcollinaQ
I7078_Cicadidae_Procollina_broad_guate
I7074_Cicadidae_Odopoea_signoreti
I7094_Cicadidae_Chinaria_mexicana_similis
I7066_Cicadidae_Borencona_aguadilla
I7068_Cicadidae_Chinaria_vivianae
I7083_Cicadidae_Zammara_ernaQ
I7089_Cicadidae_Zammara_tympanum
I7072_Cicadidae_UNKNOWN_tepui
I6995_Cicadidae_Diceroprocta_viridifascia
I6972_Cicadidae_Beameria_venosa
I7023_Cicadidae_Nosola_paradoxa
I6998_Cicadidae_Fidicina_amoena
I7034_Cicadidae_Quesada_gigas
I7073_Cicadidae_Odopoea_insignifera
I9995_PlatypleuriniUNKNOWN_UNKNOWN_UNKNOWN
I9994_PlatypleuriniUNKNOWN_UNKNOWN_UNKNOWN
I6971_Cicadidae_Azanicada_zuluensis
I7032_Cicadidae_Pycna_semiclara
I7061_Cicadidae_Ugada_limbata
I7006_Cicadidae_Hamzasp_sp
I6997_Cicadidae_Durangona_tigrina
I7000_Cicadidae_Formotosena_sp
I7004_Cicadidae_Graptopsaltria_nigrofuscata
I7009_Cicadidae_HyalessaSonata_maculaticollis
I7045_Cicadidae_Talainga_sp
I7002_Cicadidae_Gaeana_maculata
I9992_UNKNOWN_DundubiaUNKNOWN_UNKNOWN
I7056_Cicadidae_Tosena_sp
I6985_Cicadidae_Cicada_orni
I6871_Cicadidae_Distantada_thomasseti
I9978_Cicadidae_Distantada_thomasseti
I7052_Cicadidae_Thopha_saccata
I6993_Cicadidae_Cyclochila_australasiae
I9980_Cicadidae_Tamasa_tristigma
I6876_Cicadidae_Tamasa_tristigma
I6974_Cicadidae_Burbunga_distanti
I7047_Cicadidae_Talcopsaltria_olivei
I7037_Cicadidae_RATTLE_STRIPE_sp
I9982_Cicadidae_Tibicen_nrresonans
I6883_Cicadidae_Tibicen_nrresonans
I6991_Cicadidae_Cryptotympana_atrata
I6981_Cicadidae_Chremistica_blacksidespot
I9991_UNKNOWN_ChremisticaUNKNOWN_ribhoiUNKNOWN
I7043_Cicadidae_Tacua_speciosa
I7036_Cicadidae_Quintilia_weali
I7041_Cicadidae_Stagira_sp
I7051_Cicadidae_Tettigomyia_sp
I7017_Cicadidae_MaroboduusYdiella_gilloni
I6857_Cicadidae_Magicicada_decim
I9976_Cicadidae_Magicicada_decula
I7059_Cicadidae_Tryella_willsi
I6967_Cicadidae_Abroma_kedenburgi
I7048_Cicadidae_Taphura_sp
I6865_Cicadidae_Carineta_diardi
I7021_Cicadidae_Muda_sp3
I6987_Cicadidae_Cicadatra_atra
I7019_Cicadidae_Mogannia_hebes
I7027_Cicadidae_Pagiphora_annulata
I7012_Cicadidae_Katoa_green
I7065_Cicadidae_Zouga_peddi
I9973_Cicadidae_Kikihia_scutellaris
I6916_Cicadidae_Kikihia_subalpina
I6867_Cicadidae_Cicadetta_fangoana
I7007_Cicadidae_Huechys_sanguinea
I7029_Cicadidae_Parnisa_sp
I6983_Cicadidae_Chrysocicada_nrFranceaustralae
I6977_Cicadidae_Chlorocysta_suffusa
I7063_Cicadidae_Venustria_superba
I6878_Cicadidae_Tettigades_sp_Argentina
I7050_Cicadidae_Tettigades_ulnaria
I6979_Cicadidae_Chonosia_sp
I6969_Cicadidae_Acuticephala_sp
I7025_Cicadidae_Okanagana_canadensis
I7054_Cicadidae_Tibicina_haematodes
I6989_Cicadidae_Clidophleps_vagans
I6873_Cicadidae_Okanagodes_gracilis
I6874_Cicadidae_Platypedia_putnami
I6976_Cicadidae_Chilecicada_sp
I6970_Cicadidae_Aragualna_plenalinea
I7039_Cicadidae_Selymbria_stigmatica
I6879_Tettigarctidae_Tettigarcta_crinita
I9981_Tettigarctidae_Tettigarcta_crinita
I6881_Tettigarctidae_Tettigarcta_tomentosa

Exploring tree space: highly divergent starting trees



- 100 random starting trees sharing no bipartitions from ML tree
- RAxML gene partitioned analyses

Conclusions

1. Most of the tree is supported, but some rogue taxa and clades

2. Higher taxonomy needs to be reexamined in light of well-supported relationships

3. Additional sensitivity analyses needed

Questions?