Phylogenomic estimation of the Hemiptera phylogeny and the co-diversification with plants

Christopher L. Owen
Postdoctoral Researcher
George Washington University
Computational Biology Institute



Outline

Introduction

Review previous Hemiptera phylogeny estimates

Methods

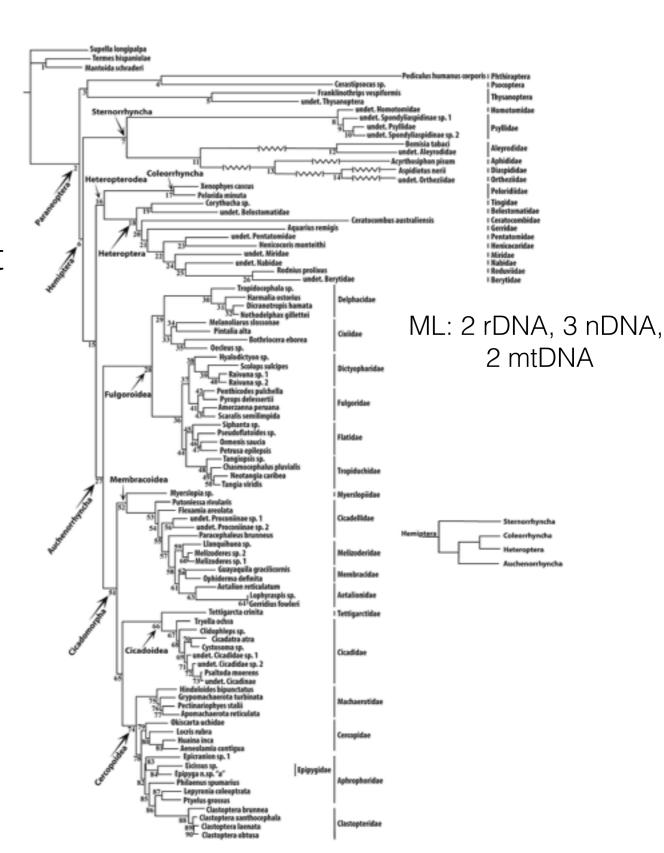
- Dataset assembly and curation
- Preliminary dating and diversification analyses

Results

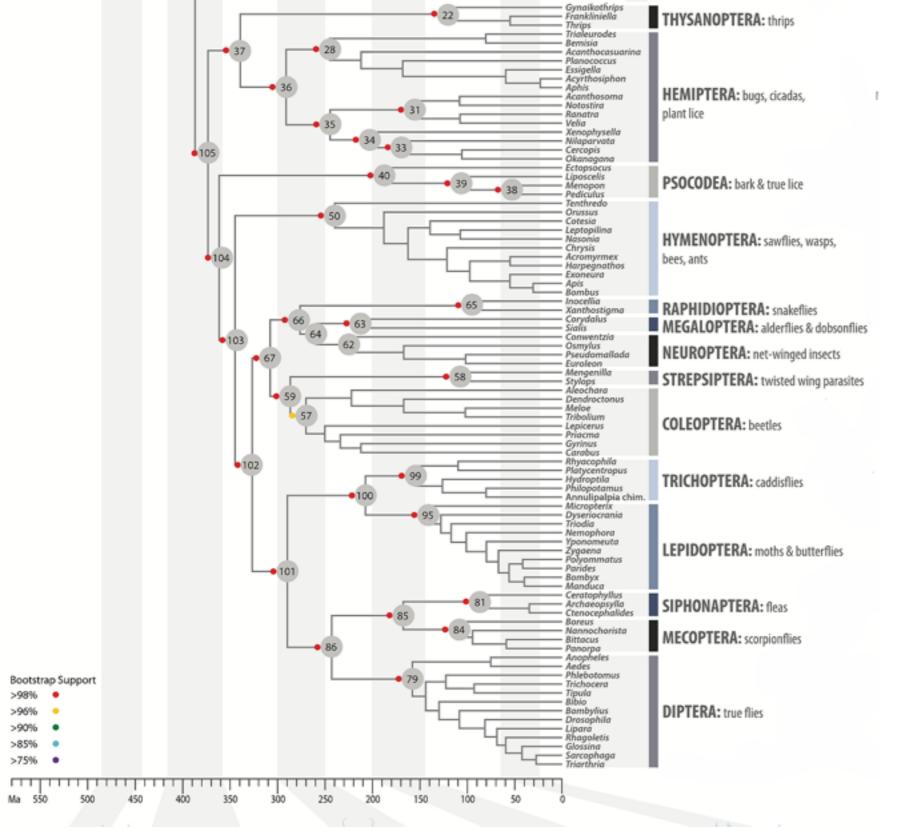
- Phylogeny of Hemiptera
- Gene tree/species tree discordance/concordance
- Preliminary chronogram
- Preliminary diversification results

Recent Hemiptera Phylogenies: Cryan and Urban 2012 J. Syst. Ent.

- Well-supported backbone for most higher taxa
- Lacking support for Hemiptera
- Lacking support for Paraneoptera
- Molecular evidence for long branches in Sternorrhyncha



Recent Hemiptera Phylogenies: Misof et al. 2014 Science



- 1,478 orthologs
- Lots of curation and analyses!
- 100 BS for
 Psocodea sister to
 Holometabola;
 poor support
 quartet mapping

Methods

Transcriptome assembly, ortholog identification, and phylogenetic analyses

Transcriptome assembly

- 57 Paraneoptera transcriptomes and NCBI ESTs
- Transcriptomes assembled from raw NGS reads
- Quality control->Error correction->Trinity

Ortholog Identification

- OrthoDB: R. prolixus, P. humanus, A. pisum
- HaMStR pipeline

Alignments, trimming, phylogenetic analyses

Mafft->GBlocks->Prottest3->PhyloBayes & RAxML->Decisiveness

Divergence time estimation and diversification rates

treePL -> Medusa

Results

Pediculus humanus 4099 Rhodnius prolixus 4099 Acyrthosiphon pisum 4099 Triatoma rubida Triatoma infestans 3860 Trialeurodes vaporariorum 3425 Tibicen_tibicen 999 Thrips_tabaci 3948 Sogatella fucifera Sitobion avenae 3825 Schizaphis_graminum 3210 Riptortus pedestris 232 Rhopalosiphum_padi 602 Popplepsalta sp 2666 1424 Platypedia_putnami Planococcus citri Philaenus spumarius strict 767 Peregrinus maidis 688 Pemphigus_spyrothecae 320 Pachypsylla_venusta 3620 Oncopeltus fasciatus 2846 471 Oncometopia_nigricans Nilaparvata lugens 3283 Myzus_persicae 311 Megalurothrips_sjostedti 1492 Magicicada_septendecim 124 Macrosiphum euphorbiae Maconellicoccus hirsutus 281 Lygus_lineolaris 3416 Lygus_hesperus Lygaeus_kalamii 3757 Liposcelis_bostrychophila 3967 Laodelphax_striatella 773 Kerria lacca 3890 Homalodisca_vitripennis 409 Graphocephala atropunctata 109 Graminella nigrifrons 3867 Gerris buenoi 2445 Frankliniella tritici 3921 Frankliniella occidentalis 3502 Ericerus_pela 3374 Echinothrips_americanus 3960 1575 Diaphorina_citri Dialeurodes citri 2958 Clavigralla tomentosicollis 3224 Cimex_lectularis 2718 1026 Cacopsylla_pruni Boisea trivittata 3319 Bemisia_tabaci 3843 Bactericera_cockerelli 3728 Atrapsalta_sp 3704 Arma chinensis 3945 3931 Apolygus lucorum Aphis nerii 3958 2508 Aphis gossypii Aphis glycines 2815 Anoplocnemis curvipes 1947 500 1000 Single-copy orthologs

Taxon & Ortholog Sampling

Outgroups

Thysanoptera: 5 speciesPhthiraptera: 1 speciesPsocoptera: 1 species

Auchenorrhyncha

Cicadoidea: 5 speciesCercopoidea: 1 species

Membracoidea: 4 species

• Fulgoroidea: 4 species

<u>Heteroptera</u>

Pentatomoidea: 1 species

Coreoidea: 4 species

Lygaeoidea: 2 species

Cimicoidea: 1 species

Miroidea: 3 species

Reduvioidea: 3 species

• Gerroidea: 1 species

Sternorrhyncha

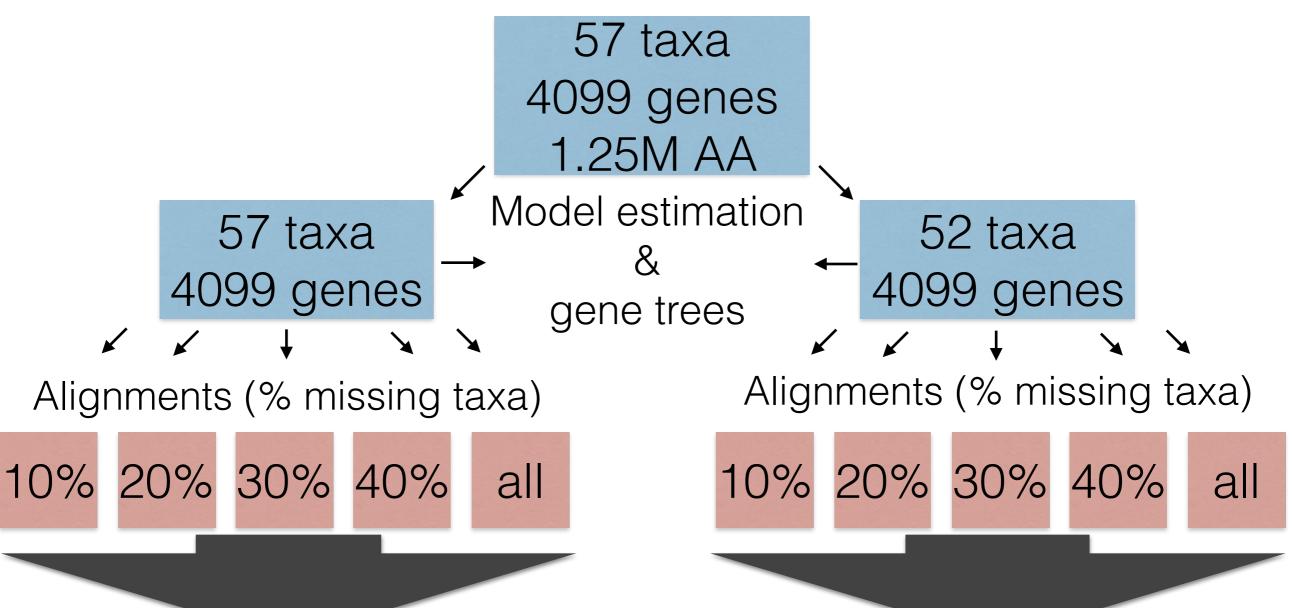
Coccoidea: 4 species

Aphidoidea: 10 species

• Psylloidea: 4 species

• Aleyrodidae: 3 species

Datasets and gene removal: "Choosing the best genes for the job" (Collins 2005)



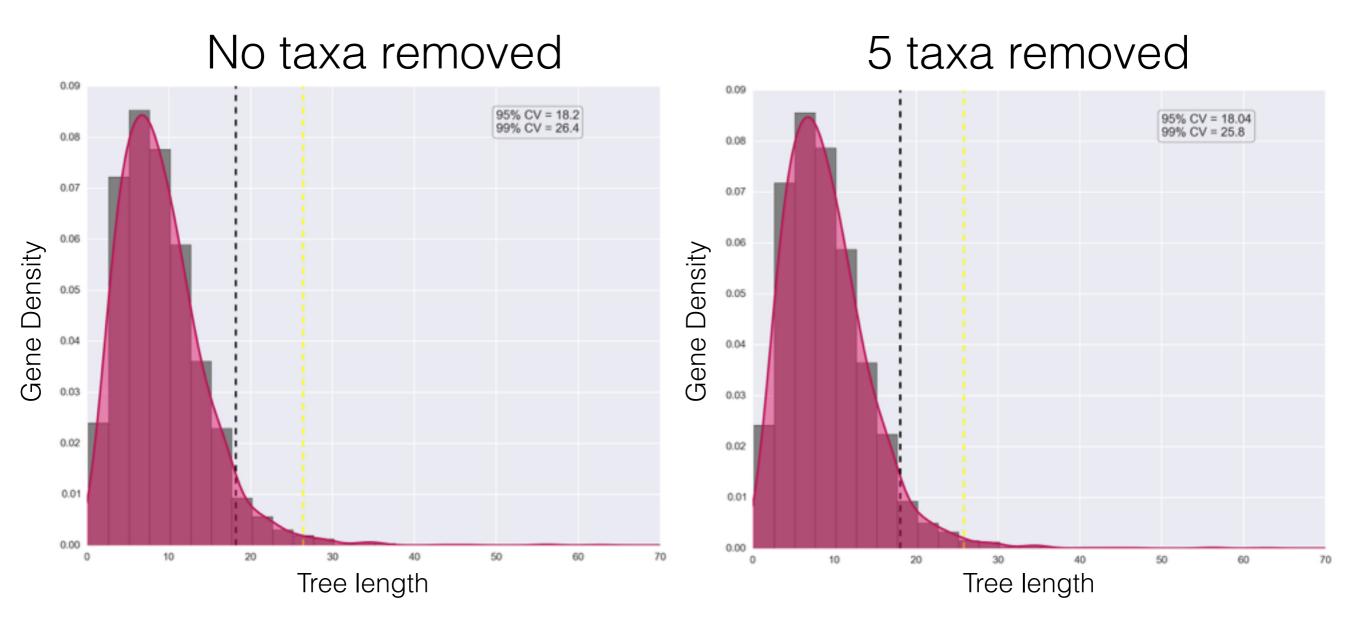
Choosing the best genes:

- 1. Rate of evolution
- 2. Among site rate variation
- 3. Compositional bias
- 4. Long branches

Choosing the best genes:

- Rate of evolution
- 2. Among site rate variation
- 3. Compositional bias
- 4. Long branches

Choosing the best genes: rate of evolution



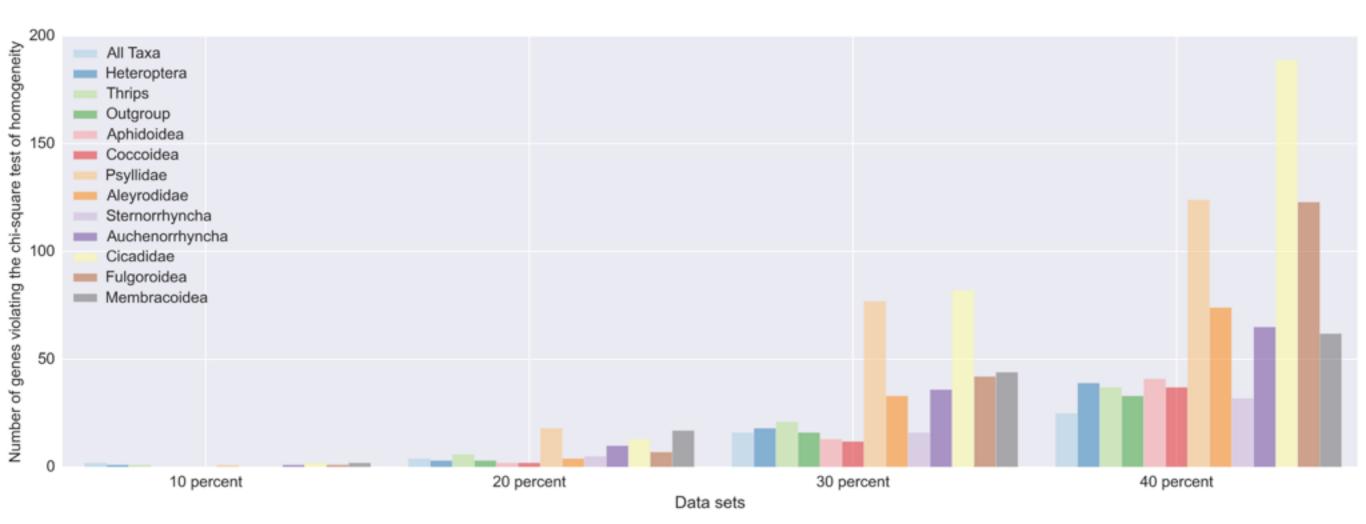
- Excluded genes with high rates of evolution (Oakley et al. 2013 MBE; Philippe et al. 2005)
- Arbitrary cutoff of TL 10

Choosing the best genes: among site rate variation



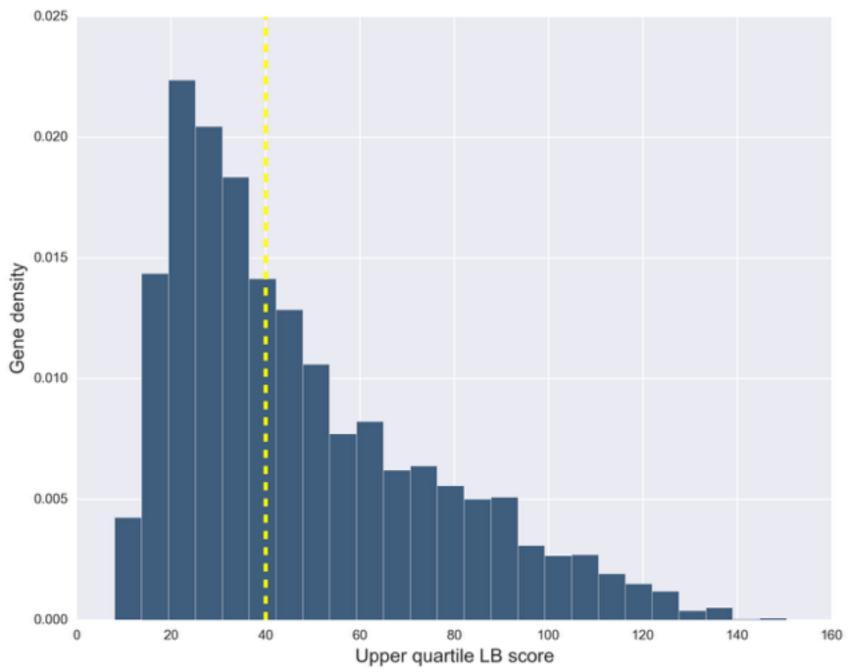
removed genes alpha value < 1 (Yang 1994)

Choosing the best genes: compositional bias



- Chi-square test among taxa and higher taxa (p<0.5)
- suffers from type 2 error (Foster 2004)
- Foster 2004; Galtier & Gouy 1995; Lockhart et al. 1992

Choosing the best genes: long branches

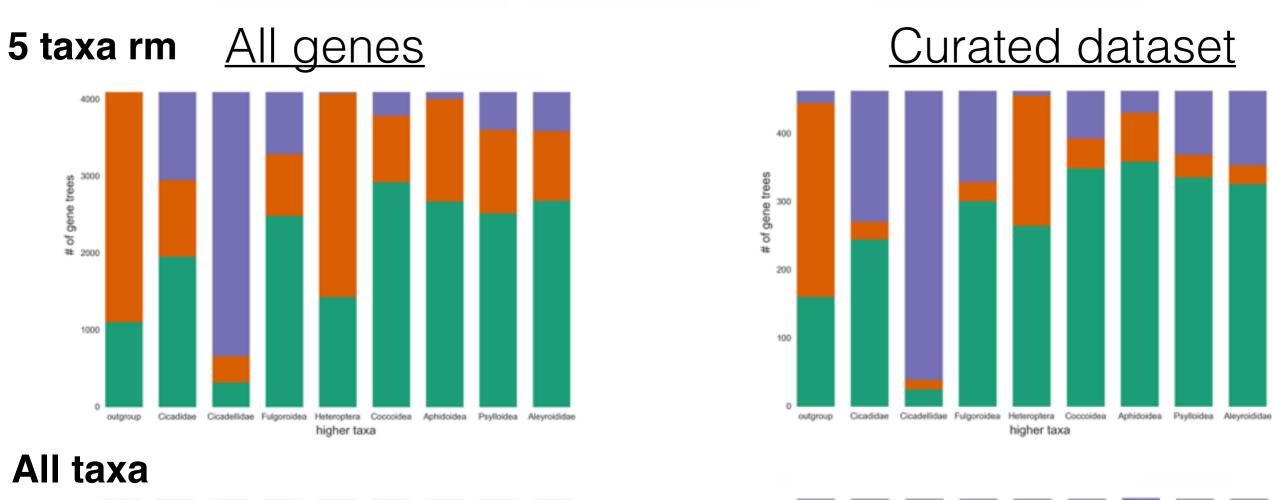


$$LB_i = \left(\frac{\overline{PD^i}}{\overline{PD_a}} - 1\right) * 100$$

- LB score upper quartile of each partition
- unrooted gene trees
- Brinkman & Philippe 2008; Bergsten 2005

Comparison of monophyly among datasets

non-monophyletic



coccoidea aphidoidea psylloidea aleyroididae

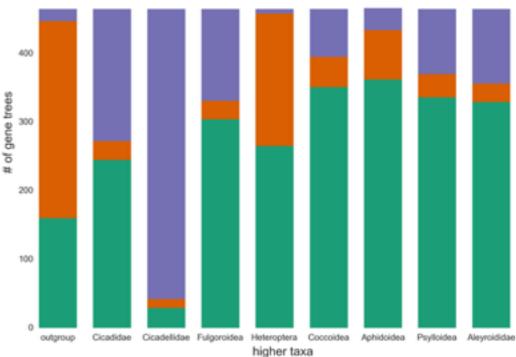
monophyletic

of gene trees

1000

cicadellidae fulgoroidea

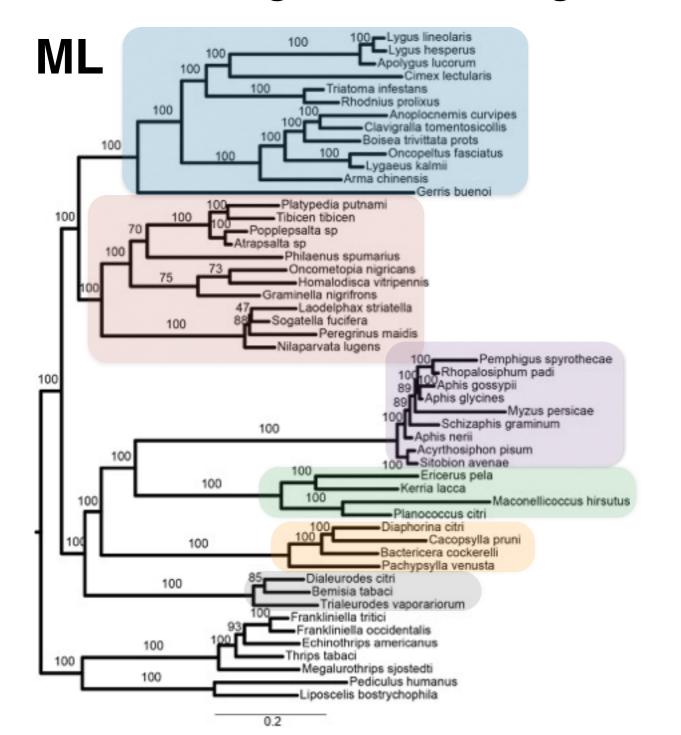
higher taxa

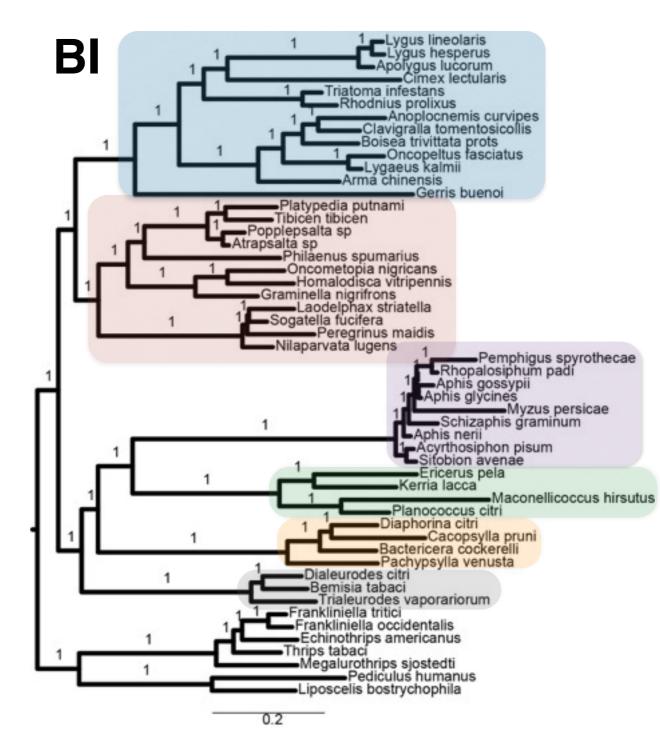


single sp.

Hemiptera phylogeny: most curated dataset

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

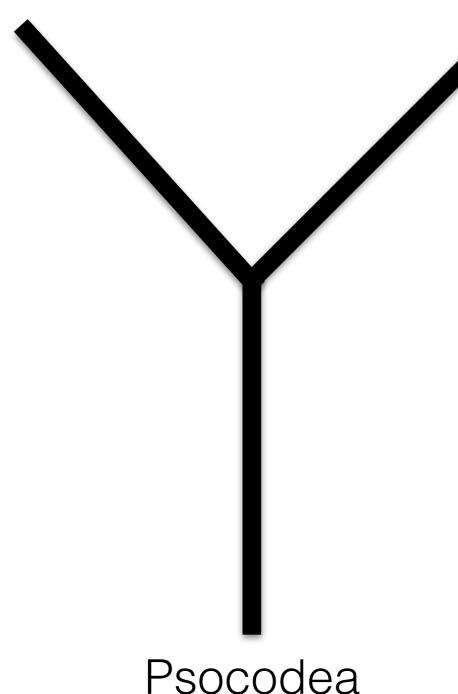




Monophyly of Paraneoptera

Thysanoptera

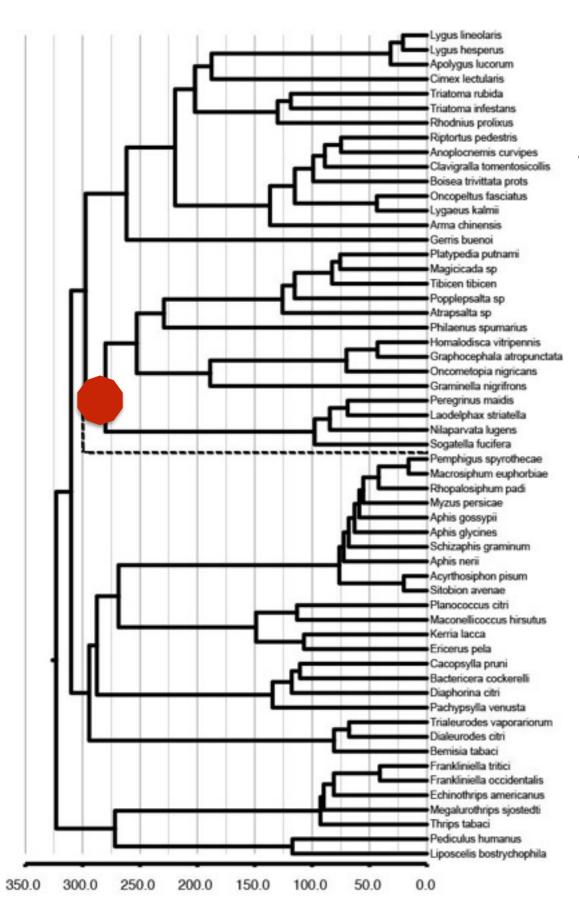
Hemiptera



Dataset: 52 taxa; 463 genes

- 149 gene trees all three lineages are monophyletic
- 3/149 Psocodea more closely related to Thysanoptera
- 146/149 Hemiptera more closely related to Thysanoptera
- Distance from Psocodea to Hemiptera ~2x greater than the distance from Thysanoptera to Hemiptera

Divergence times and diversification



Divergence times: treePL

- calibrations: Li et al. 2012; Misof et al. 2014
- a little older than Misof et al. 2014
- need to validate fossil congruence

Diversification: Medusa

- Fit BD models to nodes
- compare models with AIC
- identified one change in diversification rate leading to Coleorrhyncha

Hemiptera phylogeny and diversification: future direction

