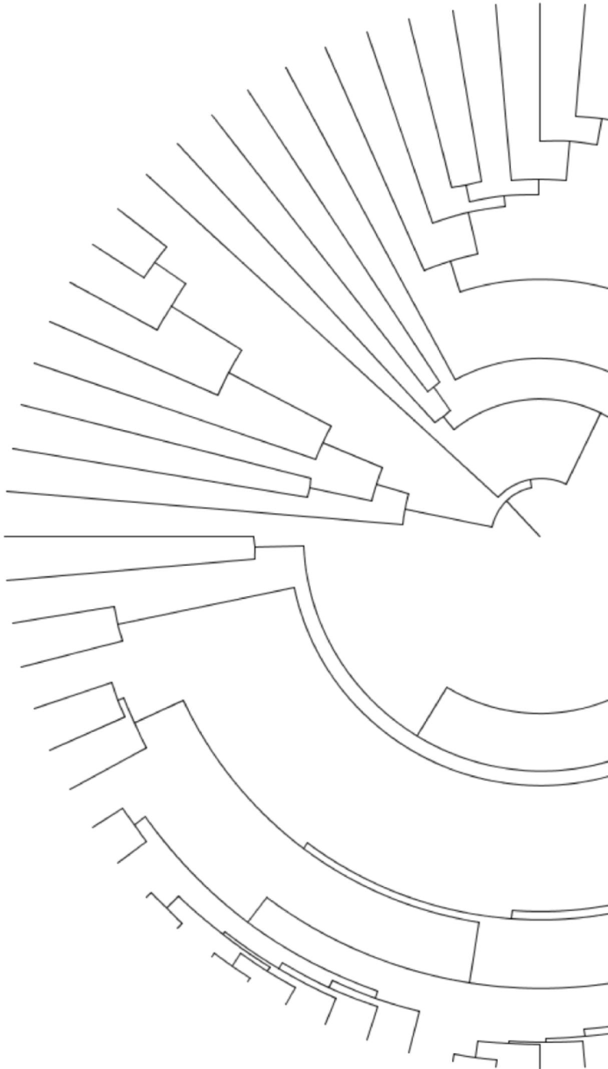
Large-scale gene family analysis of 76 Arthropods

i5K webinar / September 5, 2018



Gregg Thomas

@greggwcthomas

Indiana University

The genomic basis of Arthropod diversity

https://www.biorxiv.org/content/early/2018/08/04/382945

Why and how did we do it?

76 species

Tetranychus urticae	Trombidiformes
Metaseiulus occidentalis	Mesostigmata
Ixodes scapularis	Ixodida
Centruroides sculpturatus	Scorpiones
Loxosceles reclusa	Araneae
Stegodyphus mimosarum	
Latrodectus hesperus	
Parasteatoda tepidariorum	
Strigamia maritima	Geophilomorpha
Hyalella azteca	Amphipoda
Eurytemora affinis	Calanoida
Daphnia pulex	Cladocera
Catajapyx aquilonaris	Diplura
Ladona fulva	Odonata
Ephemera danica	Ephemeroptera
Blatella germanica	Blattodea
Zootermopsis nevadensis	Isoptera
Pediculus humanus	Phthiraptera
Frankliniella occidentalis	Thysanoptera
Acyrthosiphon pisum	Hemiptera
Pachypsylla venusta	
Homalodisca vitripennis	
Gerris buenoi	
Cimex lectularius	
Halyomorpha halys	
Oncopeltus fasciatus	
Athalia rosae	Hymenoptea
Cephus cinctus	
Orussus abietinus	
Nasonia vitripennis	
Copidosoma floridanum	
Trichogramma pretiosum	
Harpegnathos saltator	
Linepithema humile	
Camponotus floridanus	
Pogonomyrmex barbatus	
Cardiocondyla obscurior	
Solenopsis invicta	
Atta cephalotes	
Acromyrmex echinatior	
Dufourea novaeangliae	
Lasioglossum albipes	
Megachile rotundata	
Habropoda laboriosa	
Eufriesea mexicana	
Apis mellifera	
Apis florea	
Melipona quadrifasciata	
Bombus impatiens	
Bombus terrestris	
Agrilus planipennis	Coleoptera
Onthophagus taurus	
Tribolium castaneum	
Dendroctonus ponderosae	
Anoplophora glabripennis	
Leptinotarsa decemlineata	
Limnephilus lunatus	Trichoptera
Plutella xylostella	Lepidoptera
Bombyx mori	
Manduca sexta	
Heliconius melpomene	
Danaus plexippus	
Aedes aegypti	Diptera
Culex quinquefasciatus	
Anopheles albimanus	
Anopheles gambiae	
Anopheles funestus	
Lutzomyia longipalpis	
Mayetiola destructor	
Drosophila grimshawi	
Drosophila pseudoobscura	
Drosophila melanogaster	
Ceratitis capitata	
Glossina morsitans	
Lucilia cuprina	
Musca domestica	

21 orders

The genomic basis of Arthropod diversity

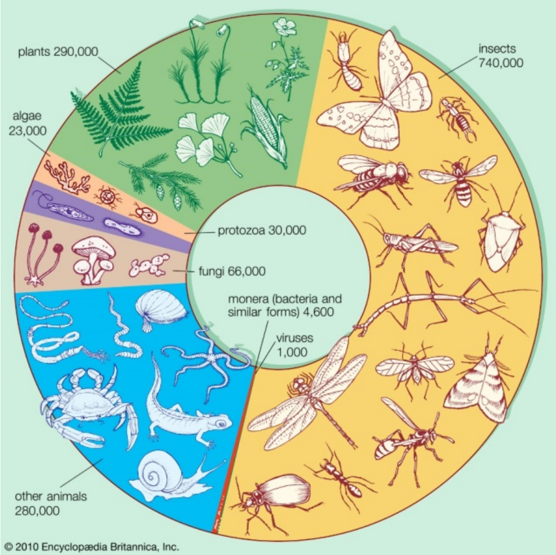
https://www.biorxiv.org/content/early/2018/08/04/382945

Why and how did we do it?

76 species

Tetranychus urticae	Trombidiformes
Metaseiulus occidentalis	Mesostigmata
Ixodes scapularis	Ixodida
Centruroides sculpturatus	Scorpiones
Loxosceles reclusa	Araneae
Stegodyphus mimosarum	
Latrodectus hesperus	
Parasteatoda tepidariorum	
Strigamia maritima	Geophilomorpha
Hyalella azteca	Amphipoda
Eurytemora affinis	Calanoida
Daphnia pulex	Cladocera
Catajapyx aquilonaris	Diplura
Ladona fulva	Odonata
Ephemera danica	Ephemeroptera
Blatella germanica	Blattodea
Zootermopsis nevadensis	Isoptera
Pediculus humanus	Phthiraptera
Frankliniella occidentalis	Thysanoptera
Acyrthosiphon pisum	Hemiptera
Pachypsylla venusta	
Homalodisca vitripennis	
Gerris buenoi	
Cimex lectularius	
Halyomorpha halys	
Oncopeltus fasciatus	
Athalia rosae	Hymenoptea
Cephus cinctus	
Orussus abietinus	
Nasonia vitripennis	
Copidosoma floridanum	
Trichogramma pretiosum	
Harpegnathos saltator	
Linepithema humile	
Camponotus floridanus	
Pogonomyrmex barbatus	
Cardiocondyla obscurior	
Solenopsis invicta	
Atta cephalotes	
Acromyrmex echinatior	
Dufourea novaeangliae	
Lasioglossum albipes	
Megachile rotundata	
Habropoda laboriosa	
Eufriesea mexicana	
Apis mellifera	
Apis florea	
Melipona quadrifasciata	
Bombus impatiens	
Bombus terrestris	
Agrilus planipennis	Coleoptera
Onthophagus taurus	
Tribolium castaneum	
Dendroctonus ponderosae	
Anoplophora glabripennis	
Leptinotarsa decemlineata	
Limnephilus lunatus	Trichoptera
Plutella xylostella	Lepidoptera
Bombyx mori	
Manduca sexta	
Heliconius melpomene	
Danaus plexippus	
Aedes aegypti	Diptera
Culex quinquefasciatus	
Anopheles albimanus	
Anopheles gambiae	
Anopheles funestus	
Lutzomyia longipalpis	
Mayetiola destructor	
Drosophila grimshawi	
Drosophila pseudoobscura	
Drosophila melanogaster	
Ceratitis capitata	
Glossina morsitans	
Lucilia cuprina	
Musca domestica	

21 orders



© 2010 Encyclopædia Britannica, Inc.

The genomic basis of Arthropod diversity

https://www.biorxiv.org/content/early/2018/08/04/382945

Why and how did we do it?

76 species

21 orders

Tetranychus urticae	Trombidiformes
Metaseiulus occidentalis	Mesostigmata
Ixodes scapularis	Ixodida
Centruroides sculpturatus	Scorpiones
Loxosceles reclusa	Araneae
Stegodyphus mimosarum	
Latrodectus hesperus	
Parasteatoda tepidariorum	
Strigamia maritima	Geophilomorpha
Hyalella azteca	Amphipoda
Eurytemora affinis	Calanoida
Daphnia pulex	Cladocera
Catajapyx aquilonaris	Diplura
Ladona fulva	Odonata
Ephemera danica	Ephemeroptera
Blatella germanica	Blattodea
Zootermopsis nevadensis	Isoptera
Pediculus humanus	Phthiraptera
Frankliniella occidentalis	Thysanoptera
Acyrthosiphon pisum	Hemiptera
Pachypsylla venusta	
Homalodisca vitripennis	
Gerris buenoi	
Cimex lectularius	
Halyomorpha halys	
Oncopeltus fasciatus	
Athalia rosae	Hymenoptea
Cephus cinctus	
Orussus abietinus	
Nasonia vitripennis	
Copidosoma floridanum	
Trichogramma pretiosum	
Harpegnathos saltator	
Linepithema humile	
Camponotus floridanus	
Pogonomyrmex barbatus	
Cardiocondyla obscurior	
Solenopsis invicta	
Atta cephalotes	
Acromyrmex echinatior	
Dufourea novaeangliae	
Lasioglossum albipes	
Megachile rotundata	
Habropoda laboriosa	
Eufriesea mexicana	
Apis mellifera	
Apis florea	
Melipona quadrifasciata	
Bombus impatiens	
Bombus terrestris	
Agrilus planipennis	Coleoptera
Onthophagus taurus	
Tribolium castaneum	
Dendroctonus ponderosae	
Anoplophora glabripennis	
Leptinotarsa decemlineata	
Limnephilus lunatus	Trichoptera
Plutella xylostella	Lepidoptera
Bombyx mori	
Manduca sexta	
Heliconius melpomene	
Danaus plexippus	
Aedes aegypti	Diptera
Culex quinquefasciatus	
Anopheles albimanus	
Anopheles gambiae	
Anopheles funestus	
Lutzomyia longipalpis	
Mayetiola destructor	
Drosophila grimshawi	
Drosophila pseudoobscura	
Drosophila melanogaster	
Ceratitis capitata	
Glossina morsitans	
Lucilia cuprina	
Musca domestica	

FiveThirtyEight

Politics Sports **Science & Health** Economics Culture

MAY 2, 2017 AT 10:00 AM

The Bugs Of The World Could Squish Us All

And we'd deserve it.

By Maggie Koerth-Baker

Filed under Science Question From A Toddler

© 2010 Encyclopædia Britannica, Inc.

https://fivethirtyeight.com/features/the-bugs-of-the-world-could-squish-us-all/

Arthropods exhibit vast phenotypic diversity

Aedes aegypti
(Diptera)

Bombus terrestris
(Hymenoptera)

Latrodectus hesperus
(Araneae)

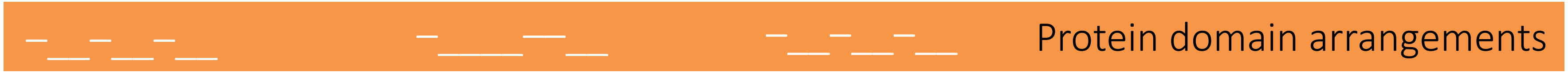
Whole-genome sequencing reveals vast molecular differences

...AAGGCCA...	...AAGTCCA...	...AAGTCCA...	Nucleotide substitutions
Aedes aegypti (Diptera)	*Bombus terrestris* (Hymenoptera)	*Latrodectus hesperus* (Araneae)	

Whole-genome sequencing reveals vast molecular differences

4	4	2	Gene copy number variation
...AAGGCCA...	...AAGTCCA...	...AAGTCCA...	Nucleotide substitutions
Aedes aegypti (Diptera)	*Bombus terrestris* (Hymenoptera)	*Latrodectus hesperus* (Araneae)	

Whole-genome sequencing reveals vast molecular differences



			Protein domain arrangements
4	4	2	Gene copy number variation
...AAGGCCA...	...AAGTCCA...	...AAGTCCA...	Nucleotide substitutions
Aedes aegypti (Diptera)	*Bombus terrestris* (Hymenoptera)	*Latrodectus hesperus* (Araneae)	

Whole-genome sequencing reveals vast molecular differences

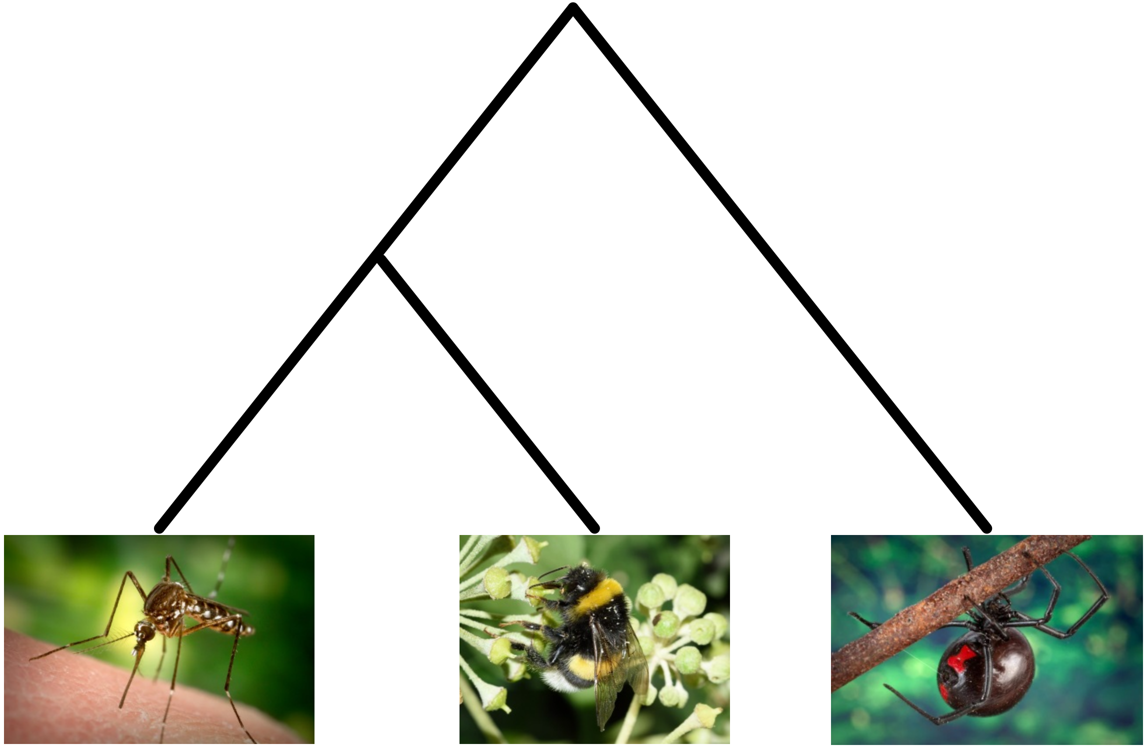
–__–__–__	–_____—__	–__–__–__	Protein domain arrangements
4	4	2	Gene copy number variation
...AAGGCCA...	...AAGTCCA...	...AAGTCCA...	Nucleotide substitutions
Aedes aegypti (Diptera)	*Bombus terrestris* (Hymenoptera)	*Latrodectus hesperus* (Araneae)	

Whole-genome sequencing reveals vast molecular differences

‾__‾__‾__	‾_____‾__	‾__‾__‾__	Protein domain arrangements
4	4	2	Gene copy number variation
...AAGGCCA...	...AAGTCCA...	...AAGTCCA...	Nucleotide substitutions
Aedes aegypti (Diptera)	*Bombus terrestris* (Hymenoptera)	*Latrodectus hesperus* (Araneae)	

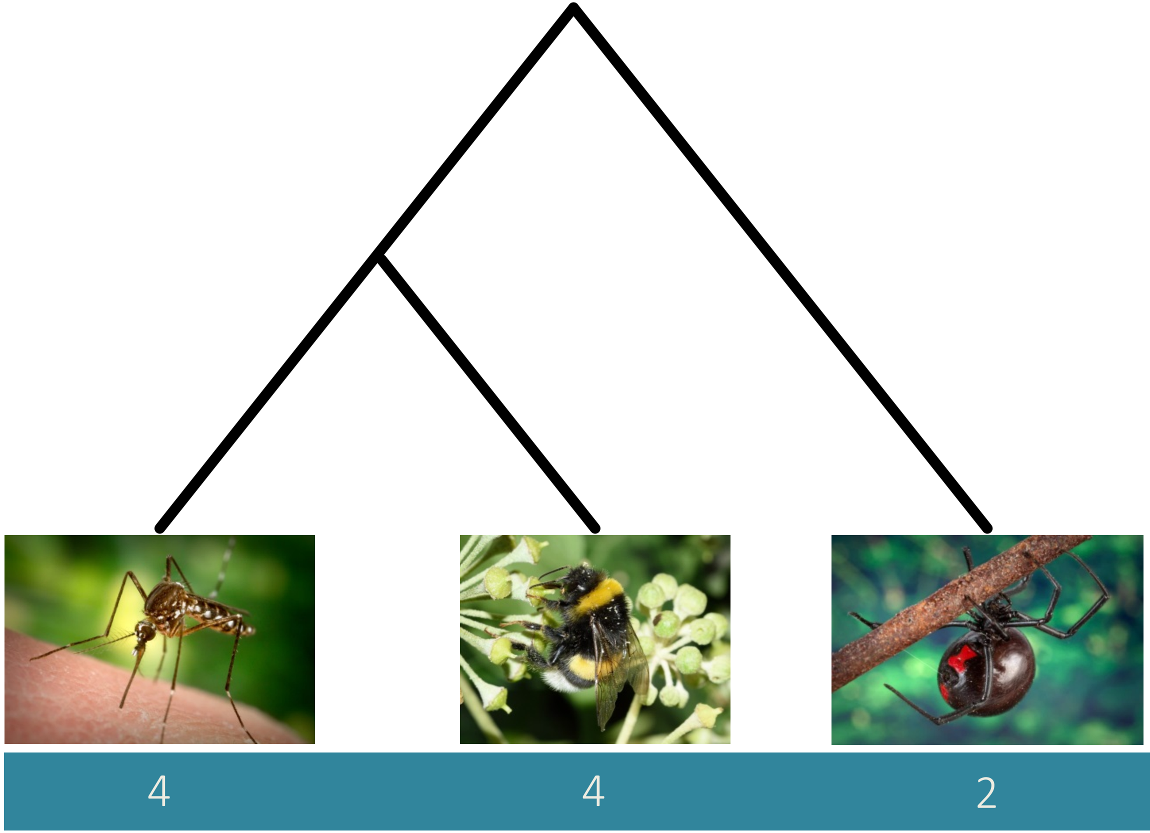
How can we begin to understand what changes are interesting or important?

Phylogenies act as a framework for asking these types of questions



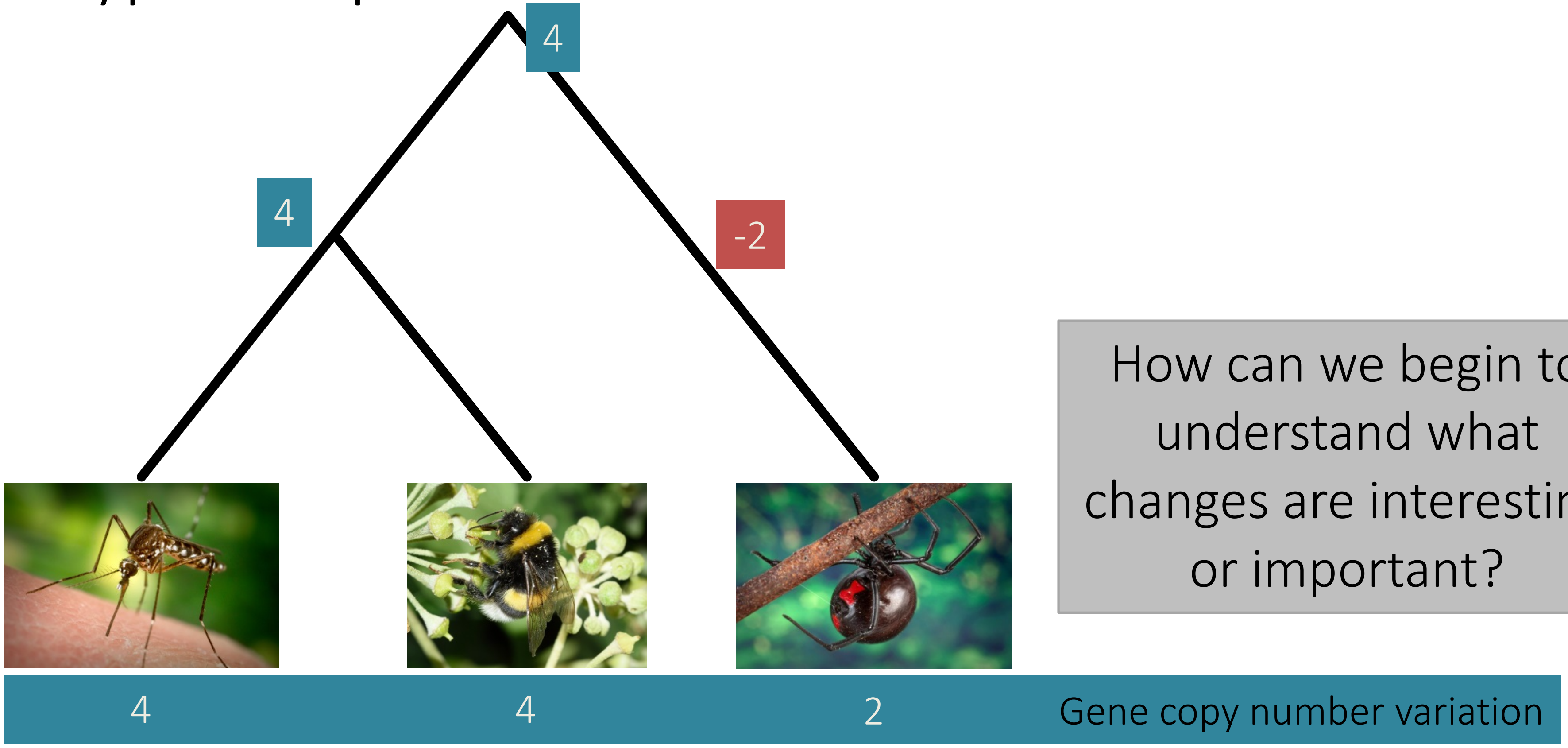
How can we begin to understand what changes are interesting or important?

Phylogenies act as a framework for asking these types of questions



How can we begin to understand what changes are interesting or important?

Phylogenies act as a framework for asking these types of questions



How can we begin to understand what changes are interesting or important?

Which molecular changes lead to interesting phenotypic differences?

1. Sequence and annotate many genomes

Stephen Richards
Monica Poelchau

Which molecular changes lead to interesting phenotypic differences?

1. Sequence and annotate many genomes

Stephen Richards
Monica Poelchau

2. Determine orthology of sequences

Rob Waterhouse
Evgeny Zdobnov
Panagiotis Ioannidis

Which molecular changes lead to interesting phenotypic differences?

1. Sequence and annotate many genomes
2. Determine orthology of sequences
3. Determine phylogeny of species

Stephen Richards
Monica Poelchau

Rob Waterhouse
Evgeny Zdobnov
Panagiotis Ioannidis

Which molecular changes lead to interesting phenotypic differences?

1. Sequence and annotate many genomes
2. Determine orthology of sequences
3. Determine phylogeny of species
4. Map orthology onto phylogeny to reconstruct the evolutionary history of all loci

Stephen Richards
Monica Poelchau

Rob Waterhouse
Evgeny Zdobnov
Panagiotis Ioannidis

Elias Dohmen
Karl Glastad
Yiyuan Li

Which molecular changes lead to interesting phenotypic differences?

1. Sequence and annotate many genomes
2. Determine orthology of sequences
3. Determine phylogeny of species
4. Map orthology onto phylogeny to reconstruct the evolutionary history of all loci

Today's topics

1. Determining the Arthropod phylogeny

2. Reconstructing ancestral gene counts

3. Using the i5k gene family web site

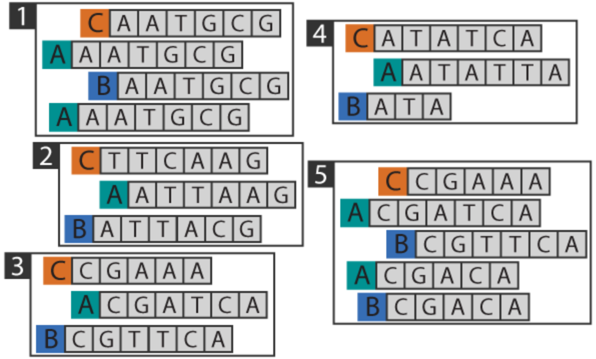
Today's topics

1. Determining the Arthropod phylogeny

2. Reconstructing ancestral gene counts

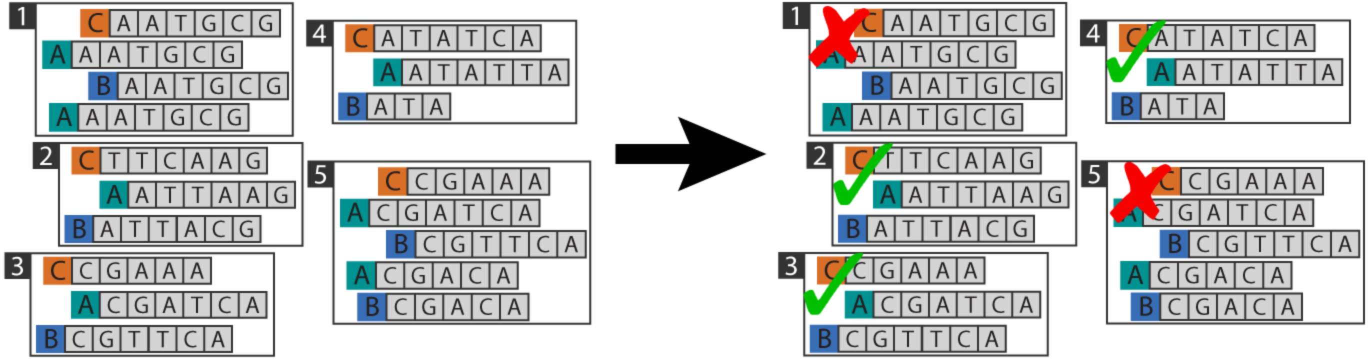
3. Using the i5k gene family web site

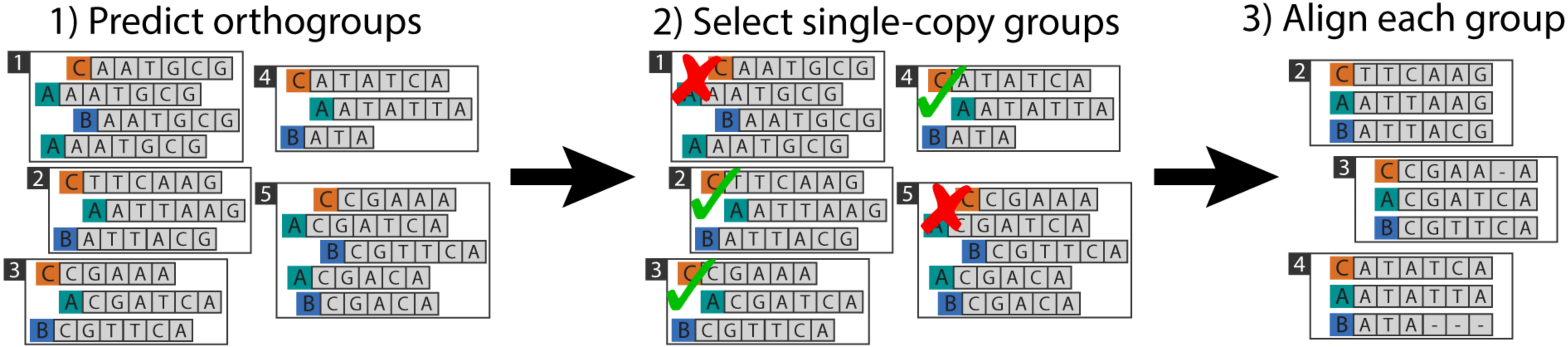
1) Predict orthogroups

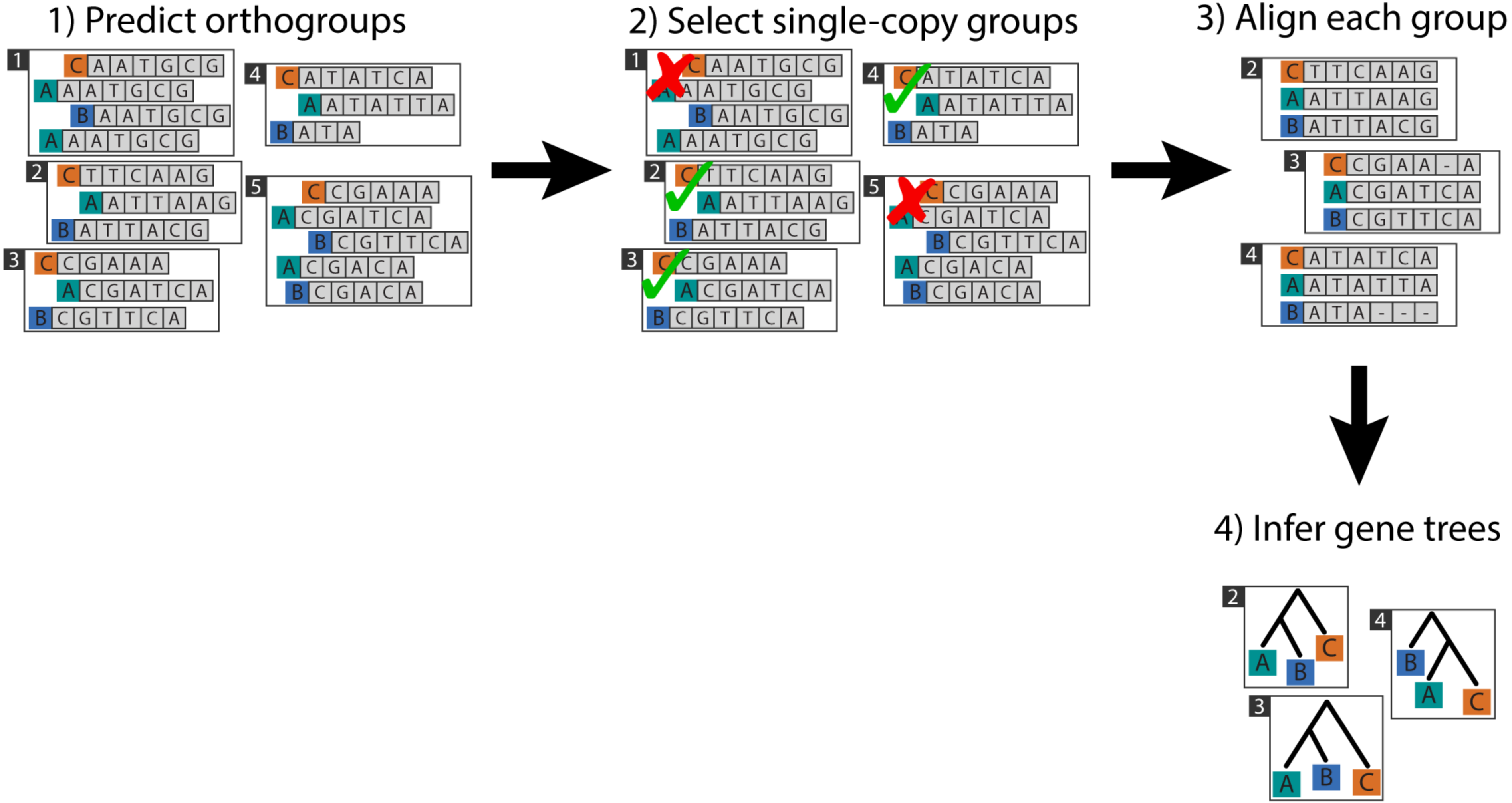


1) Predict orthogroups

2) Select single-copy groups




1) Predict orthogroups
1
C A A T G C G
A A A T G C G
B A A T G C G
A A A T G C G
4
C A T A T C A
A A T A T T A
B A T A
2
C T T C A A G
A A T T A A G
B A T T A C G
5
C C G A A A
A C G A T C A
B C G T T C A
A C G A C A
B C G A C A
3
C C G A A A
A C G A T C A
B C G T T C A
2) Select single-copy groups
1
C A A T G C G
A A A T G C G
B A A T G C G
A A A T G C G
4
C A T A T C A
A A T A T T A
B A T A
2
C T T C A A G
A A T T A A G
B A T T A C G
5
C C G A A A
A C G A T C A
B C G T T C A
A C G A C A
B C G A C A
3
C C G A A A
A C G A T C A
B C G T T C A
3) Align each group
2
C T T C A A G
A A T T A A G
B A T T A C G
3
C C G A A - A
A C G A T C A
B C G T T C A
4
C A T A T C A
A A T A T T A
B A T A - - -

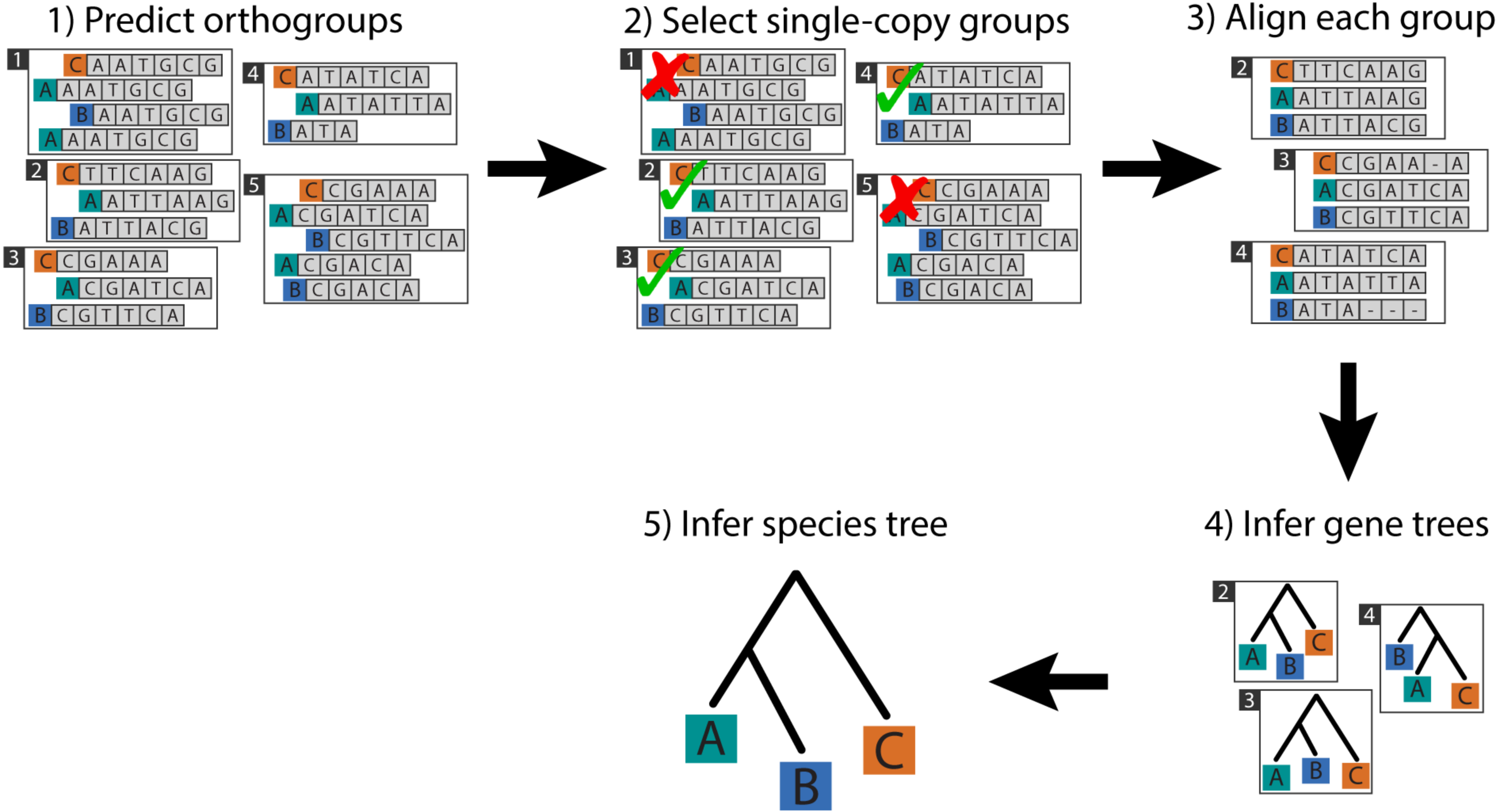


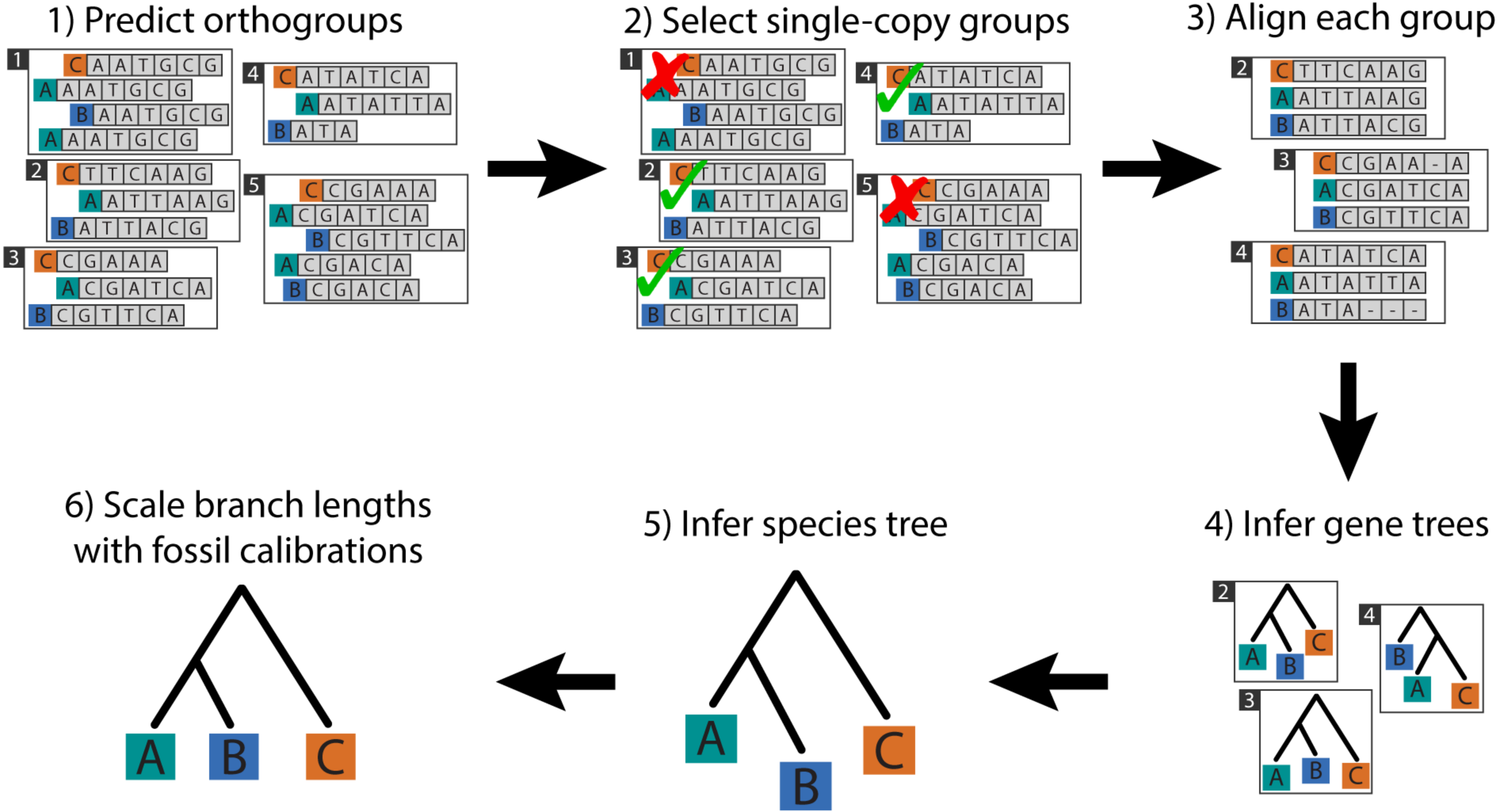
1) Predict orthogroups

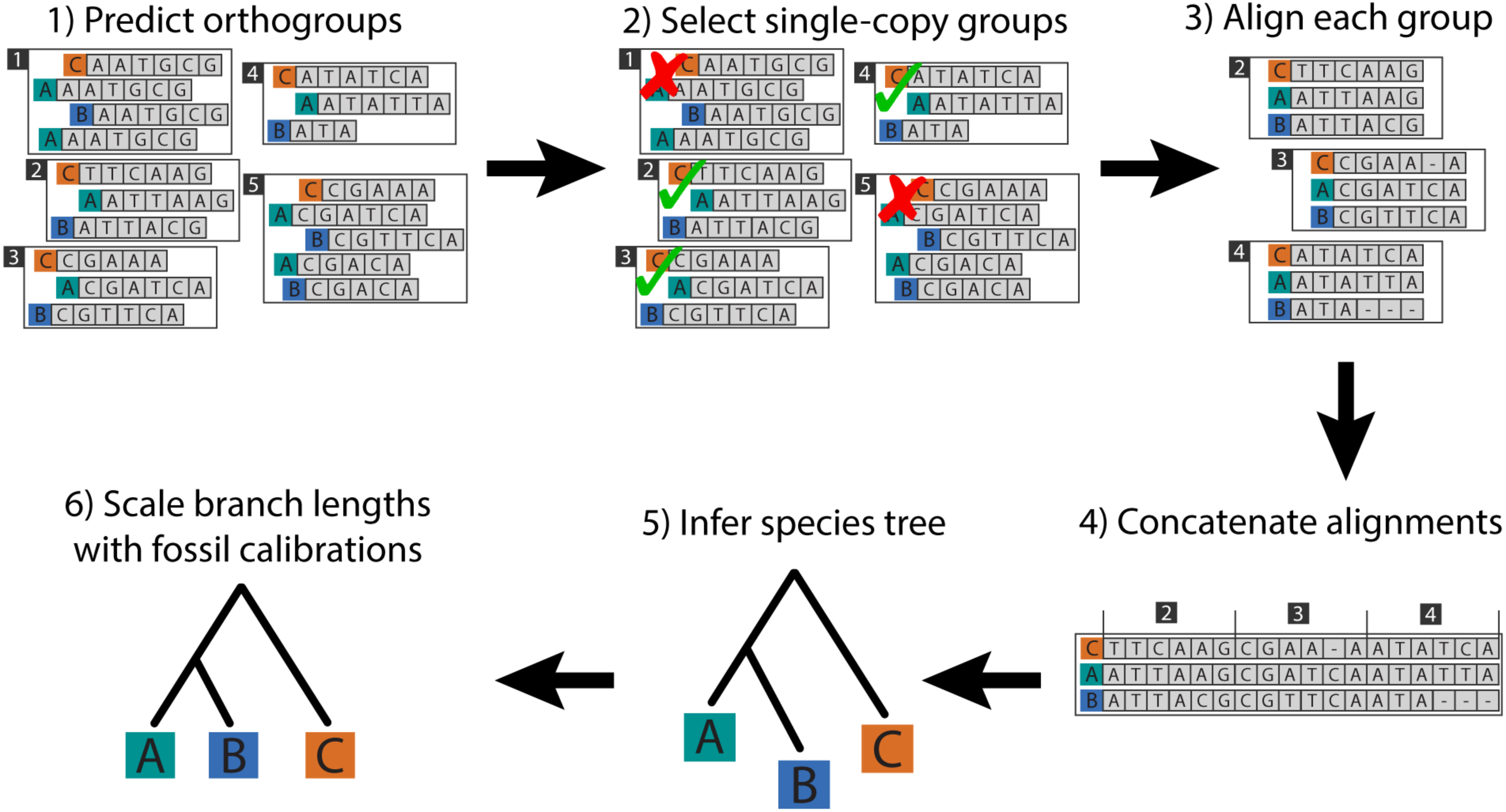
2) Select single-copy groups

3) Align each group

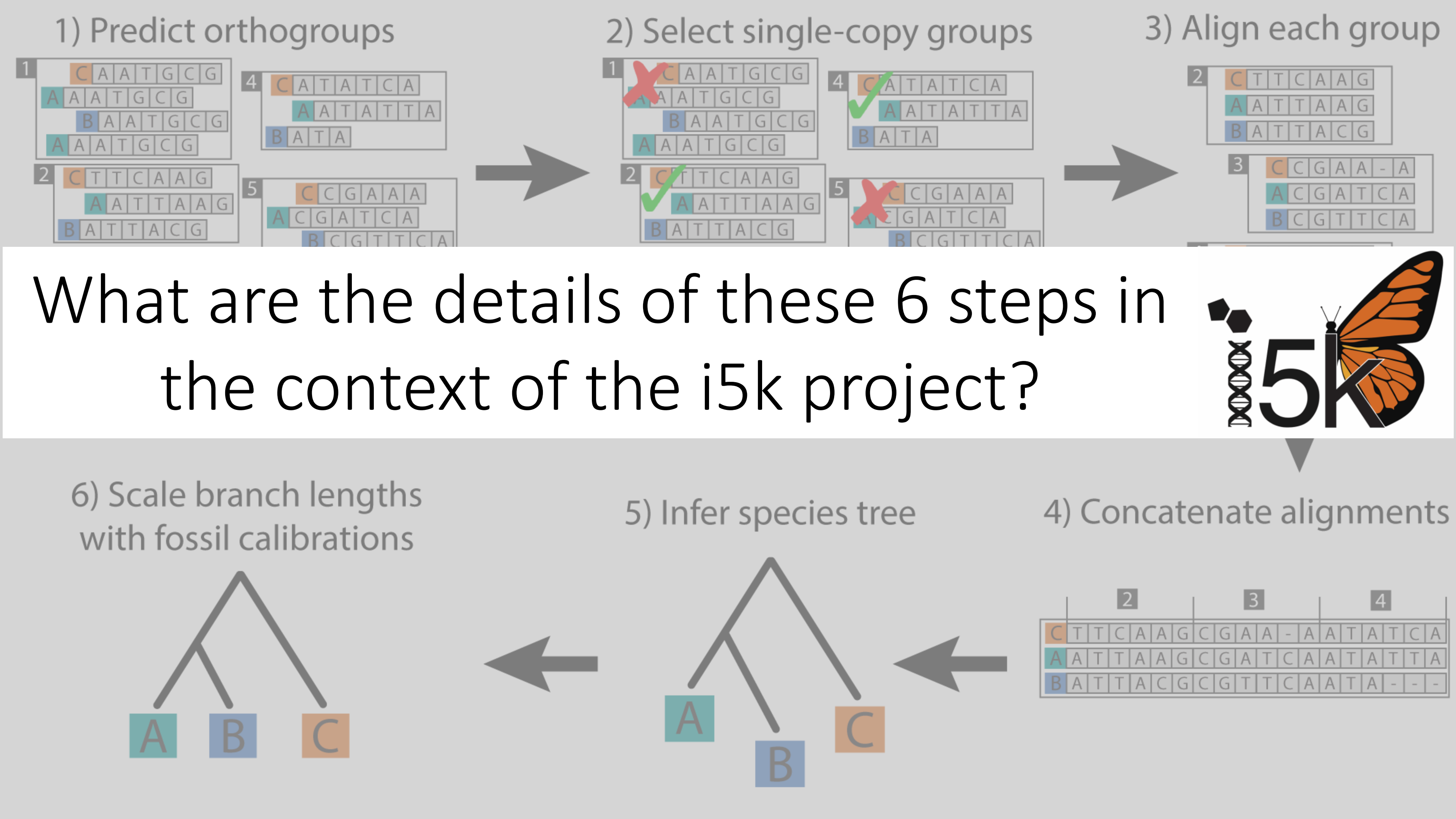
4) Infer gene trees


1) Predict orthogroups
1 C AATGCG / A AATGCG / B AATGCG / A AATGCG
2 C TTCAAG / A ATTAAG / B ATTACG
3 C CGAAA / A CGATCA / B CGTTCA
4 C ATATCA / A ATATTA / B ATA
5 C CGAAA / A CGATCA / B CGTTCA / A CGACA / B CGACA
2) Select single-copy groups
1 ✗
2 ✓
3 ✓
4 ✓
5 ✗
3) Align each group
2 C TTCAAG / A ATTAAG / B ATTACG
3 C CGAA-A / A CGATCA / B CGTTCA
4 C ATATCA / A ATATTA / B ATA---
4) Infer gene trees
2 A B C
3 A B C
4 B A C
5) Infer species tree
A B C

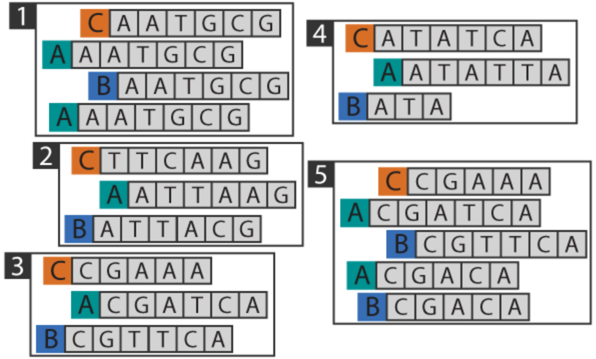

1) Predict orthogroups
2) Select single-copy groups
3) Align each group
4) Infer gene trees
5) Infer species tree
6) Scale branch lengths with fossil calibrations
1
2
3
4
5
A
B
C


1) Predict orthogroups
2) Select single-copy groups
3) Align each group
4) Concatenate alignments
5) Infer species tree
6) Scale branch lengths with fossil calibrations
1
2
3
4
5
A
B
C

What are the details of these 6 steps in the context of the i5k project?

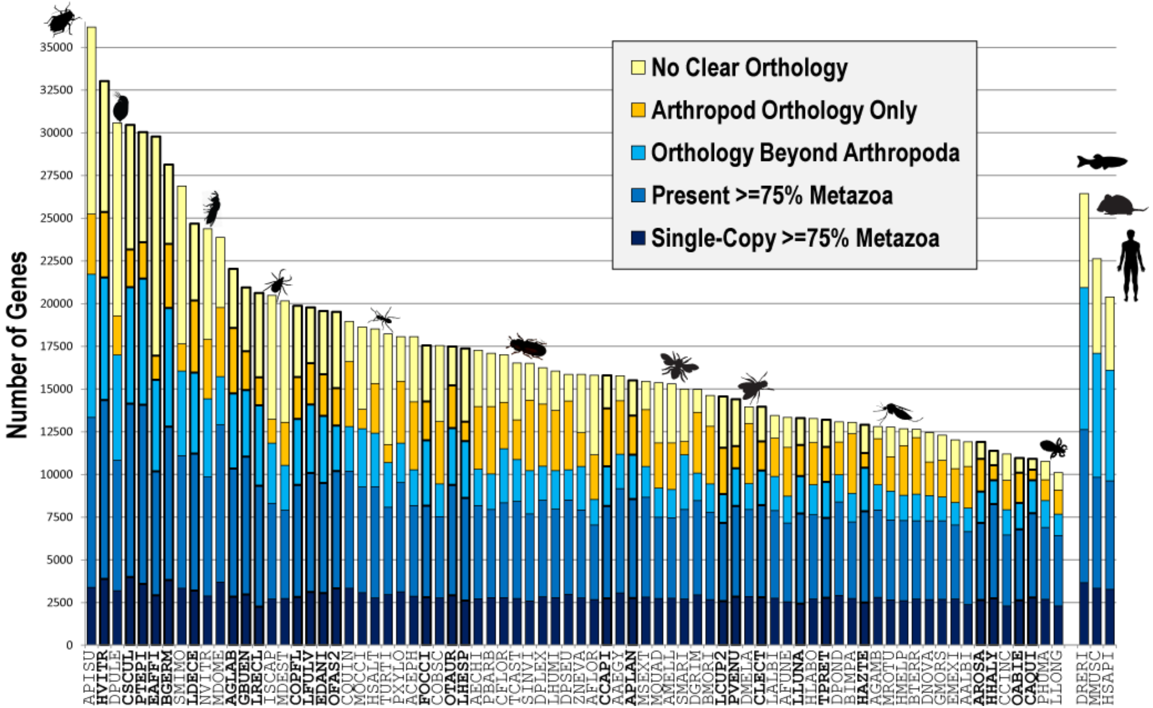

1) Predict orthogroups
2) Select single-copy groups
3) Align each group
4) Concatenate alignments
5) Infer species tree
6) Scale branch lengths with fossil calibrations
1
2
3
4
5
A
B
C
C T T C A A G C G A A - A A T A T C A
A A T T A A G C G A T C A A T A T T A
B A T T A C G C G T T C A A T A - - -

1) Predict orthogroups



1) Predict orthogroups – OrthoDB

Rob Waterhouse
Evgeny Zdobnov
Panagiotis Ioannidis

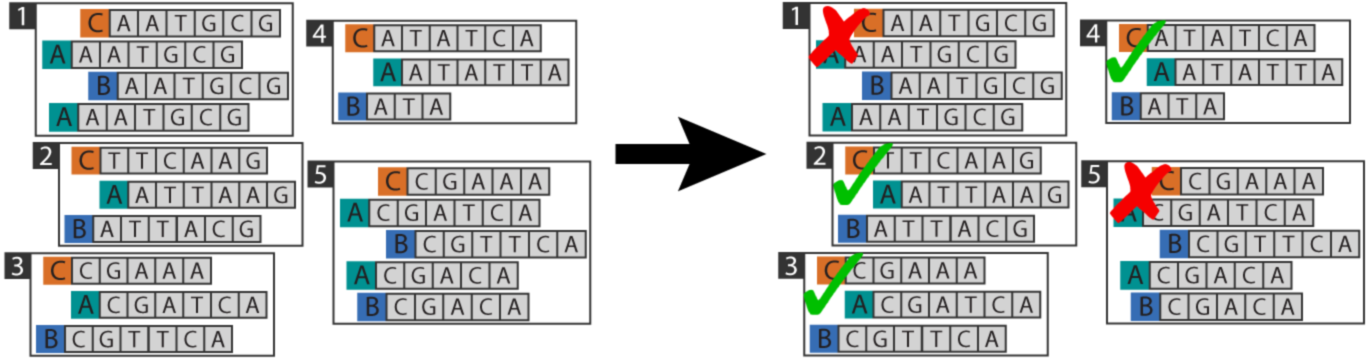


38,195 ortho-groups across 76 arthropod species

(See i5k webinar from Feb. 1, 2017: http://i5k.github.io/webinar)

1) Predict orthogroups

2) Select single-copy groups



2) Select single-copy orthogroups

How many single-copy orthologs in our 38,195 groups?

2) Select single-copy orthogroups

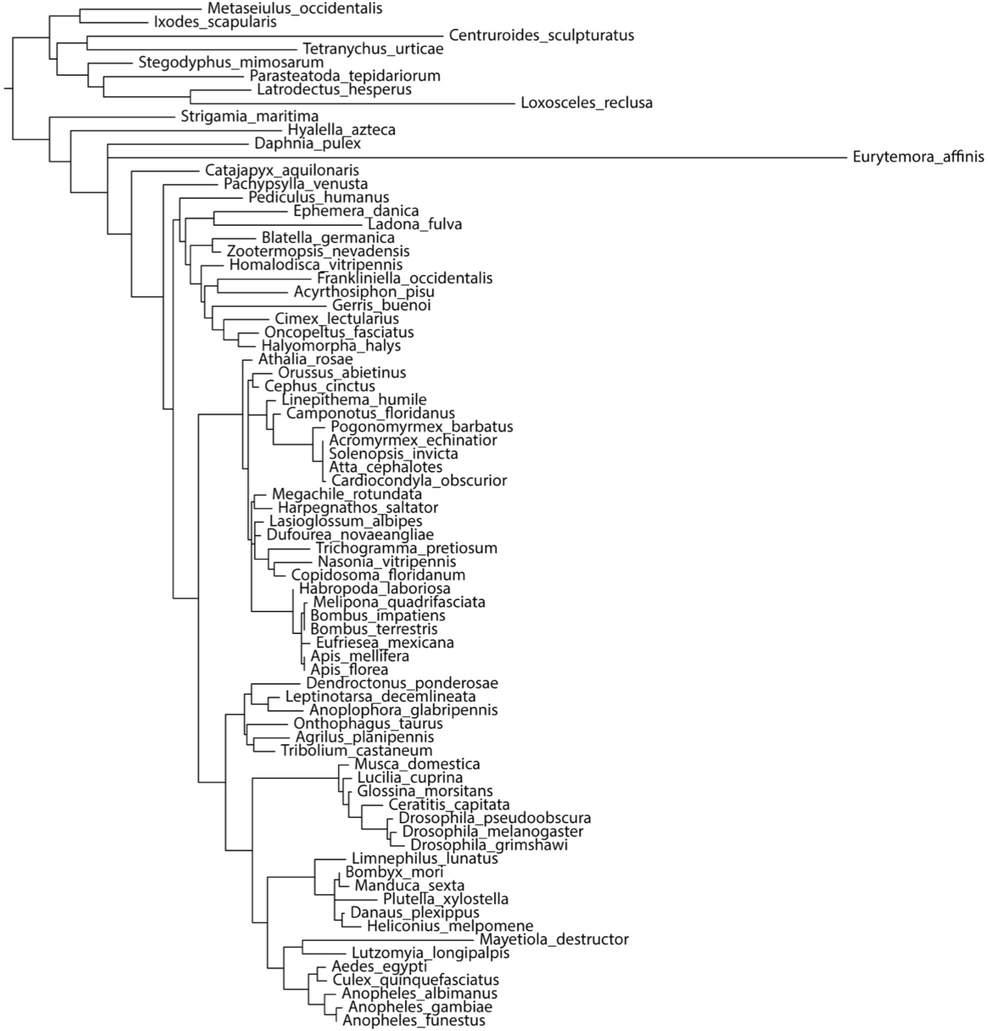
How many single-copy orthologs in our 38,195 groups?

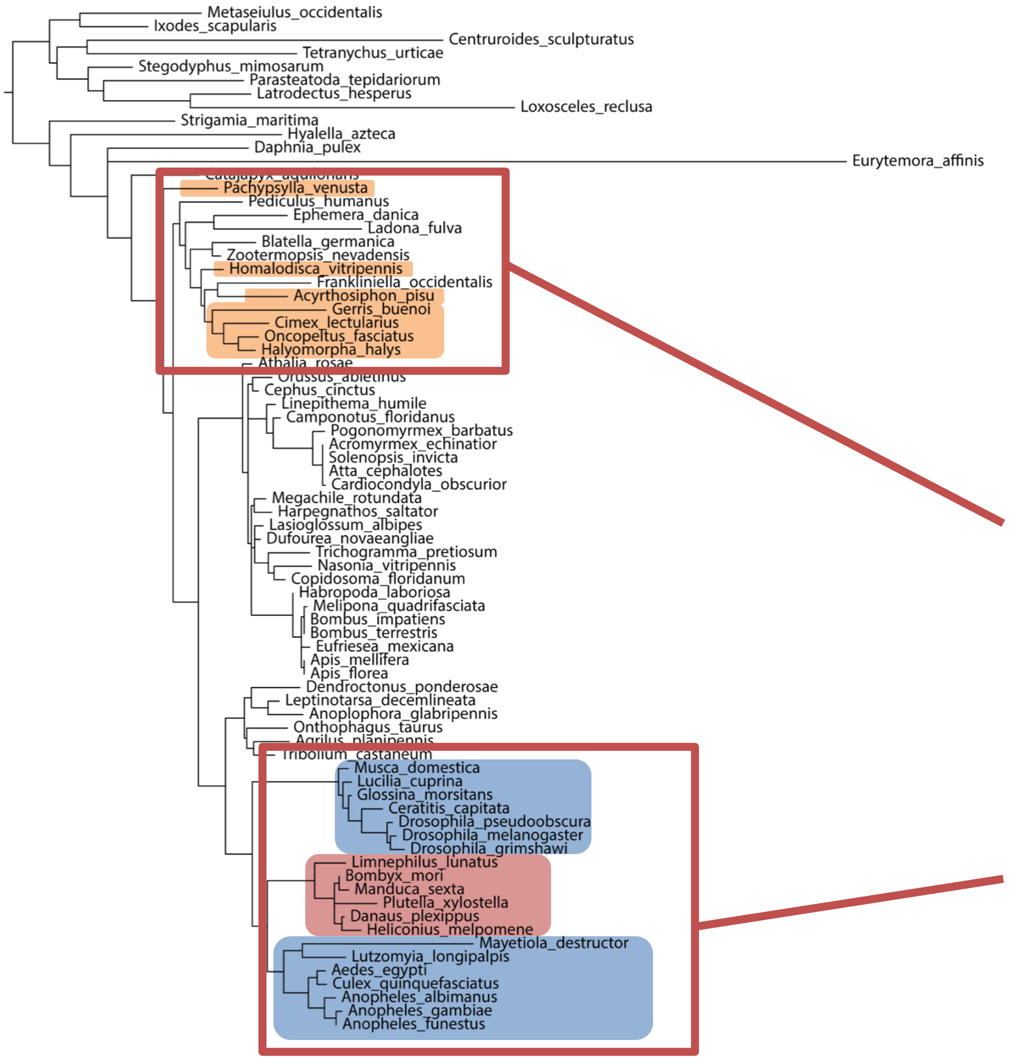
0

1 family single copy in all but one species
(2 copies in Plutella xylostella)

EOG8DFS3J

Single-copy in all but one species (2 copies in Plutella xylostella)



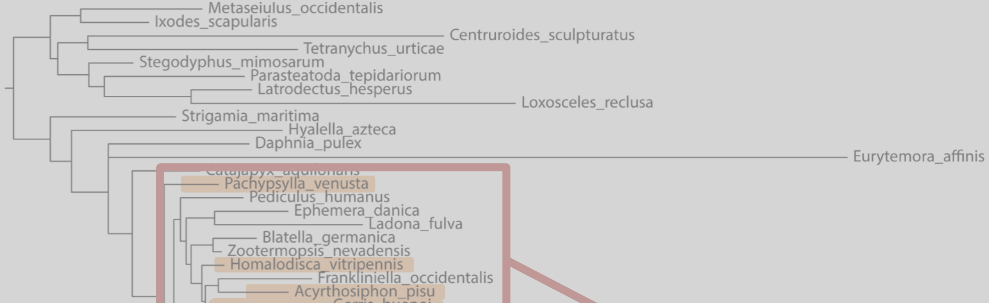


EOG8DFS3J

Single-copy in all but one species (2 copies in Plutella xylostella)

Problem: Hemiptera not monophyletic

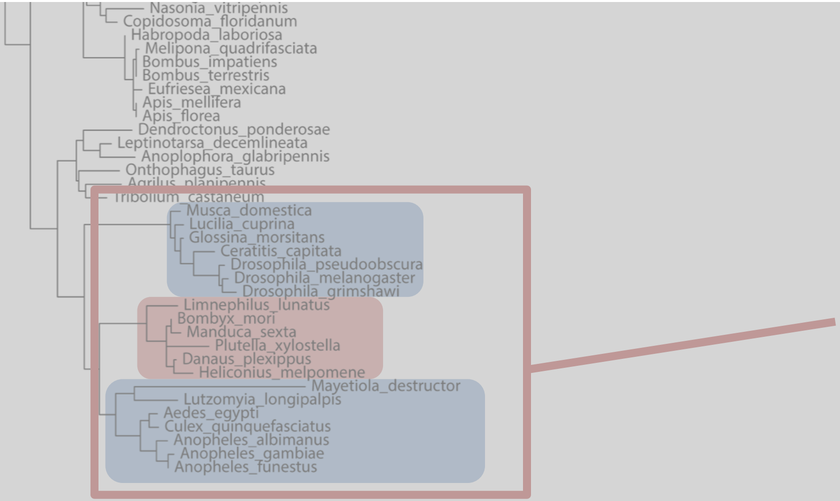
Problem: Lepidoptera and Trichoptera nested within Diptera



EOG8DFS3J

Single-copy in all but one species

How can we turn our **species** rich data into **sequence** rich data?



monophyletic

Problem: Lepidoptera and Trichoptera nested within Diptera

Construct a backbone tree among **orders** rather than **species**

76 species

Tetranychus urticae	Trombidiformes
Metaseiulus occidentalis	Mesostigmata
Ixodes scapularis	Ixodida
Centruroides sculpturatus	Scorpiones
Loxosceles reclusa	Araneae
Stegodyphus mimosarum	
Latrodectus hesperus	
Parasteatoda tepidariorum	
Strigamia maritima	Geophilomorpha
Hyalella azteca	Amphipoda
Eurytemora affinis	Calanoida
Daphnia pulex	Cladocera
Catajapyx aquilonaris	Diplura
Ladona fulva	Odonata
Ephemera danica	Ephemeroptera
Blatella germanica	Blattodea
Zootermopsis nevadensis	Isoptera
Pediculus humanus	Phthiraptera
Frankliniella occidentalis	Thysanoptera
Acyrthosiphon pisum	Hemiptera
Pachypsylla venusta	
Homalodisca vitripennis	
Gerris buenoi	
Cimex lectularius	
Halyomorpha halys	
Oncopeltus fasciatus	
Athalia rosae	Hymenoptea
Cephus cinctus	
Orussus abietinus	
Nasonia vitripennis	
Copidosoma floridanum	
Trichogramma pretiosum	
Harpegnathos saltator	
Linepithema humile	
Camponotus floridanus	
Pogonomyrmex barbatus	
Cardiocondyla obscurior	
Solenopsis invicta	
Atta cephalotes	
Acromyrmex echinatior	
Dufourea novaeangliae	
Lasioglossum albipes	
Megachile rotundata	
Habropoda laboriosa	
Eufriesea mexicana	
Apis mellifera	
Apis florea	
Melipona quadrifasciata	
Bombus impatiens	
Bombus terrestris	
Agrilus planipennis	Coleoptera
Onthophagus taurus	
Tribolium castaneum	
Dendroctonus ponderosae	
Anoplophora glabripennis	
Leptinotarsa decemlineata	
Limnephilus lunatus	Trichoptera
Plutella xylostella	Lepidoptera
Bombyx mori	
Manduca sexta	
Heliconius melpomene	
Danaus plexippus	
Aedes aegypti	Diptera
Culex quinquefasciatus	
Anopheles albimanus	
Anopheles gambiae	
Anopheles funestus	
Lutzomyia longipalpis	
Mayetiola destructor	
Drosophila grimshawi	
Drosophila pseudoobscura	
Drosophila melanogaster	
Ceratitis capitata	
Glossina morsitans	
Lucilia cuprina	
Musca domestica	

21 orders

76 species

21 orders

Tetranychus urticae	1	Trombidiformes
Metaseiulus occidentalis	1	Mesostigmata
Ixodes scapularis	1	Ixodida
Centruroides sculpturatus	1	Scorpiones
Loxosceles reclusa		Araneae
Stegodyphus mimosarum		
Latrodectus hesperus		
Parasteatoda tepidariorum		
Strigamia maritima	1	Geophilomorpha
Hyalella azteca	1	Amphipoda
Eurytemora affinis	1	Calanoida
Daphnia pulex	1	Cladocera
Catajapyx aquilonaris	1	Diplura
Ladona fulva	1	Odonata
Ephemera danica	1	Ephemeroptera
Blatella germanica	1	Blattodea
Zootermopsis nevadensis	1	Isoptera
Pediculus humanus	1	Phthiraptera
Frankliniella occidentalis	1	Thysanoptera
Acyrthosiphon pisum		Hemiptera
Pachypsylla venusta		
Homalodisca vitripennis		
Gerris buenoi		
Cimex lectularius		
Halyomorpha halys		
Oncopeltus fasciatus		
Athalia rosae		Hymenoptea
Cephus cinctus		
Orussus abietinus		
Nasonia vitripennis		
Copidosoma floridanum		
Trichogramma pretiosum		
Harpegnathos saltator		
Linepithema humile		
Camponotus floridanus		
Pogonomyrmex barbatus		
Cardiocondyla obscurior		
Solenopsis invicta		
Atta cephalotes		
Acromyrmex echinatior		
Dufourea novaeangliae		
Lasioglossum albipes		
Megachile rotundata		
Habropoda laboriosa		
Eufriesea mexicana		
Apis mellifera		
Apis florea		
Melipona quadrifasciata		
Bombus impatiens		
Bombus terrestris		
Agrilus planipennis		Coleoptera
Onthophagus taurus		
Tribolium castaneum		
Dendroctonus ponderosae		
Anoplophora glabripennis		
Leptinotarsa decemlineata		
Limnephilus lunatus	1	Trichoptera
Plutella xylostella		Lepidoptera
Bombyx mori		
Manduca sexta		
Heliconius melpomene		
Danaus plexippus		
Aedes aegypti		Diptera
Culex quinquefasciatus		
Anopheles albimanus		
Anopheles gambiae		
Anopheles funestus		
Lutzomyia longipalpis		
Mayetiola destructor		
Drosophila grimshawi		
Drosophila pseudoobscura		
Drosophila melanogaster		
Ceratitis capitata		
Glossina morsitans		
Lucilia cuprina		
Musca domestica		

Construct a backbone tree among **orders** rather than **species**

"Single-copy" orthologs are now those that:

1. Are single-copy in ALL the orders represented by a single species.

Construct a backbone tree among **orders** rather than **species**

76 species

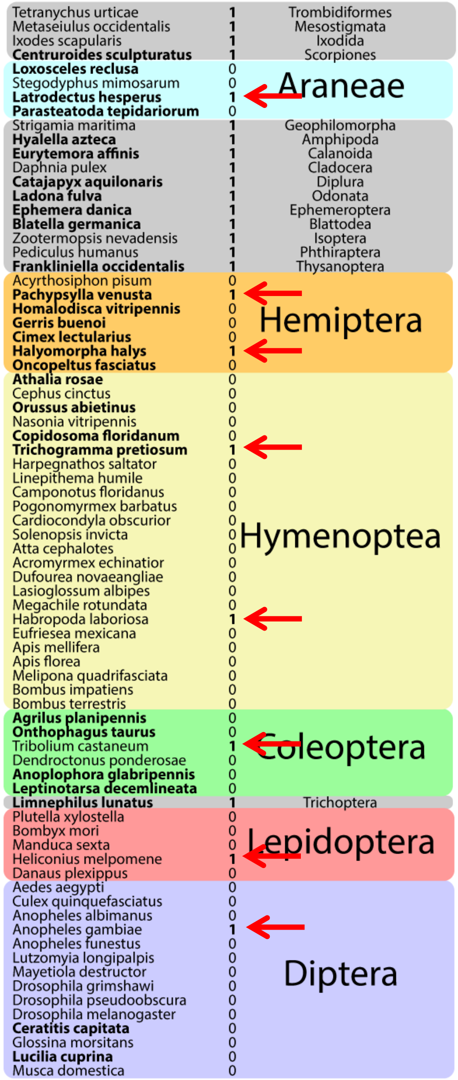
21 orders

Tetranychus urticae	1	Trombidiformes
Metaseiulus occidentalis	1	Mesostigmata
Ixodes scapularis	1	Ixodida
Centruroides sculpturatus	1	Scorpiones
Loxosceles reclusa	0	Araneae
Stegodyphus mimosarum	0	
Latrodectus hesperus	1	
Parasteatoda tepidariorum	0	
Strigamia maritima	1	Geophilomorpha
Hyalella azteca	1	Amphipoda
Eurytemora affinis	1	Calanoida
Daphnia pulex	1	Cladocera
Catajapyx aquilonaris	1	Diplura
Ladona fulva	1	Odonata
Ephemera danica	1	Ephemeroptera
Blatella germanica	1	Blattodea
Zootermopsis nevadensis	1	Isoptera
Pediculus humanus	1	Phthiraptera
Frankliniella occidentalis	1	Thysanoptera
Acyrthosiphon pisum	0	Hemiptera
Pachypsylla venusta	1	
Homalodisca vitripennis	0	
Gerris buenoi	0	
Cimex lectularius	0	
Halyomorpha halys	0	
Oncopeltus fasciatus	0	
Athalia rosae	0	Hymenoptea
Cephus cinctus	0	
Orussus abietinus	0	
Nasonia vitripennis	0	
Copidosoma floridanum	0	
Trichogramma pretiosum	1	
Harpegnathos saltator	0	
Linepithema humile	0	
Camponotus floridanus	0	
Pogonomyrmex barbatus	0	
Cardiocondyla obscurior	0	
Solenopsis invicta	0	
Atta cephalotes	0	
Acromyrmex echinatior	0	
Dufourea novaeangliae	0	
Lasioglossum albipes	0	
Megachile rotundata	0	
Habropoda laboriosa	0	
Eufriesea mexicana	0	
Apis mellifera	0	
Apis florea	0	
Melipona quadrifasciata	0	
Bombus impatiens	0	
Bombus terrestris	0	
Agrilus planipennis	0	Coleoptera
Onthophagus taurus	0	
Tribolium castaneum	1	
Dendroctonus ponderosae	0	
Anoplophora glabripennis	0	
Leptinotarsa decemlineata	0	
Limnephilus lunatus	1	Trichoptera
Plutella xylostella	0	Lepidoptera
Bombyx mori	0	
Manduca sexta	0	
Heliconius melpomene	1	
Danaus plexippus	0	
Aedes aegypti	0	Diptera
Culex quinquefasciatus	0	
Anopheles albimanus	0	
Anopheles gambiae	1	
Anopheles funestus	0	
Lutzomyia longipalpis	0	
Mayetiola destructor	0	
Drosophila grimshawi	0	
Drosophila pseudoobscura	0	
Drosophila melanogaster	0	
Ceratitis capitata	0	
Glossina morsitans	0	
Lucilia cuprina	0	
Musca domestica	0	

"Single-copy" orthologs are now those that:

1. Are single-copy in ALL the orders represented by a single species.
2. Have at least ONE species that is single-copy in each of the 6 multi-species orders.

Construct a backbone tree among **orders** rather than **species**



76 species

Tetranychus urticae	1	Trombidiformes
Metaseiulus occidentalis	1	Mesostigmata
Ixodes scapularis	1	Ixodida
Centruroides sculpturatus	1	Scorpiones
Loxosceles reclusa	0	Araneae
Stegodyphus mimosarum	0	
Latrodectus hesperus	1	
Parasteatoda tepidariorum	0	
Strigamia maritima	1	Geophilomorpha
Hyalella azteca	1	Amphipoda
Eurytemora affinis	1	Calanoida
Daphnia pulex	1	Cladocera
Catajapyx aquilonaris	1	Diplura
Ladona fulva	1	Odonata
Ephemera danica	1	Ephemeroptera
Blatella germanica	1	Blattodea
Zootermopsis nevadensis	1	Isoptera
Pediculus humanus	1	Phthiraptera
Frankliniella occidentalis	1	Thysanoptera
Acyrthosiphon pisum	0	Hemiptera
Pachypsylla venusta	1	
Homalodisca vitripennis	0	
Gerris buenoi	0	
Cimex lectularius	0	
Halyomorpha halys	1	
Oncopeltus fasciatus	0	
Athalia rosae	0	Hymenoptea
Cephus cinctus	0	
Orussus abietinus	0	
Nasonia vitripennis	0	
Copidosoma floridanum	0	
Trichogramma pretiosum	1	
Harpegnathos saltator	0	
Linepithema humile	0	
Camponotus floridanus	0	
Pogonomyrmex barbatus	0	
Cardiocondyla obscurior	0	
Solenopsis invicta	0	
Atta cephalotes	0	
Acromyrmex echinatior	0	
Dufourea novaeangliae	0	
Lasioglossum albipes	0	
Megachile rotundata	0	
Habropoda laboriosa	1	
Eufriesea mexicana	0	
Apis mellifera	0	
Apis florea	0	
Melipona quadrifasciata	0	
Bombus impatiens	0	
Bombus terrestris	0	
Agrilus planipennis	0	Coleoptera
Onthophagus taurus	0	
Tribolium castaneum	1	
Dendroctonus ponderosae	0	
Anoplophora glabripennis	0	
Leptinotarsa decemlineata	0	
Limnephilus lunatus	1	Trichoptera
Plutella xylostella	0	Lepidoptera
Bombyx mori	0	
Manduca sexta	0	
Heliconius melpomene	1	
Danaus plexippus	0	
Aedes aegypti	0	Diptera
Culex quinquefasciatus	0	
Anopheles albimanus	0	
Anopheles gambiae	1	
Anopheles funestus	0	
Lutzomyia longipalpis	0	
Mayetiola destructor	0	
Drosophila grimshawi	0	
Drosophila pseudoobscura	0	
Drosophila melanogaster	0	
Ceratitis capitata	0	
Glossina morsitans	0	
Lucilia cuprina	0	
Musca domestica	0	

21 orders

"Single-copy" orthologs are now those that:

1. Are single-copy in ALL the orders represented by a single species.

2. Have at least ONE species that is single-copy in each of the 6 multi-species orders.

Construct a backbone tree among **orders** rather than **species**

76 species

Tetranychus urticae	**1**	Trombidiformes
Metaseiulus occidentalis	**1**	Mesostigmata
Ixodes scapularis	**1**	Ixodida
Centruroides sculpturatus	**1**	Scorpiones
Loxosceles reclusa	0	Araneae
Stegodyphus mimosarum	0	
Latrodectus hesperus	**1**	
Parasteatoda tepidariorum	0	
Strigamia maritima	**1**	Geophilomorpha
Hyalella azteca	**1**	Amphipoda
Eurytemora affinis	**1**	Calanoida
Daphnia pulex	**1**	Cladocera
Catajapyx aquilonaris	**1**	Diplura
Ladona fulva	**1**	Odonata
Ephemera danica	**1**	Ephemeroptera
Blatella germanica	**1**	Blattodea
Zootermopsis nevadensis	**1**	Isoptera
Pediculus humanus	**1**	Phthiraptera
Frankliniella occidentalis	**1**	Thysanoptera
Acyrthosiphon pisum	0	Hemiptera
Pachypsylla venusta	**1**	
Homalodisca vitripennis	0	
Gerris buenoi	0	
Cimex lectularius	0	
Halyomorpha halys	**1**	
Oncopeltus fasciatus	0	
Athalia rosae	0	Hymenoptea
Cephus cinctus	0	
Orussus abietinus	0	
Nasonia vitripennis	0	
Copidosoma floridanum	0	
Trichogramma pretiosum	**1**	
Harpegnathos saltator	0	
Linepithema humile	0	
Camponotus floridanus	0	
Pogonomyrmex barbatus	0	
Cardiocondyla obscurior	0	
Solenopsis invicta	0	
Atta cephalotes	0	
Acromyrmex echinatior	0	
Dufourea novaeangliae	0	
Lasioglossum albipes	0	
Megachile rotundata	0	
Habropoda laboriosa	**1**	
Eufriesea mexicana	0	
Apis mellifera	6	
Apis florea	6	
Melipona quadrifasciata	4	
Bombus impatiens	5	
Bombus terrestris	5	
Agrilus planipennis	0	Coleoptera
Onthophagus taurus	0	
Tribolium castaneum	**1**	
Dendroctonus ponderosae	0	
Anoplophora glabripennis	0	
Leptinotarsa decemlineata	0	
Limnephilus lunatus	**1**	Trichoptera
Plutella xylostella	0	Lepidoptera
Bombyx mori	0	
Manduca sexta	0	
Heliconius melpomene	**1**	
Danaus plexippus	0	
Aedes aegypti	0	Diptera
Culex quinquefasciatus	0	
Anopheles albimanus	0	
Anopheles gambiae	**1**	
Anopheles funestus	0	
Lutzomyia longipalpis	0	
Mayetiola destructor	0	
Drosophila grimshawi	0	
Drosophila pseudoobscura	0	
Drosophila melanogaster	0	
Ceratitis capitata	0	
Glossina morsitans	0	
Lucilia cuprina	0	
Musca domestica	0	

21 orders

✓

"Single-copy" orthologs are now those that:

1. Are single-copy in ALL the orders represented by a single species.
2. Have at least ONE species that is single-copy in each of the 6 multi-species orders.

Construct a backbone tree among **orders** rather than **species**

76 species

Tetranychus urticae	**1**	Trombidiformes
Metaseiulus occidentalis	**1**	Mesostigmata
Ixodes scapularis	**1**	Ixodida
Centruroides sculpturatus	**1**	Scorpiones
Loxosceles reclusa	0	Araneae
Stegodyphus mimosarum	0	
Latrodectus hesperus	**1**	
Parasteatoda tepidariorum	0	
Strigamia maritima	**1**	Geophilomorpha
Hyalella azteca	**1**	Amphipoda
Eurytemora affinis	**1**	Calanoida
Daphnia pulex	**1**	Cladocera
Catajapyx aquilonaris	**1**	Diplura
Ladona fulva	**1**	Odonata
Ephemera danica	**1**	Ephemeroptera
Blatella germanica	**1**	Blattodea
Zootermopsis nevadensis	**1**	Isoptera
Pediculus humanus	**1**	Phthiraptera
Frankliniella occidentalis	**1**	Thysanoptera
Acyrthosiphon pisum	0	Hemiptera
Pachypsylla venusta	0	
Homalodisca vitripennis	0	
Gerris buenoi	0	
Cimex lectularius	0	
Halyomorpha halys	0	
Oncopeltus fasciatus	0	
Athalia rosae	0	Hymenoptea
Cephus cinctus	0	
Orussus abietinus	0	
Nasonia vitripennis	0	
Copidosoma floridanum	0	
Trichogramma pretiosum	**1**	
Harpegnathos saltator	0	
Linepithema humile	0	
Camponotus floridanus	0	
Pogonomyrmex barbatus	0	
Cardiocondyla obscurior	0	
Solenopsis invicta	0	
Atta cephalotes	0	
Acromyrmex echinatior	0	
Dufourea novaeangliae	0	
Lasioglossum albipes	0	
Megachile rotundata	0	
Habropoda laboriosa	0	
Eufriesea mexicana	0	
Apis mellifera	0	
Apis florea	0	
Melipona quadrifasciata	0	
Bombus impatiens	0	
Bombus terrestris	0	
Agrilus planipennis	0	Coleoptera
Onthophagus taurus	0	
Tribolium castaneum	**1**	
Dendroctonus ponderosae	0	
Anoplophora glabripennis	0	
Leptinotarsa decemlineata	0	
Limnephilus lunatus	**1**	Trichoptera
Plutella xylostella	0	Lepidoptera
Bombyx mori	0	
Manduca sexta	0	
Heliconius melpomene	**1**	
Danaus plexippus	0	
Aedes aegypti	0	Diptera
Culex quinquefasciatus	0	
Anopheles albimanus	0	
Anopheles gambiae	**1**	
Anopheles funestus	0	
Lutzomyia longipalpis	0	
Mayetiola destructor	0	
Drosophila grimshawi	0	
Drosophila pseudoobscura	0	
Drosophila melanogaster	0	
Ceratitis capitata	0	
Glossina morsitans	0	
Lucilia cuprina	0	
Musca domestica	0	

21 orders

"Single-copy" orthologs are now those that:

1. Are single-copy in ALL the orders represented by a single species.
2. Have at least ONE species that is single-copy in each of the 6 multi-species orders.

Construct a backbone tree among **orders** rather than **species**

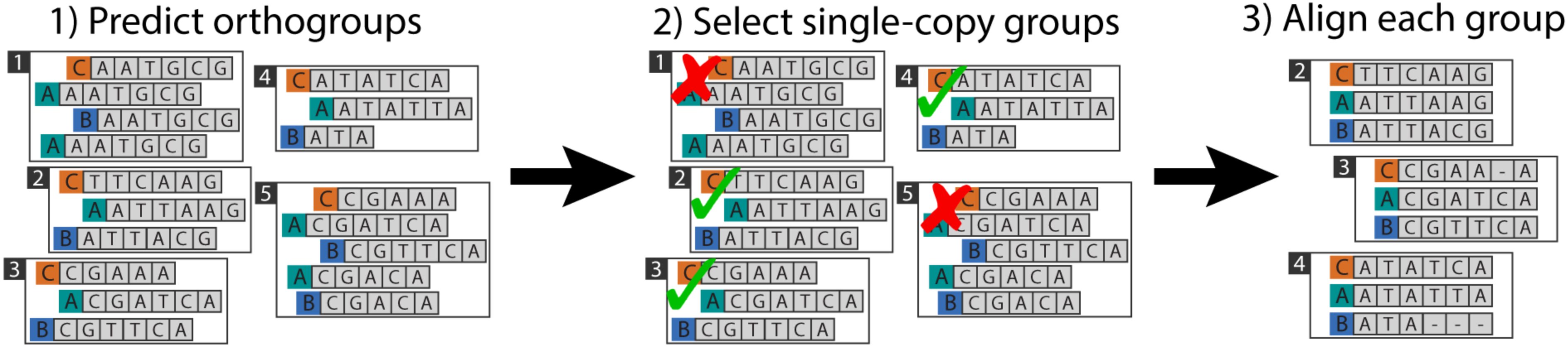
Phylum	# Orders	# single-copy orthologs
Arthropoda	21	150

Construct a backbone tree among **orders** rather than **species**

Phylum	# Orders	# single-copy orthologs
Arthropoda	21	150

Then use single-copy orthologs from the 6 multi-species orders to construct order-level trees

Order	# Species	# single-copy orthologs
Araneae	4	1627
Hemiptera	7	2053
Hymenoptera	24	2121
Coleoptera	6	3880
Lepidoptera	5	3660
Diptera	14	1324


1) Predict orthogroups
2) Select single-copy groups
3) Align each group
1
C A A T G C G
A A A T G C G
B A A T G C G
A A A T G C G
2
C T T C A A G
A A T T A A G
B A T T A C G
3
C C G A A A
A C G A T C A
B C G T T C A
4
C A T A T C A
A A T A T T A
B A T A
5
C C G A A A
A C G A T C A
B C G T T C A
A C G A C A
B C G A C A
2
C T T C A A G
A A T T A A G
B A T T A C G
3
C C G A A - A
A C G A T C A
B C G T T C A
4
C A T A T C A
A A T A T T A
B A T A - - -

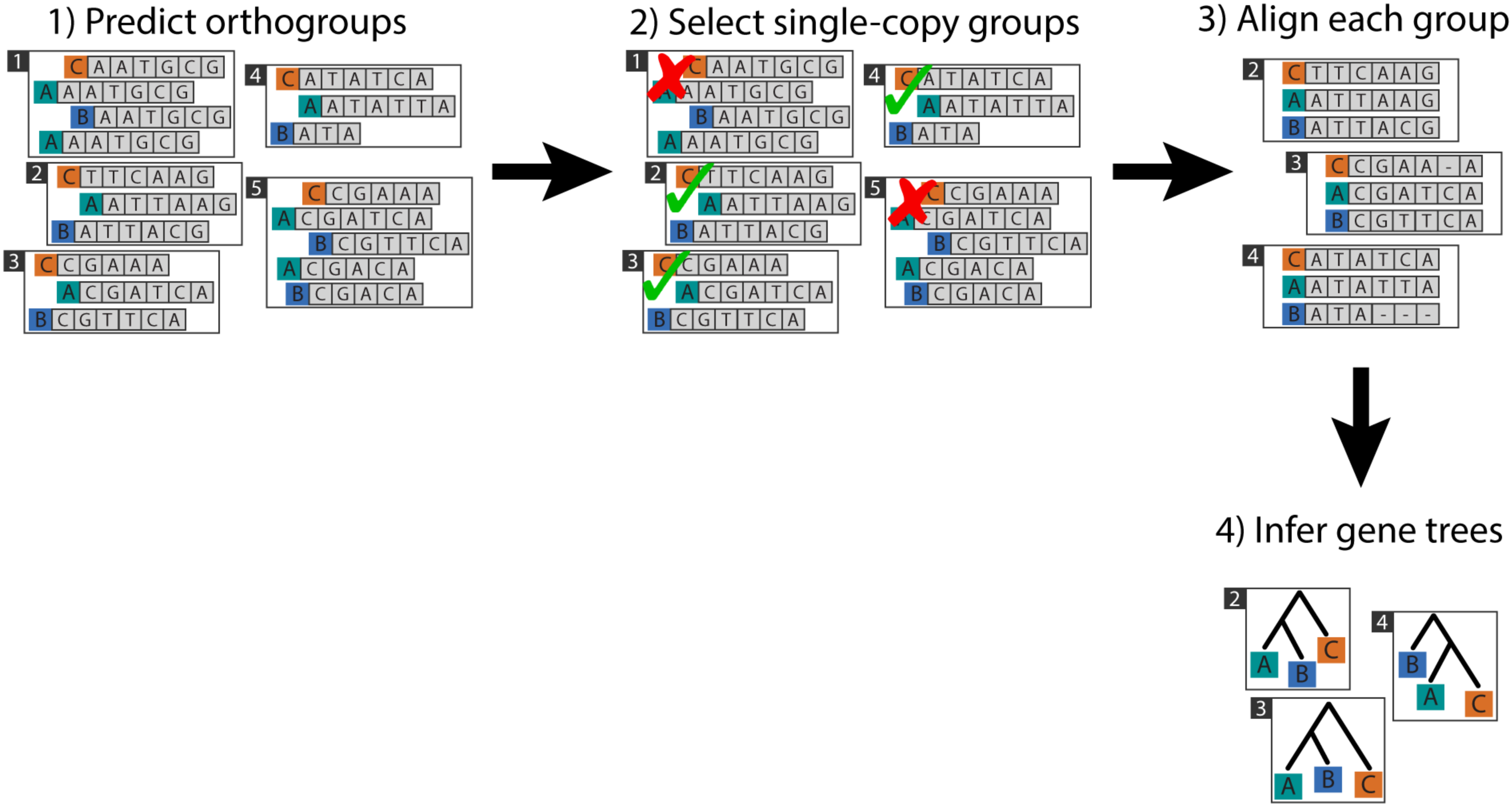
3) Align each group

Phylum	# Orders	# single-copy orthologs
Arthropoda	21	150

Order	# Species	# single-copy orthologs
Araneae	4	1627
Hemiptera	7	2053
Hymenoptera	24	2121
Coleoptera	6	3880
Lepidoptera	5	3660
Diptera	14	1324

Two alignment programs:

1. MUSCLE
2. PASTA


1) Predict orthogroups
1
C AATGCG
A AATGCG
B AATGCG
A AATGCG
2
C TTCAAG
A ATTAAG
B ATTACG
3
C CGAAA
A CGATCA
B CGTTCA
4
C ATATCA
A ATATTA
B ATA
5
C CGAAA
A CGATCA
B CGTTCA
A CGACA
B CGACA
2) Select single-copy groups
1
C AATGCG
A AATGCG
B AATGCG
A AATGCG
2
C TTCAAG
A ATTAAG
B ATTACG
3
C CGAAA
A CGATCA
B CGTTCA
4
C ATATCA
A ATATTA
B ATA
5
C CGAAA
A CGATCA
B CGTTCA
A CGACA
B CGACA
3) Align each group
2
C TTCAAG
A ATTAAG
B ATTACG
3
C CGAA-A
A CGATCA
B CGTTCA
4
C ATATCA
A ATATTA
B ATA---
4) Infer gene trees
2
A
B
C
3
A
B
C
4
B
A
C

4) Infer gene trees

Phylum	# Orders	# single-copy orthologs
Arthropoda	21	150

Order	# Species	# single-copy orthologs
Araneae	4	1627
Hemiptera	7	2053
Hymenoptera	24	2121
Coleoptera	6	3880
Lepidoptera	5	3660
Diptera	14	1324

RAxML:
with PROTGAMMAJTTF amino acid substitution model

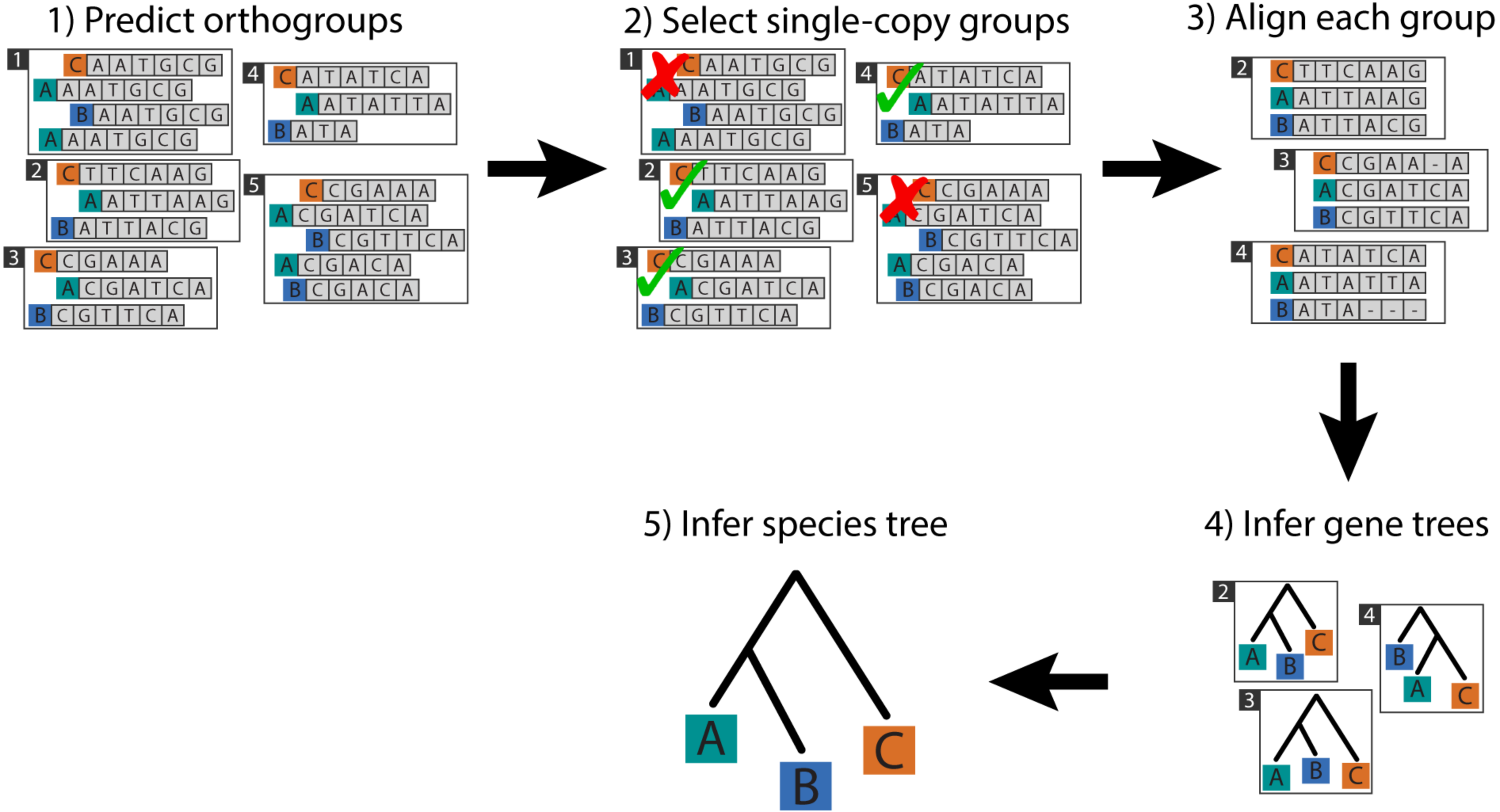
4) Infer gene trees

Phylum	# Orders	# single-copy orthologs
Arthropoda	21	150

Order	# Species	# single-copy orthologs
Araneae	4	1627
Hemiptera	7	2053
Hymenoptera	24	2121
Coleoptera	6	3880
Lepidoptera	5	3660
Diptera	14	1324

RAxML:
with PROTGAMMAJTTF amino acid substitution model

Topologies largely insensitive to substitution model


1) Predict orthogroups
1 C AATGCG
A AATGCG
B AATGCG
A AATGCG
2 C TTCAAG
A ATTAAG
B ATTACG
3 C CGAAA
A CGATCA
B CGTTCA
4 C ATATCA
A ATATTA
B ATA
5 C CGAAA
A CGATCA
B CGTTCA
A CGACA
B CGACA
2) Select single-copy groups
3) Align each group
2 C TTCAAG
A ATTAAG
B ATTACG
3 C CGAA-A
A CGATCA
B CGTTCA
4 C ATATCA
A ATATTA
B ATA---
4) Infer gene trees
2 A B C
4 B A C
3 A B C
5) Infer species tree
A B C

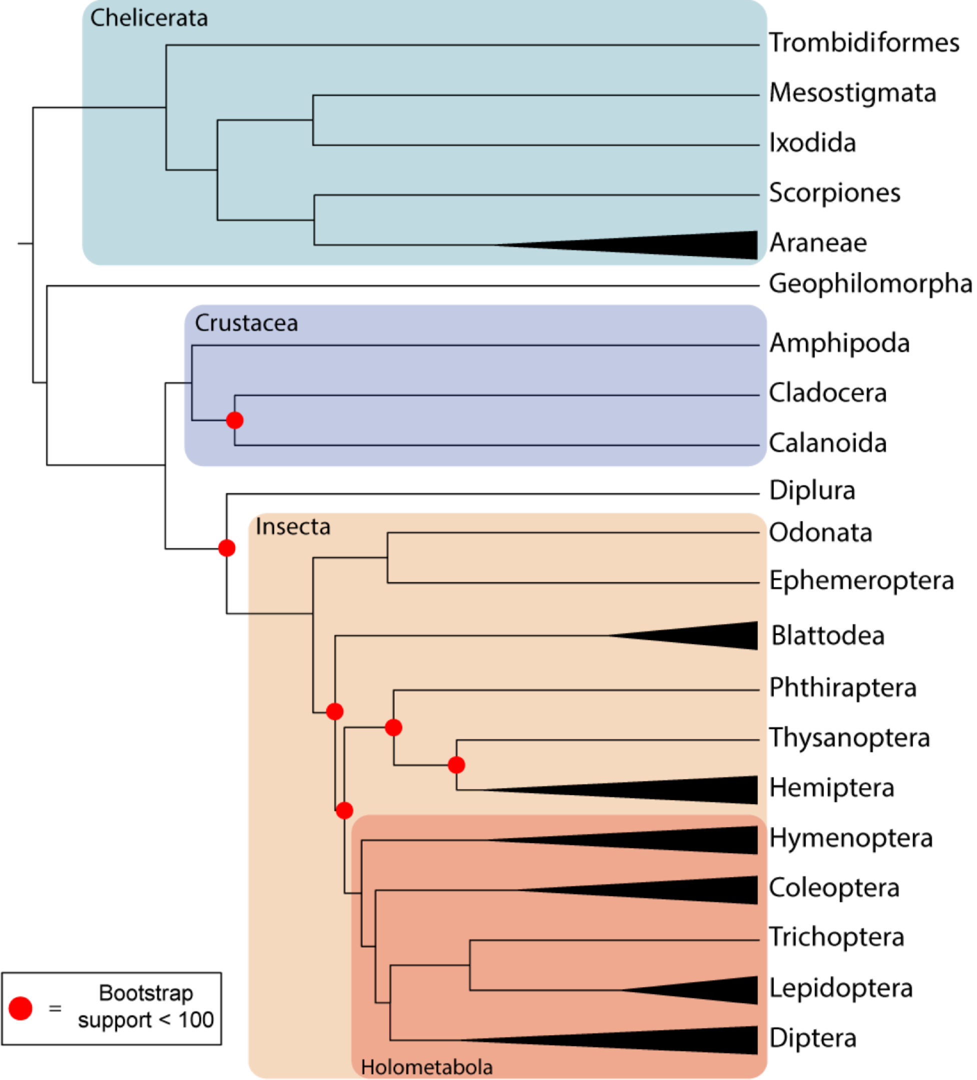
5) Infer species tree

Phylum	# Orders	# single-copy orthologs
Arthropoda	21	150

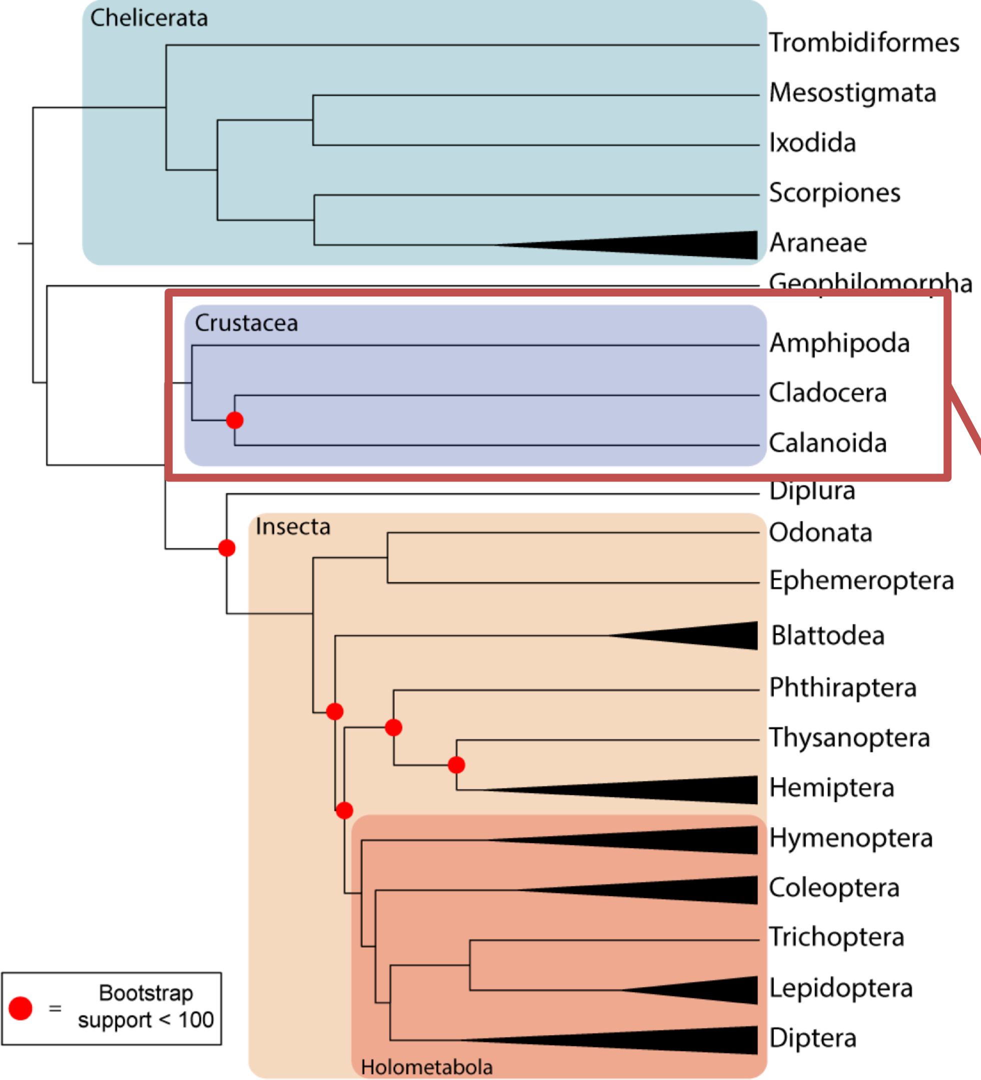
Three species tree methods:

1. Average consensus
2. Concatenation
3. ASTRAL

The backbone phylogeny



The backbone phylogeny



Monophyletic Crustacea??

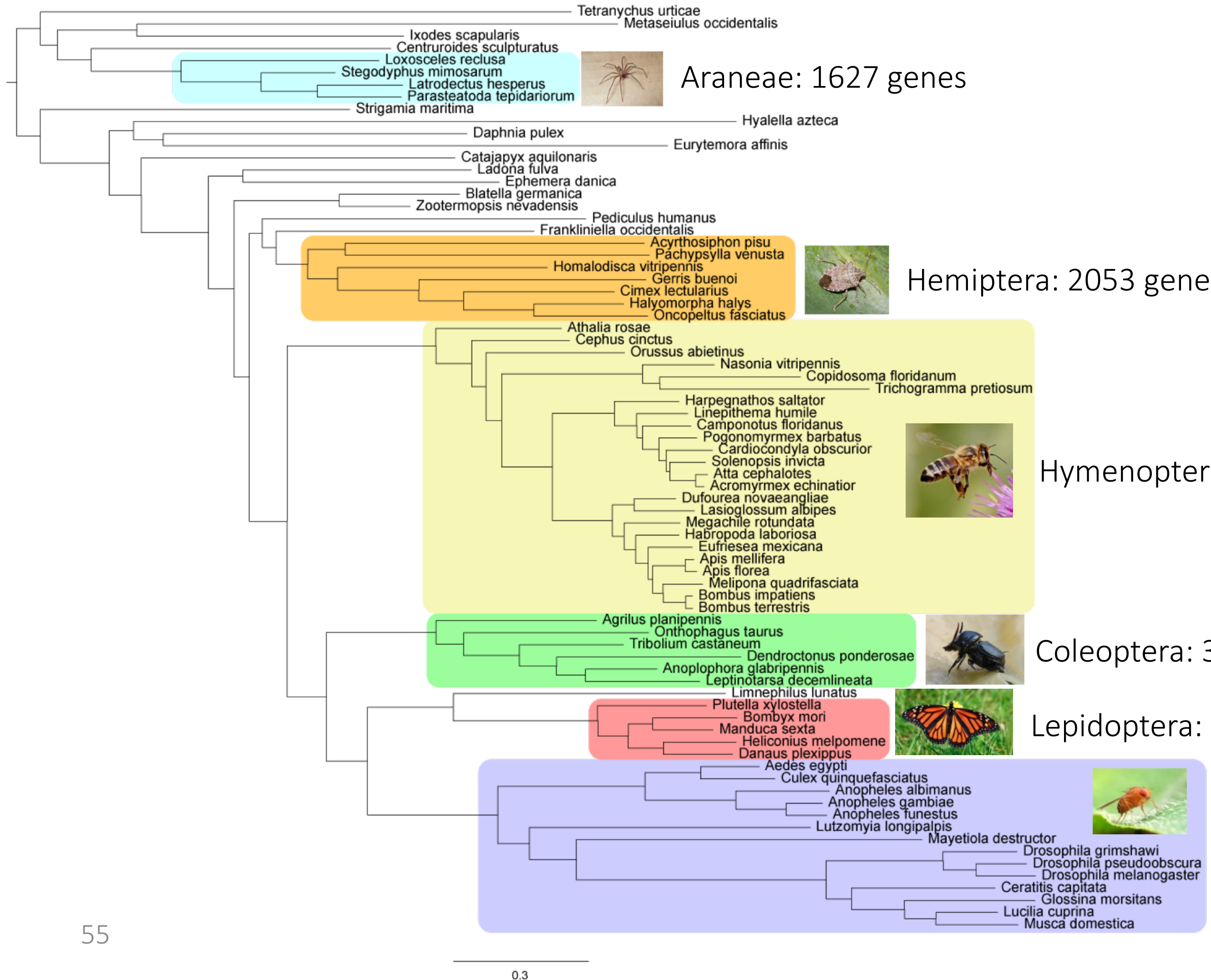
5) Infer species tree

Order	# Species	# single-copy orthologs
Araneae	4	1627
Hemiptera	7	2053
Hymenoptera	24	2121
Coleoptera	6	3880
Lepidoptera	5	3660
Diptera	14	1324

Three species tree methods:

1. Average consensus
2. Concatenation
3. ASTRAL

The Arthropod phylogeny



Araneae: 1627 genes

Hemiptera: 2053 genes

Hymenoptera: 2121 genes

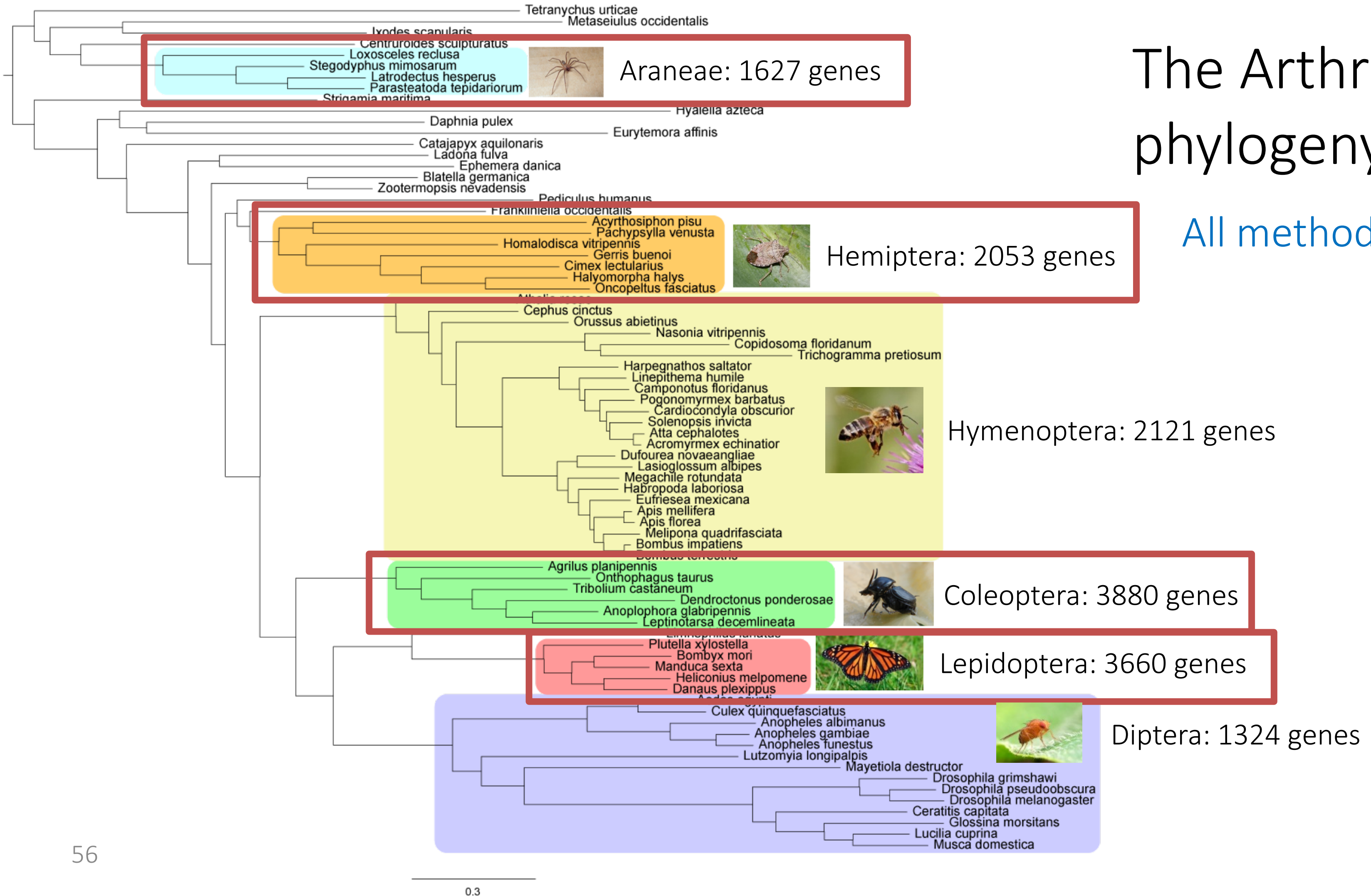
Coleoptera: 3880 genes

Lepidoptera: 3660 genes

Diptera: 1324 genes

The Arthropod phylogeny

All methods agree ✓



Araneae: 1627 genes

Hemiptera: 2053 genes

Hymenoptera: 2121 genes

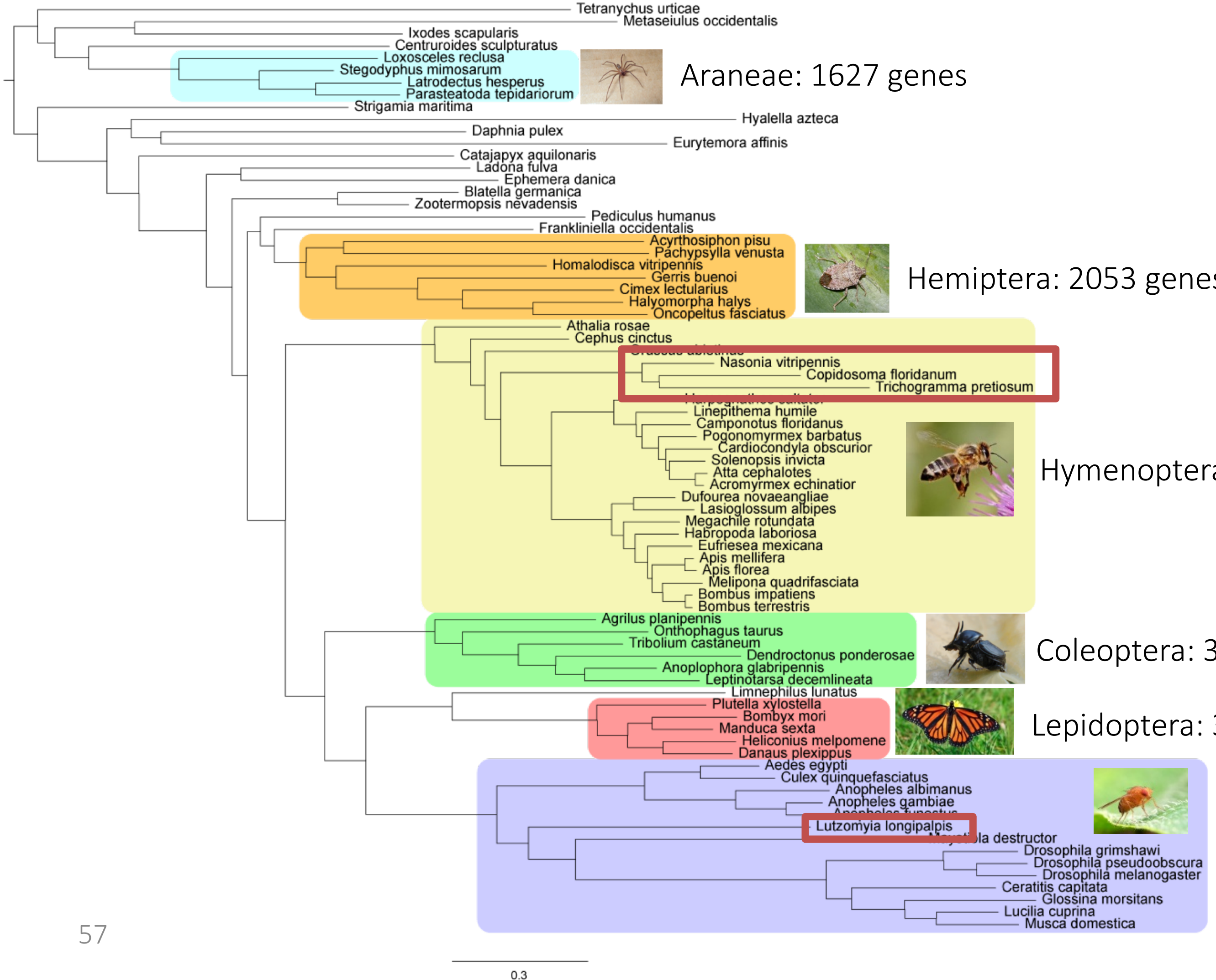
Coleoptera: 3880 genes

Lepidoptera: 3660 genes

Diptera: 1324 genes

The Arthropod phylogeny

Disagreement between methods ✗



Araneae: 1627 genes

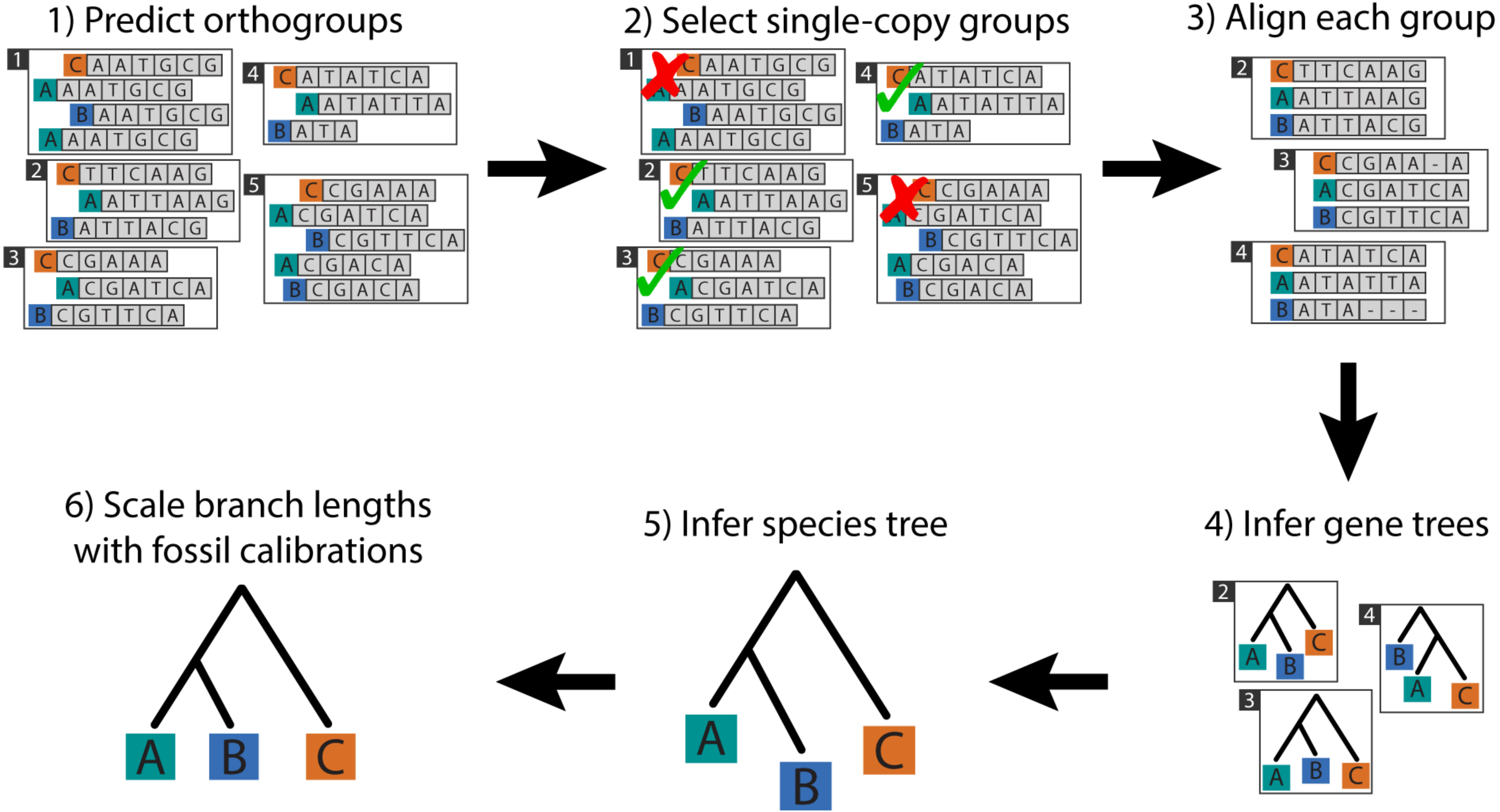
Hemiptera: 2053 genes

Hymenoptera: 2121 genes

Coleoptera: 3880 genes

Lepidoptera: 3660 genes

Diptera: 1324 genes


1) Predict orthogroups
1 CAATGCG AAATGCG BAATGCG AAATGCG
2 CTTCAAG AATTAAG BATTACG
3 CCGAAA ACGATCA BCGTTCA
4 CATATCA AATATTA BATA
5 CCGAAA ACGATCA BCGTTCA ACGACA BCGACA
2) Select single-copy groups
1 2 3 4 5
3) Align each group
2 CTTCAAG AATTAAG BATTACG
3 CCGAA-A ACGATCA BCGTTCA
4 CATATCA AATATTA BATA---
4) Infer gene trees
2 A B C
3 A B C
4 B A C
5) Infer species tree
A B C
6) Scale branch lengths with fossil calibrations
A B C

6) Scale branch lengths with fossil calibrations

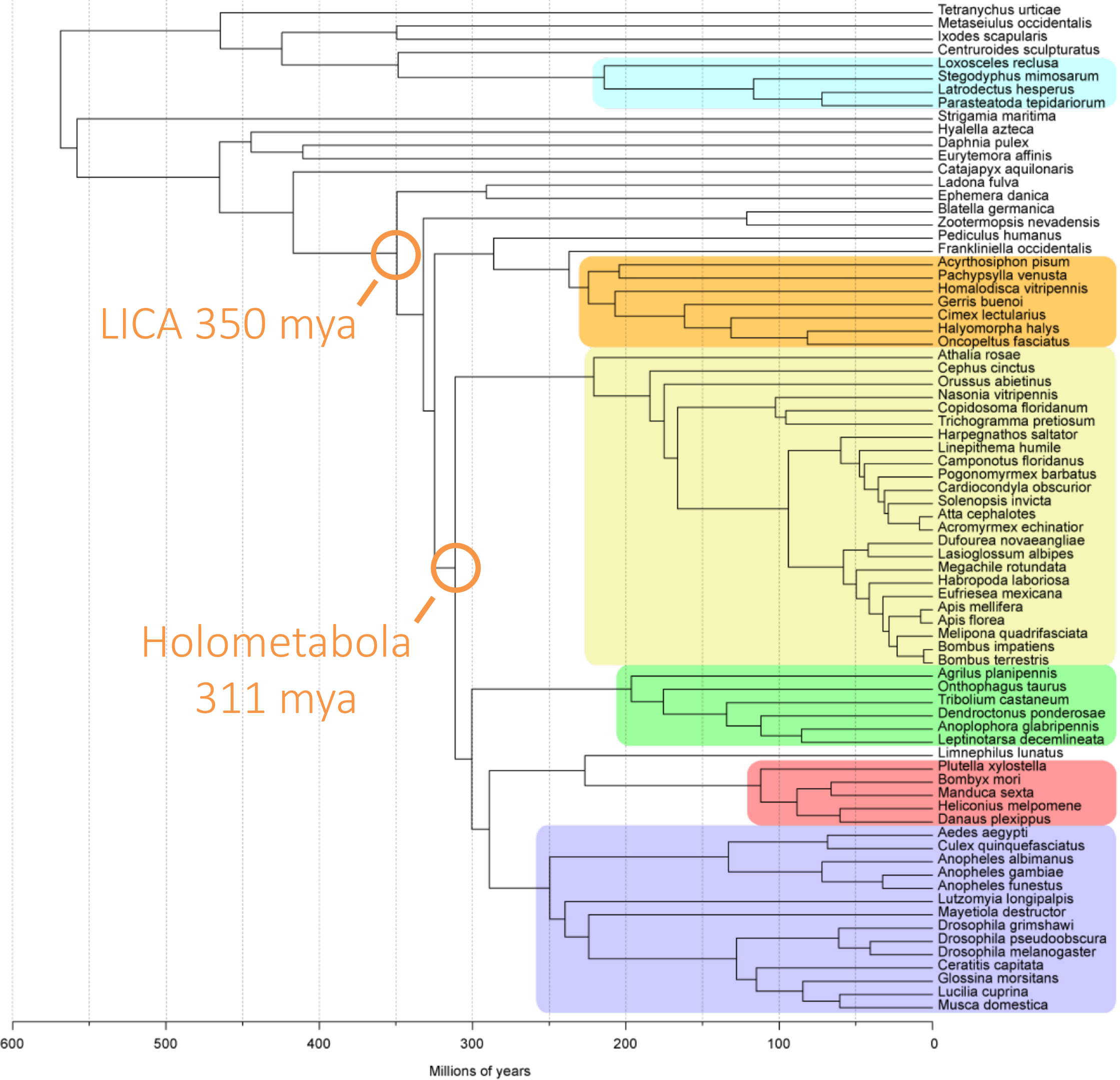
Crown group	Node	Min time	Max time
Euarthropoda	75	514	636.1
Arachnida	74	432.6	636.1
Parasitiformes	72	98.17	514
Mandibulata	67	514	636.1
Multicrustacea	64	487	636.1
Pterygota	62	322.83	521
Paleoptera	1	319.9	521
Neoptera	61	319.9	411
Blattodea	2	130.3	411
Eumetabola	60	319.9	411
Condylognatha	58	306.9	411
Hemiptera	57	306.9	411
Holometabola	51	313.7	411
Hymenoptera	25	226.4	411
Aparaglossata	50	313.7	411
Coleoptera	30	208.5	411
Mecopterida	49	271.8	411
Amphiesmenoptera	35	195.31	411
Lepidoptera	34	129.41	411
Diptera	48	240.5	411

Order	Node	Min time	Max time
Hymenoptera	HY25	89.9	93.9
Hymenoptera	HY13	23	28.4

Use r8s to smooth the tree:

Penalized likelihood method to correlate rates of evolution among branches

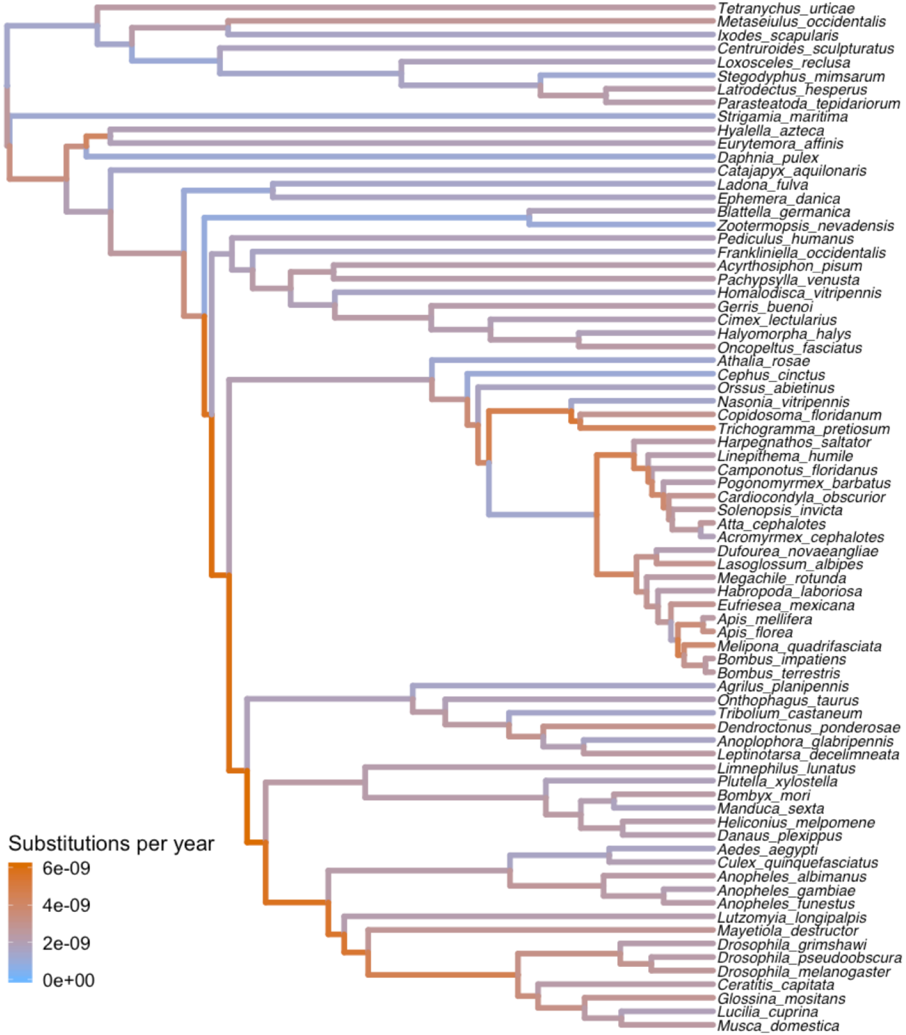
Arthropod Time Tree



LICA 350 mya

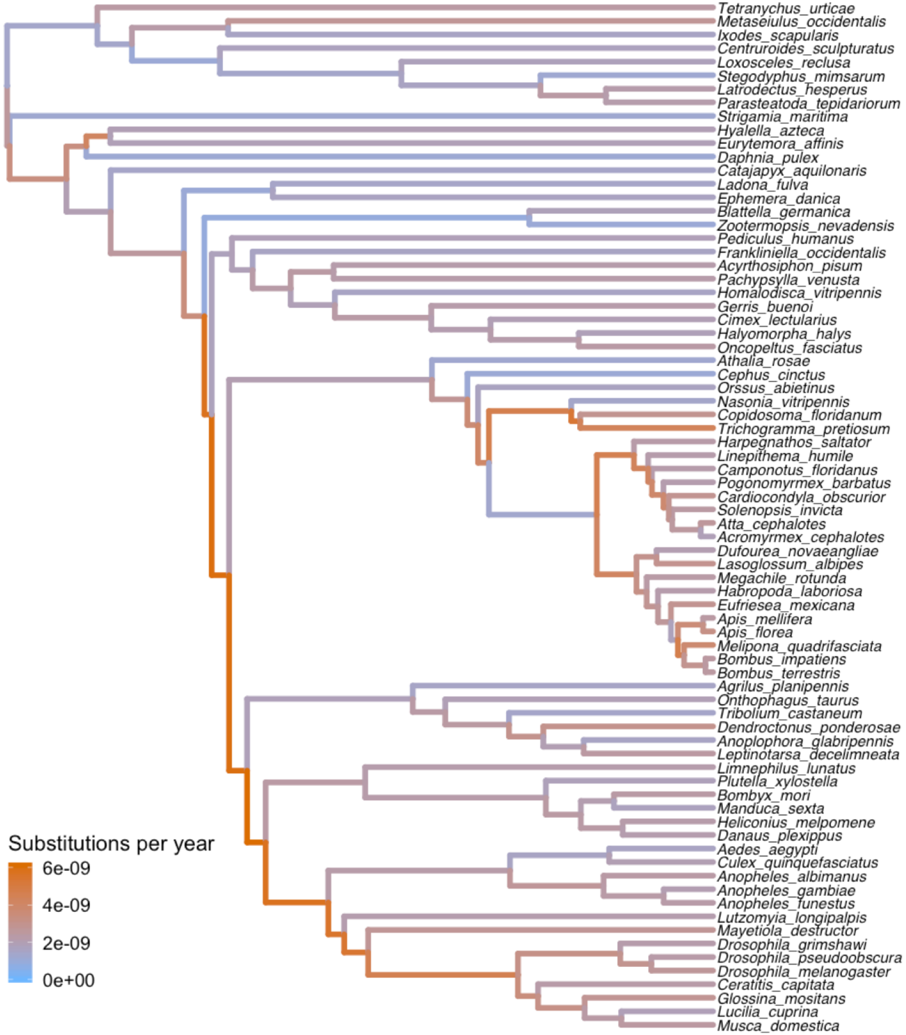
Holometabola 311 mya

Arthropod Time Tree



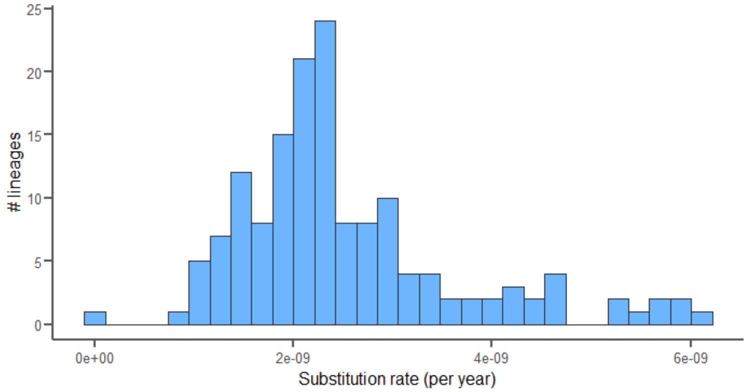
The branches of the ML tree can be scaled by time to infer substitution rates

Arthropod Time Tree

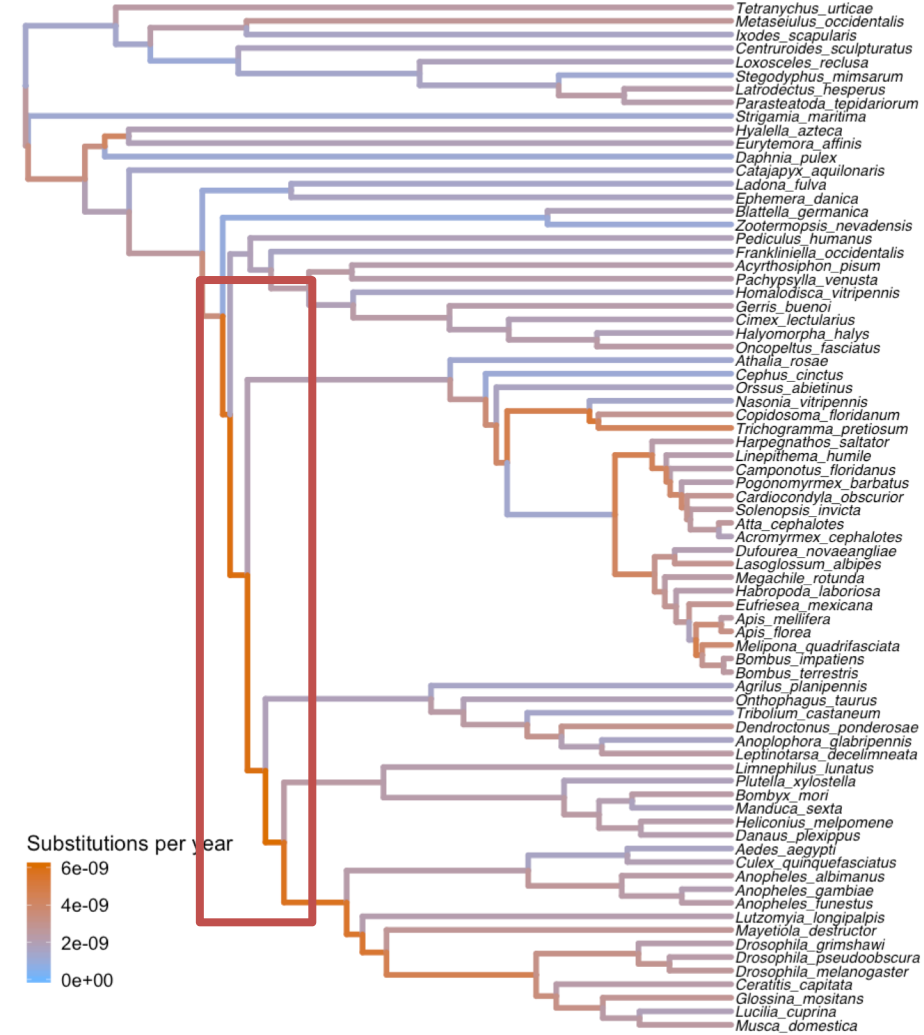


The branches of the ML tree can be scaled by time to infer substitution rates

Rates are mostly consistent across arthropods

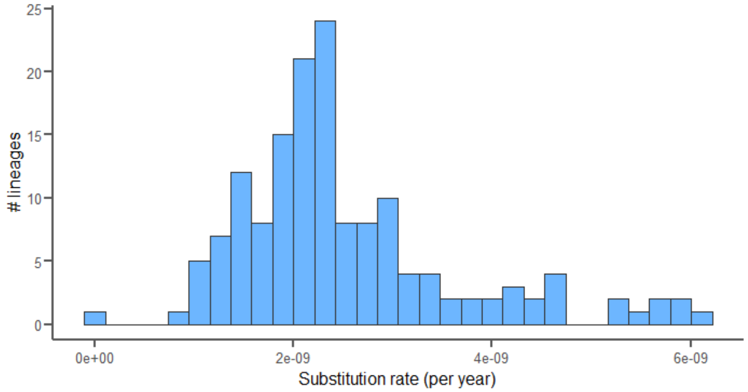


Arthropod Time Tree



The branches of the ML tree can be scaled by time to infer substitution rates

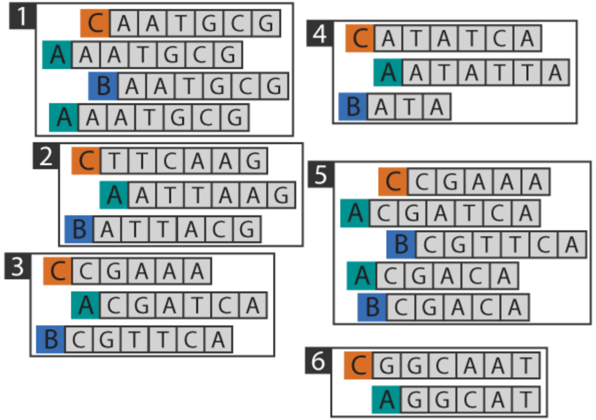
Rates are mostly consistent across arthropods

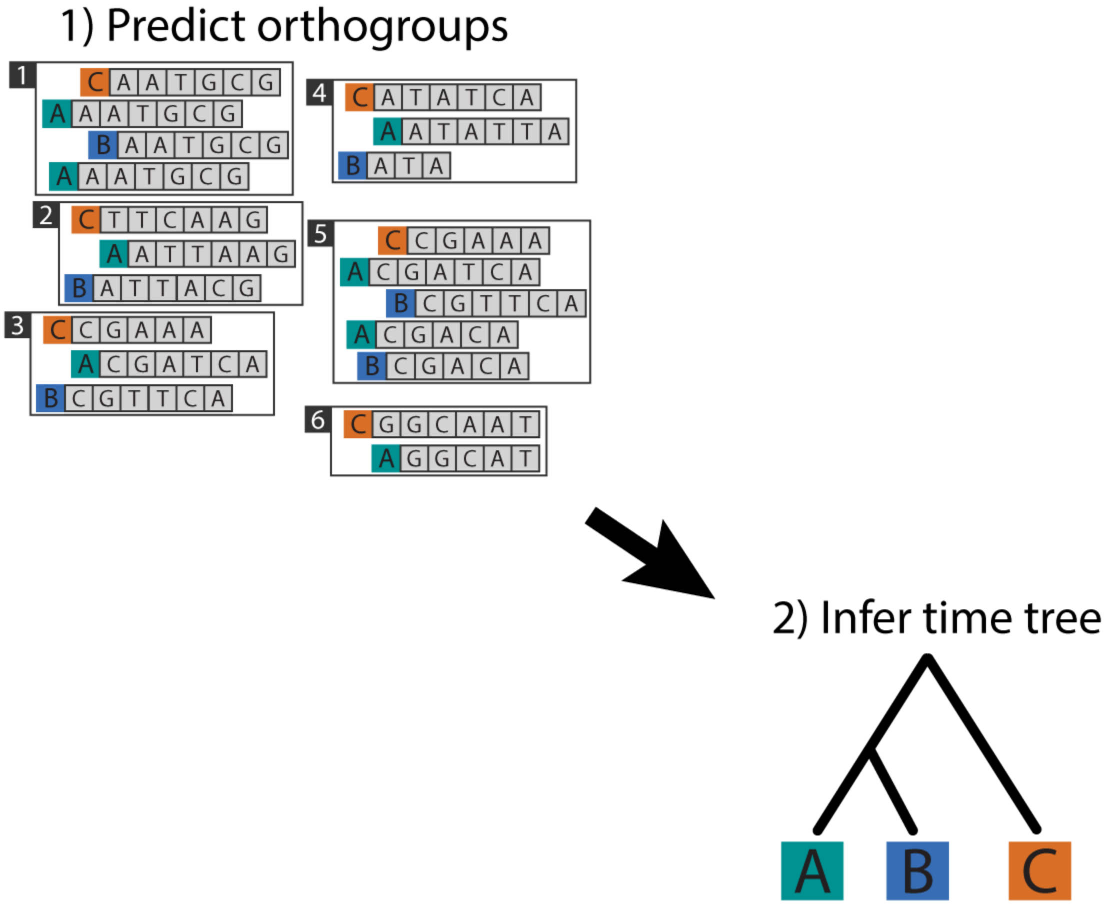


Today's topics

1. Determining the Arthropod phylogeny
2. Reconstructing ancestral gene counts
3. Using the i5k gene family web site

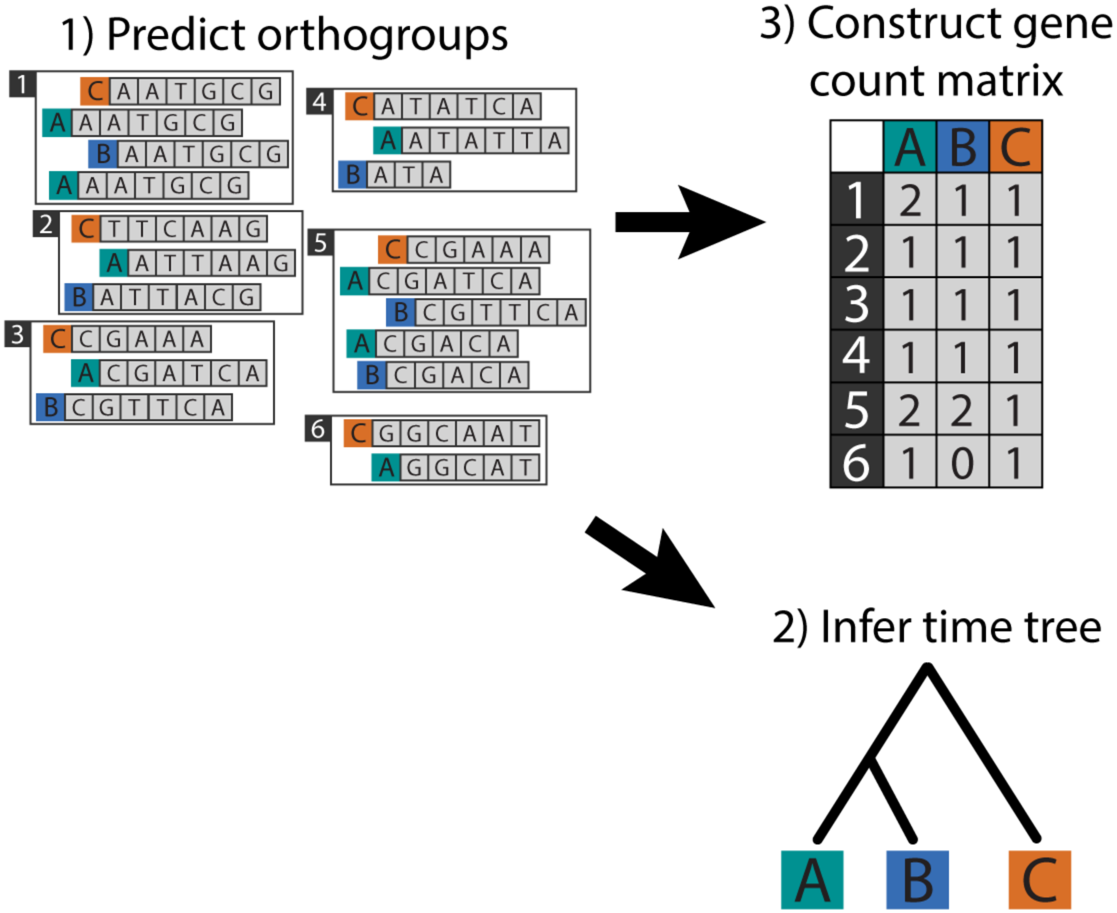
1) Predict orthogroups





1) Predict orthogroups

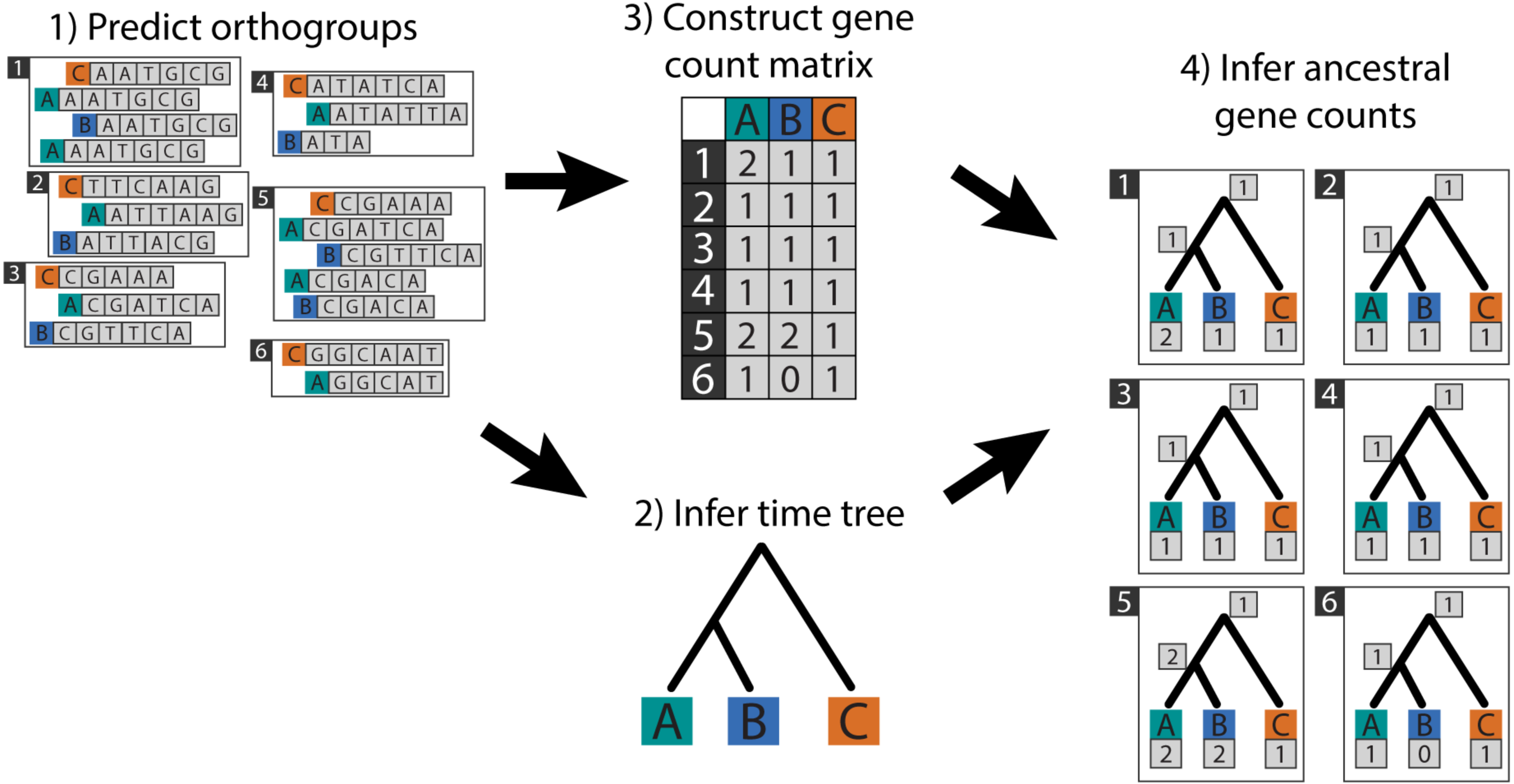
2) Infer time tree


1) Predict orthogroups

3) Construct gene count matrix

	A	B	C
1	2	1	1
2	1	1	1
3	1	1	1
4	1	1	1
5	2	2	1
6	1	0	1

2) Infer time tree



1) Predict orthogroups

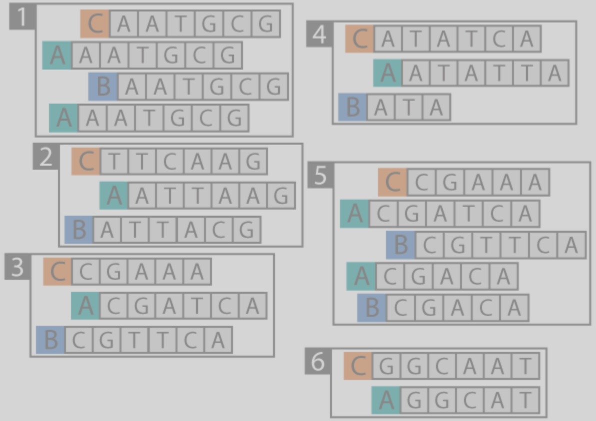
2) Infer time tree

3) Construct gene count matrix

	A	B	C
1	2	1	1
2	1	1	1
3	1	1	1
4	1	1	1
5	2	2	1
6	1	0	1

4) Infer ancestral gene counts

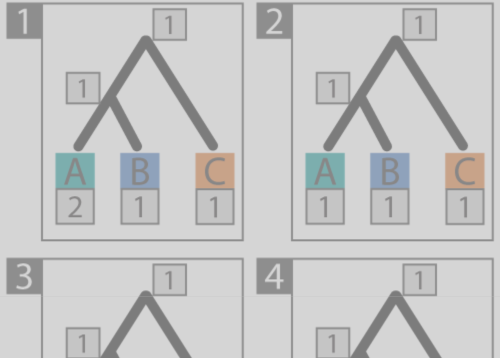
1) Predict orthogroups

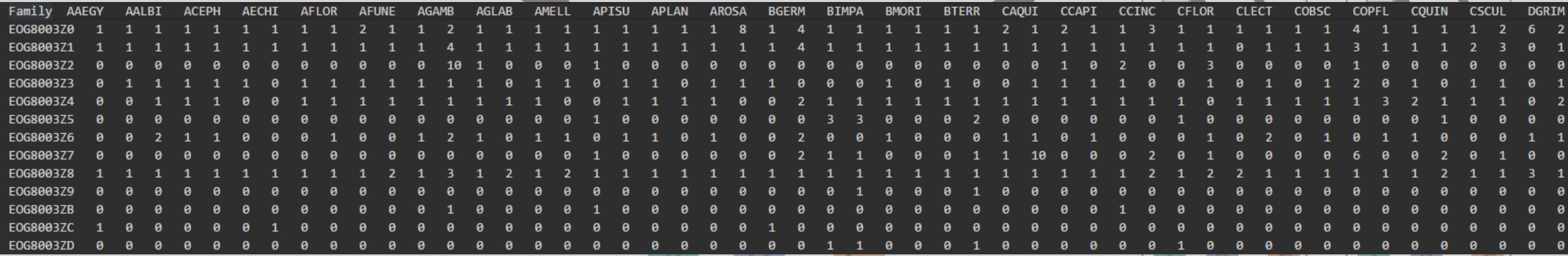


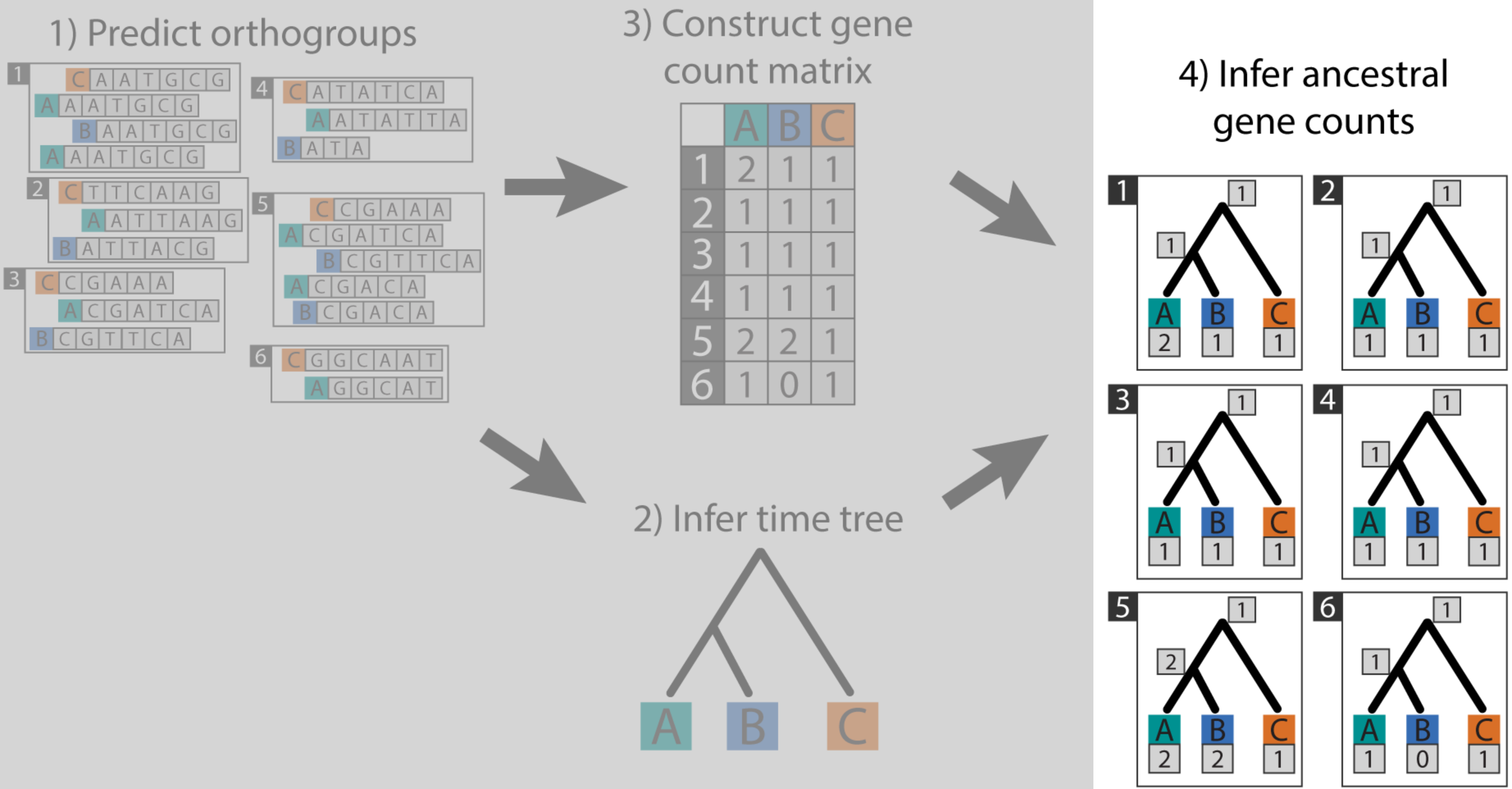
3) Construct gene count matrix

	A	B	C
1	2	1	1
2	1	1	1
3	1	1	1
4	1	1	1
5	2	2	1
6	1	0	1

4) Infer ancestral gene counts




```
Family  AAEGY   AALBI   ACEPH   AECHI   AFLOR   AFUNE   AGAMB   AGLAB   AMELL   APISU   APLAN   AROSA   BGERM   BIMPA   BMORI   BTERR   CAQUI   CCAPI   CCINC   CFLOR   CLECT   COBSC   COPFL   CQUIN   CSCUL   DGRIM
EOG8003Z0   1   1   1   1   1   1   1   1   1   2   1   1   2   1   1   1   1   1   1   1   1   1   8   1   4   1   1   1   1   1   1   2   1   2   1   1   3   1   1   1   1   1   1   4   1   1   1   1   2   6   2
EOG8003Z1   1   1   1   1   1   1   1   1   1   1   1   1   4   1   1   1   1   1   1   1   1   1   1   1   4   1   1   1   1   1   1   1   1   1   1   1   1   1   1   0   1   1   1   3   1   1   1   2   3   0   1
EOG8003Z2   0   0   0   0   0   0   0   0   0   0   0   0   10  1   0   0   0   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   1   0   2   0   0   3   0   0   0   0   1   0   0   0   0   0   0   0
EOG8003Z3   0   1   1   1   1   1   0   1   1   1   1   1   1   1   0   1   1   0   1   1   0   1   1   1   0   0   0   1   0   1   0   0   1   1   1   1   0   0   1   0   1   0   1   2   0   1   0   1   1   0   1
EOG8003Z4   0   0   1   1   1   0   0   1   1   1   1   1   1   1   1   1   0   0   1   1   1   1   0   0   2   1   1   1   1   1   1   1   1   1   1   1   1   1   0   1   1   1   1   1   3   2   1   1   1   0   2
EOG8003Z5   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   1   0   0   0   0   0   0   0   3   3   0   0   0   2   0   0   0   0   0   0   1   0   0   0   0   0   0   0   0   1   0   0   0   0
EOG8003Z6   0   0   2   1   1   0   0   0   1   0   0   1   2   1   0   1   1   0   1   1   0   1   0   0   2   0   0   1   0   0   0   1   1   0   1   0   0   0   1   0   2   0   1   0   1   1   0   0   0   1   1
EOG8003Z7   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   1   0   0   0   0   0   0   2   1   1   0   0   0   1   1   10  0   0   0   2   0   1   0   0   0   0   6   0   0   2   0   1   0   0
EOG8003Z8   1   1   1   1   1   1   1   1   1   1   2   1   3   1   2   1   2   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   2   1   2   2   1   1   1   1   1   1   2   1   1   3   1
EOG8003Z9   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   1   0   0   0   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
EOG8003ZB   0   0   0   0   0   0   0   0   0   0   0   0   1   0   0   0   0   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
EOG8003ZC   1   0   0   0   0   0   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
EOG8003ZD   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   1   1   0   0   0   1   0   0   0   0   0   0   1   0   0   0   0   0   0   0   0   0   0   0   0   0
```



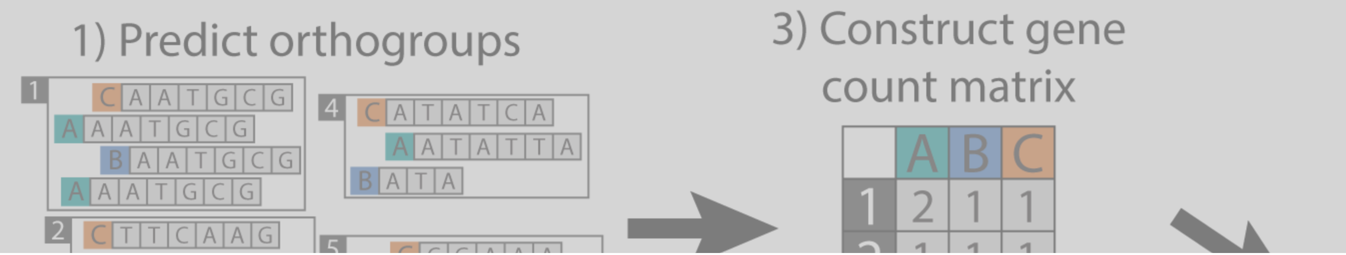
1) Predict orthogroups

3) Construct gene count matrix

	A	B	C
1	2	1	1
2	1	1	1
3	1	1	1
4	1	1	1
5	2	2	1
6	1	0	1

2) Infer time tree

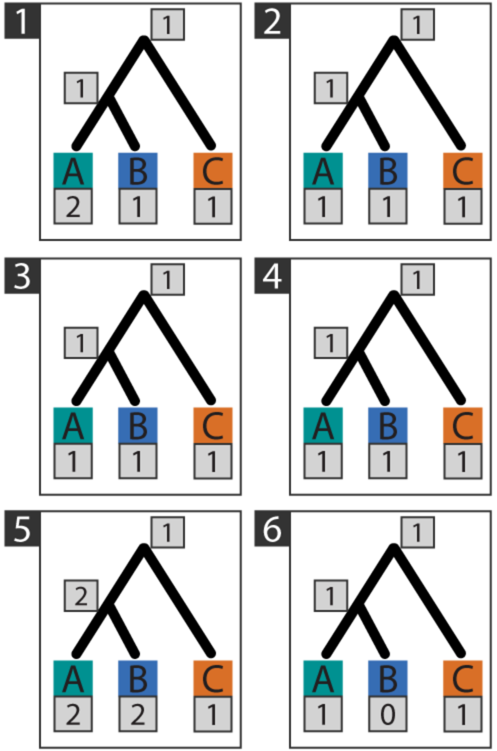
4) Infer ancestral gene counts



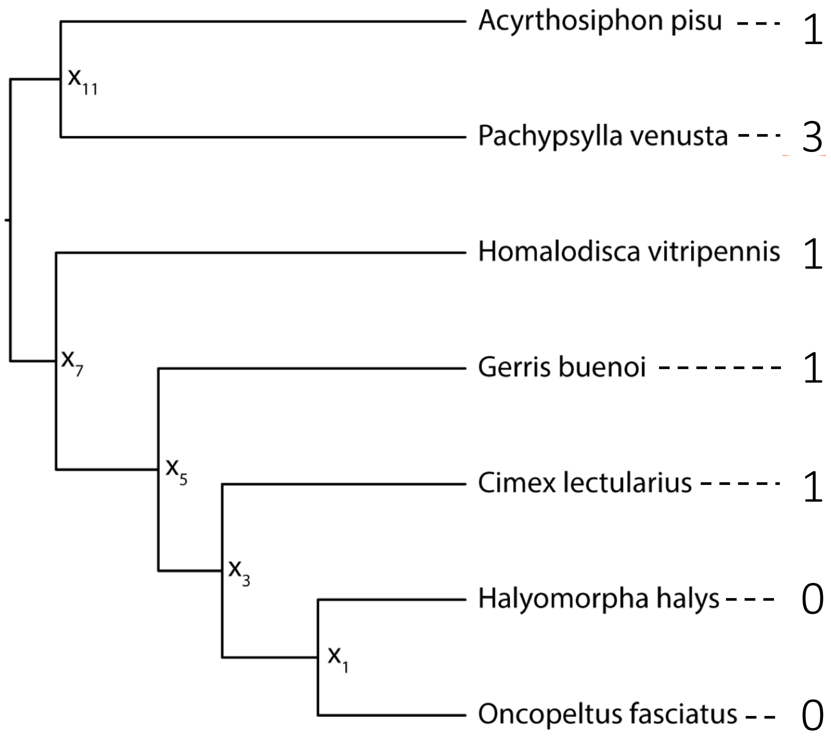
Ancestral gene counts inferred with:

1. Maximum likelihood (CAFE) for the 6 multi-species orders
2. Parsimony (Dupliphy) for all nodes

4) Infer ancestral gene counts



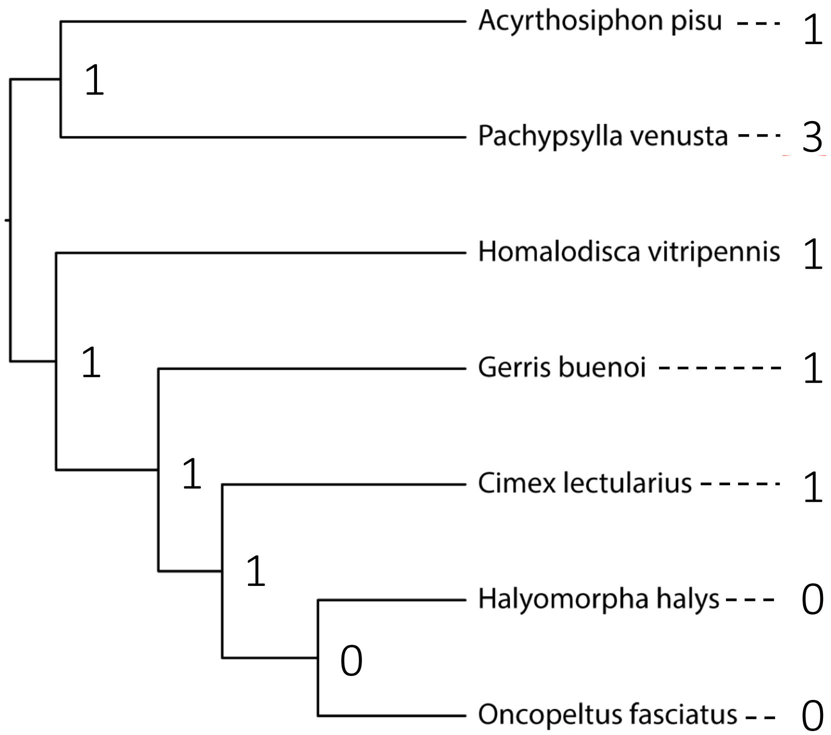
Ancestral gene counts: Example


Acyrthosiphon pisu - - - 1
x11
Pachypsylla venusta - - - 3
Homalodisca vitripennis 1
x7
Gerris buenoi - - - - - - - 1
x5
Cimex lectularius - - - - - 1
x3
Halyomorpha halys - - - 0
x1
Oncopeltus fasciatus - - 0

Tips: observed variables

x_i: hidden variables

Ancestral gene counts: Example

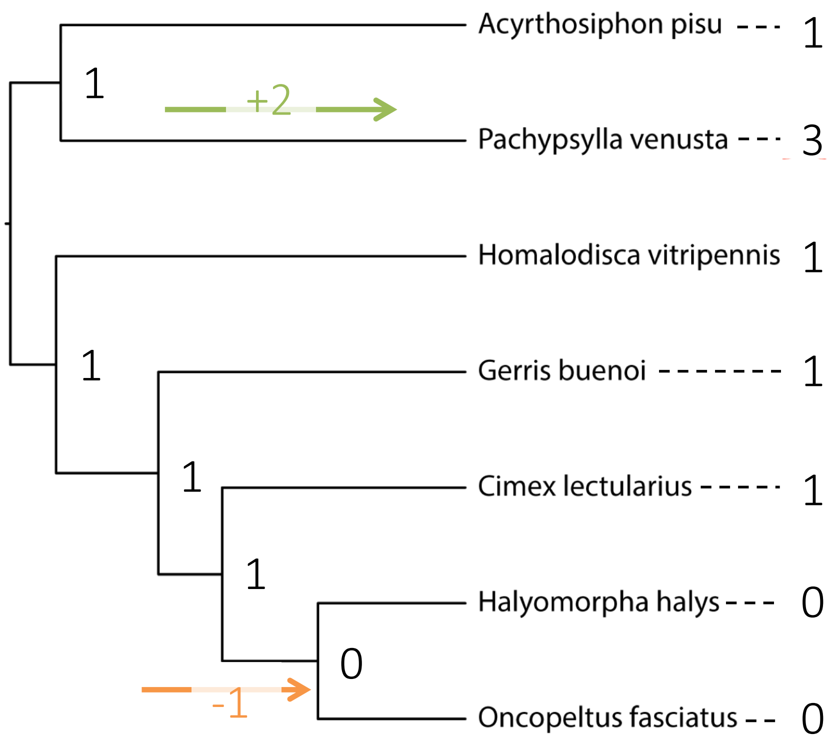

Acyrthosiphon pisu --- 1
1
Pachypsylla venusta --- 3
Homalodisca vitripennis 1
1
Gerris buenoi ------- 1
1
Cimex lectularius ----- 1
1
Halyomorpha halys --- 0
0
Oncopeltus fasciatus -- 0

Tips: observed variables

x_i: hidden variables

Our goal is to infer the states of the internal nodes of the tree

Ancestral gene counts: Example



Tips: observed variables
x_i: hidden variables

Our goal is to infer the states of the internal nodes of the tree

Then we can count changes along each lineage

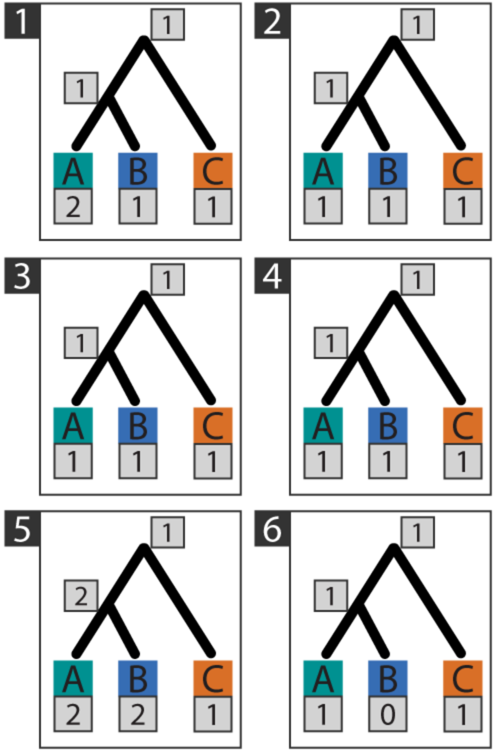
1) Predict orthogroups

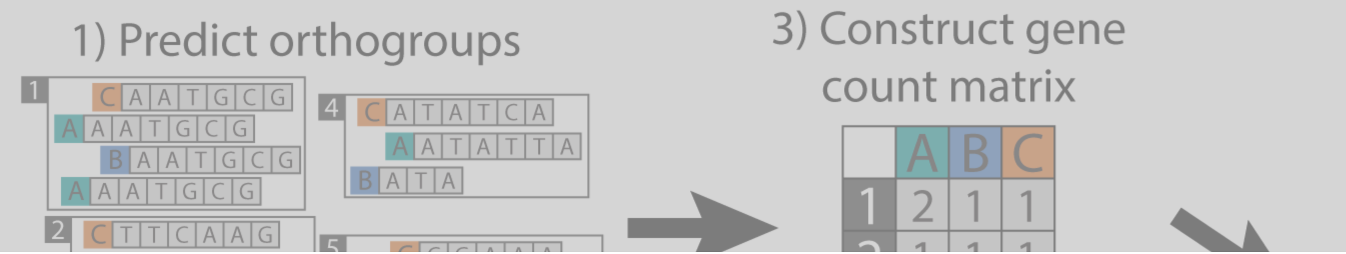
3) Construct gene count matrix

With ancestral gene counts we can:

1. Infer rates of gene gain/loss
2. Count gene gains and losses and check for rapid changes on every lineage
3. Estimate gene counts in extinct ancestors

4) Infer ancestral gene counts

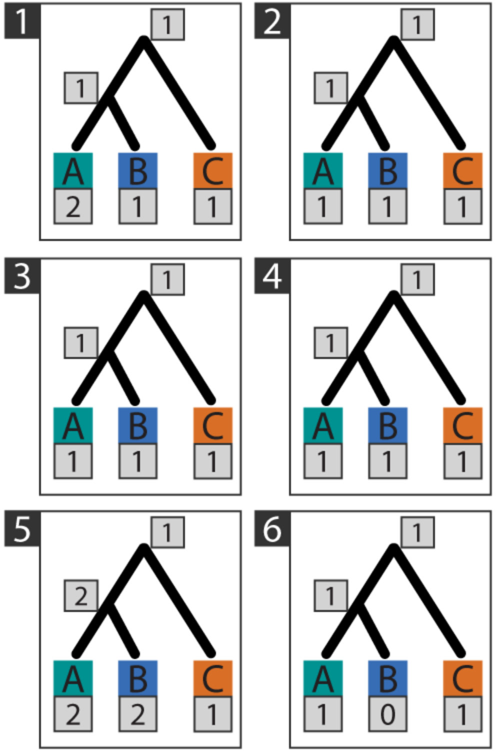




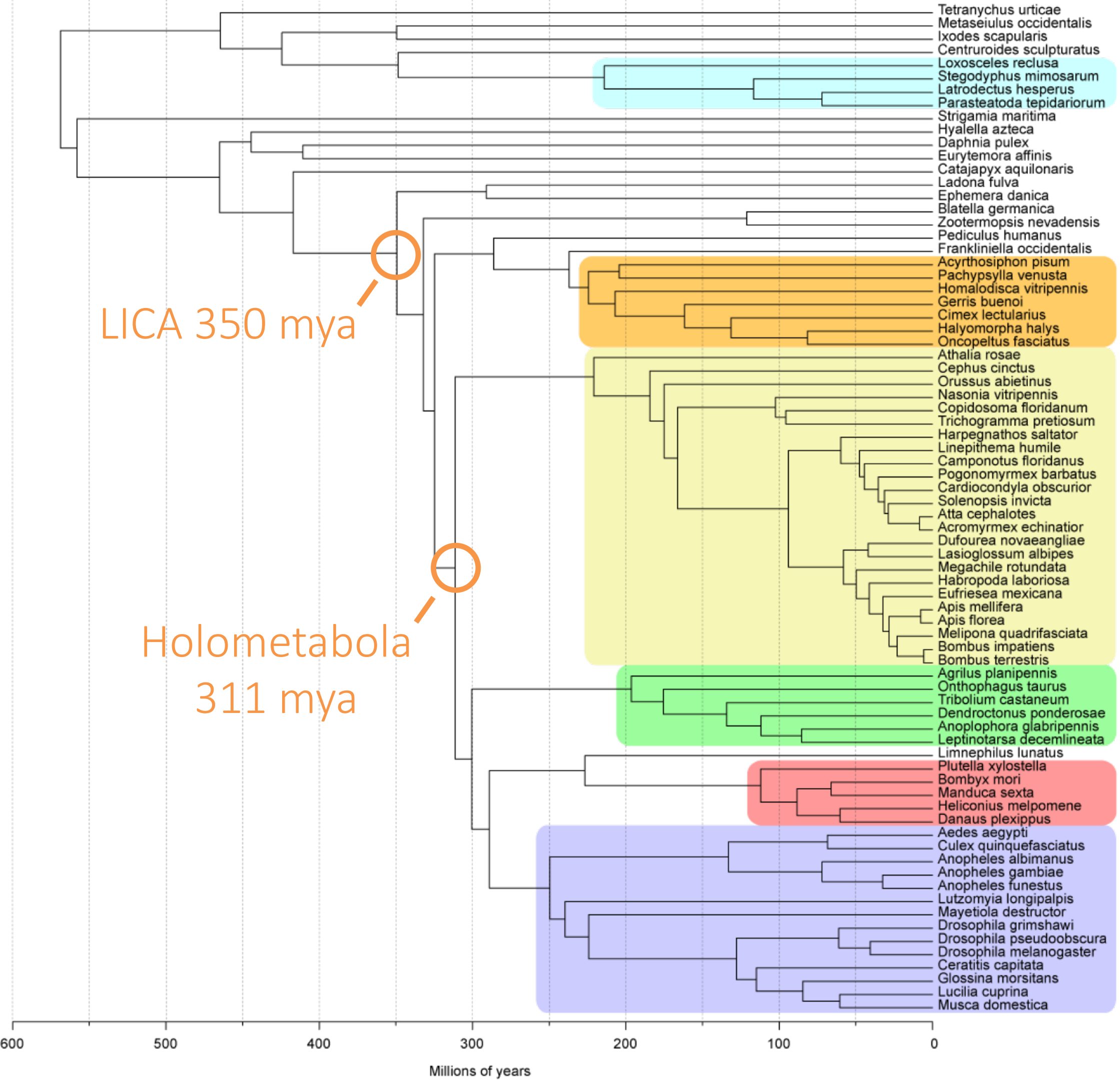
With ancestral gene counts we can:

1. Infer rates of gene gain/loss
2. Count gene gains and losses and check for rapid changes on every lineage
3. Estimate gene counts in extinct ancestors

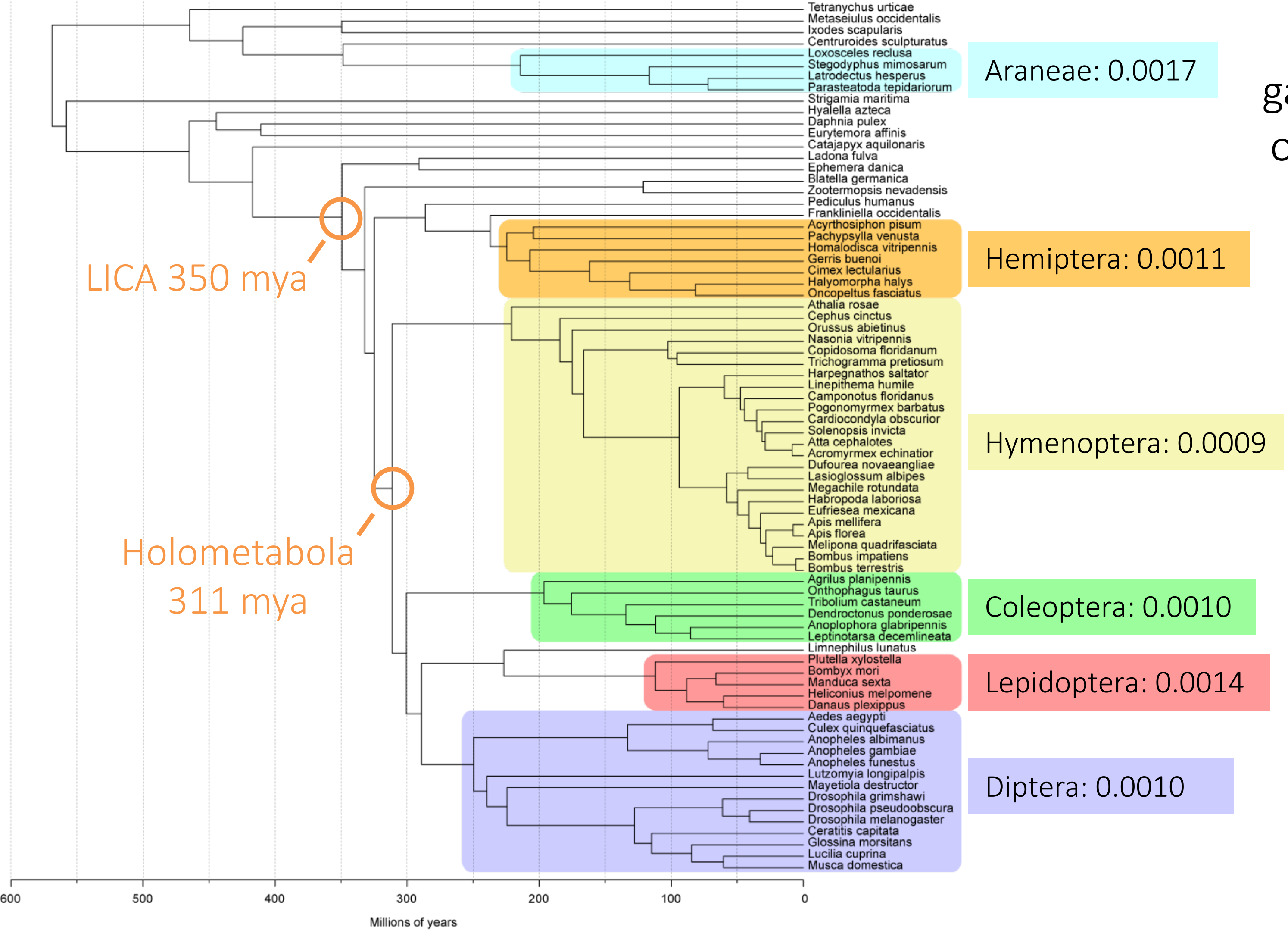
4) Infer ancestral gene counts



Arthropod Time Tree



Rates of gene gain/loss between orders are largely consistent



Araneae: 0.0017

Hemiptera: 0.0011

Hymenoptera: 0.0009

Coleoptera: 0.0010

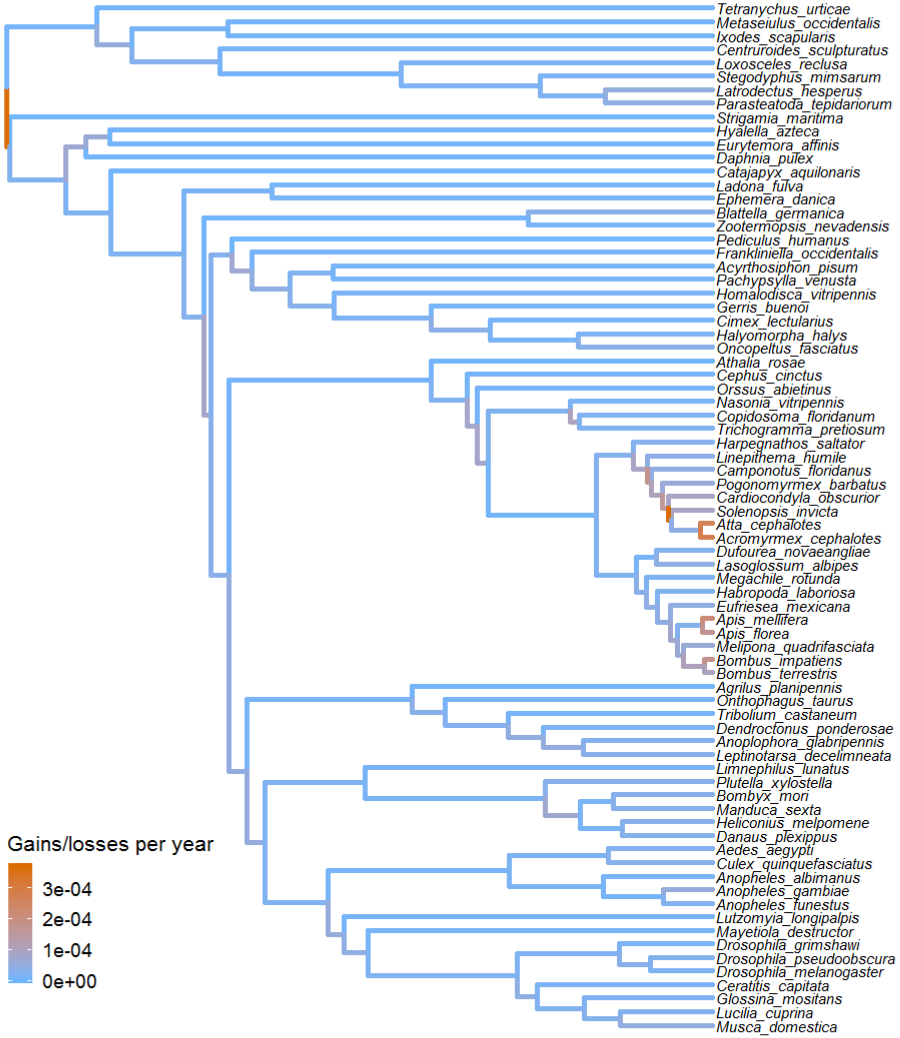
Lepidoptera: 0.0014

Diptera: 0.0010

LICA 350 mya

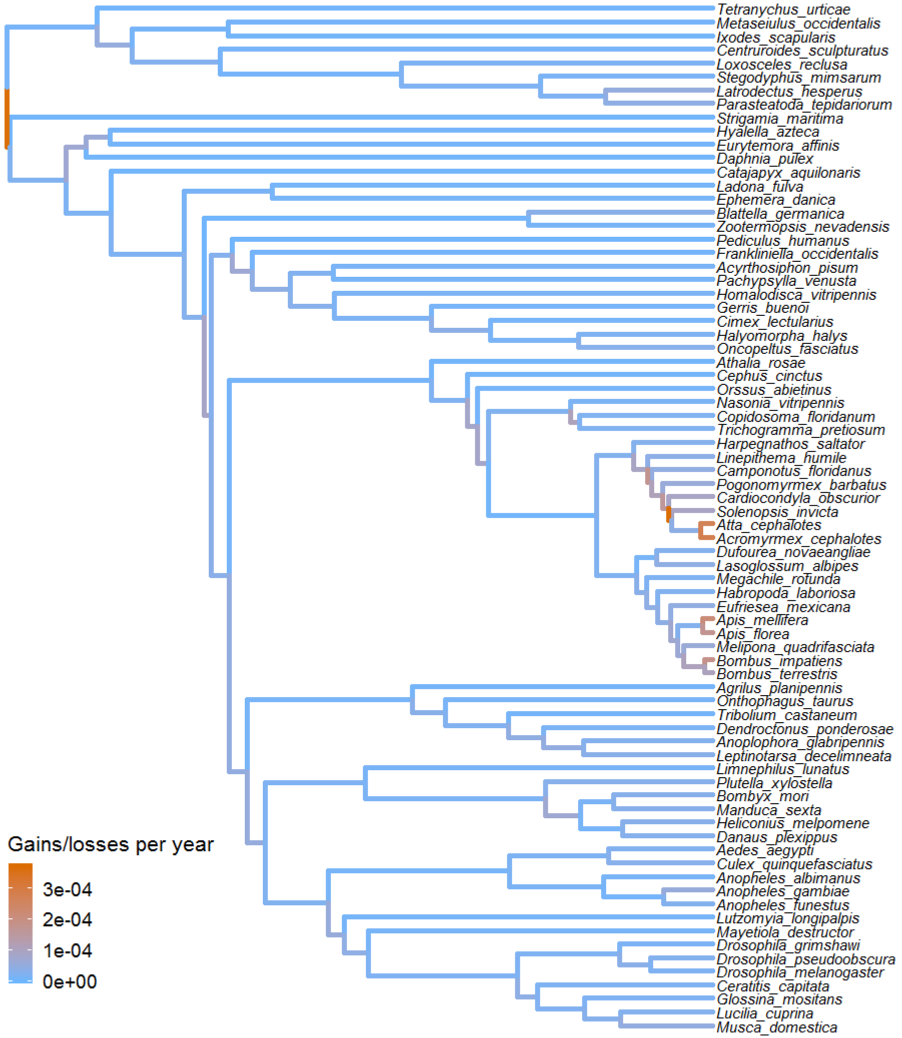
Holometabola 311 mya

Arthropod Time Tree



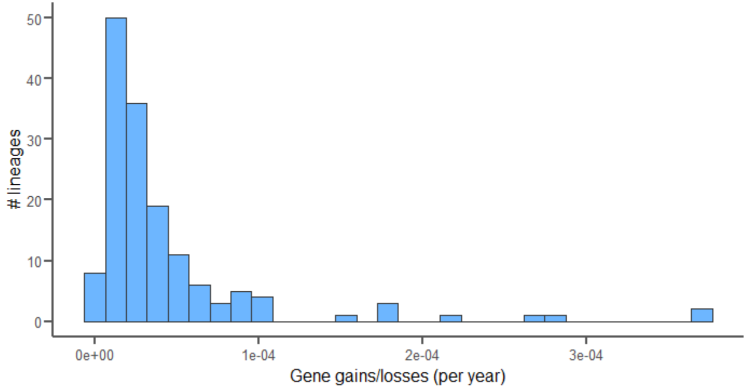
The branches of the ML tree can be scaled by time to infer lineage specific gain/loss rates

Arthropod Time Tree

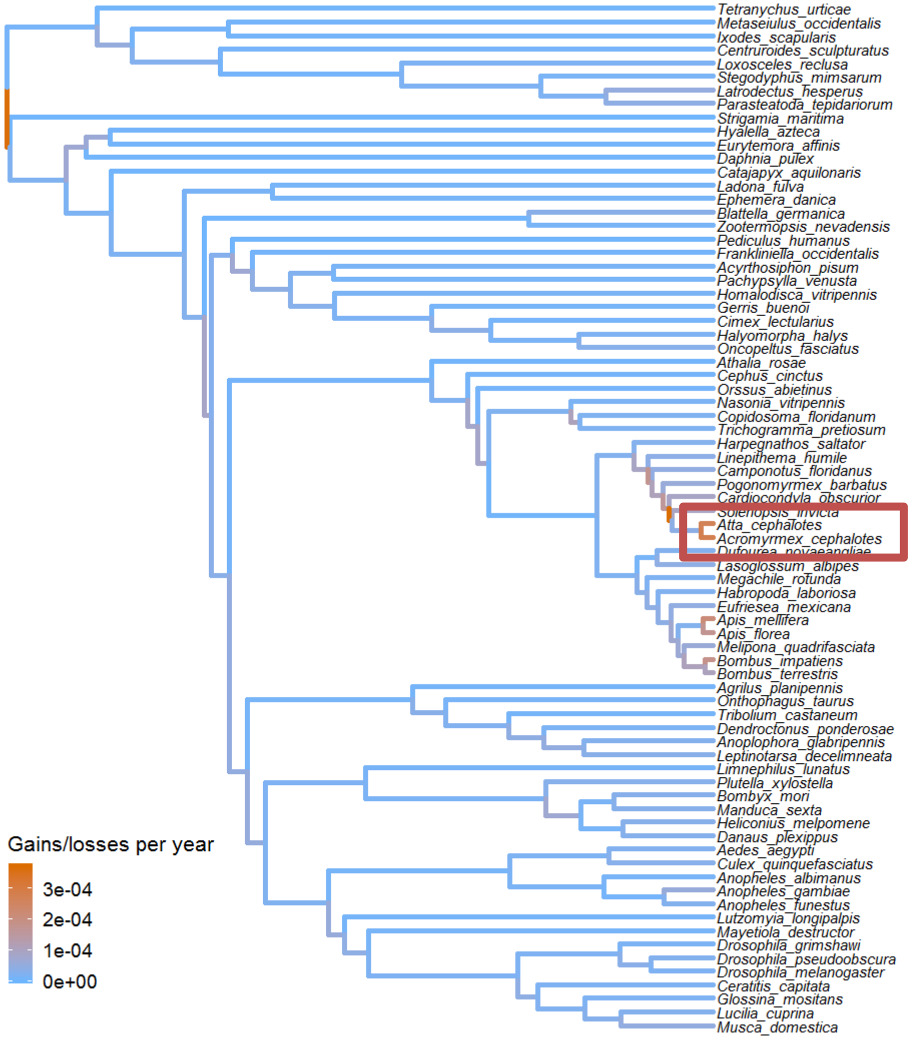


The branches of the ML tree can be scaled by time to infer lineage specific gain/loss rates

Rates are mostly consistent across arthropods

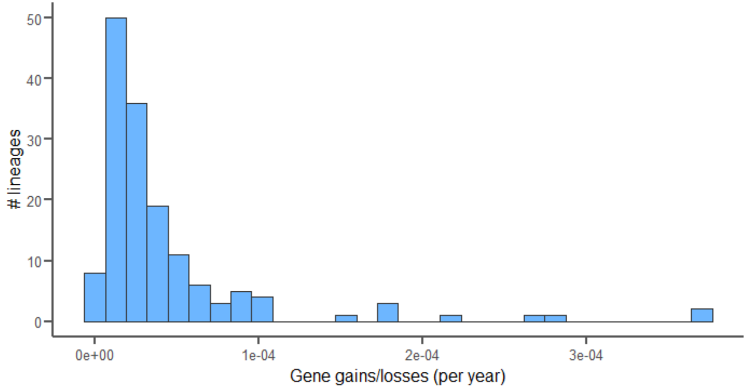


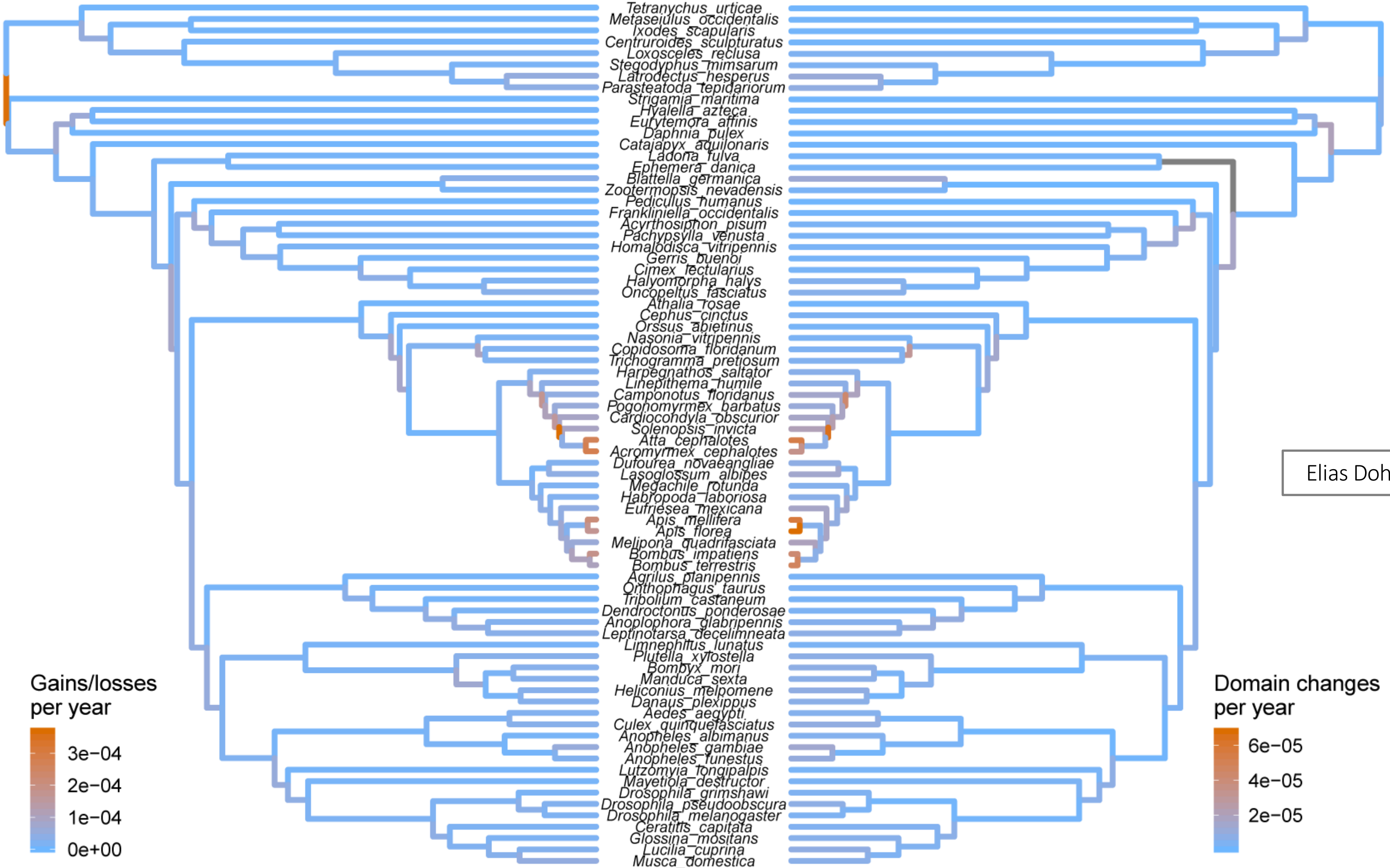
Arthropod Time Tree



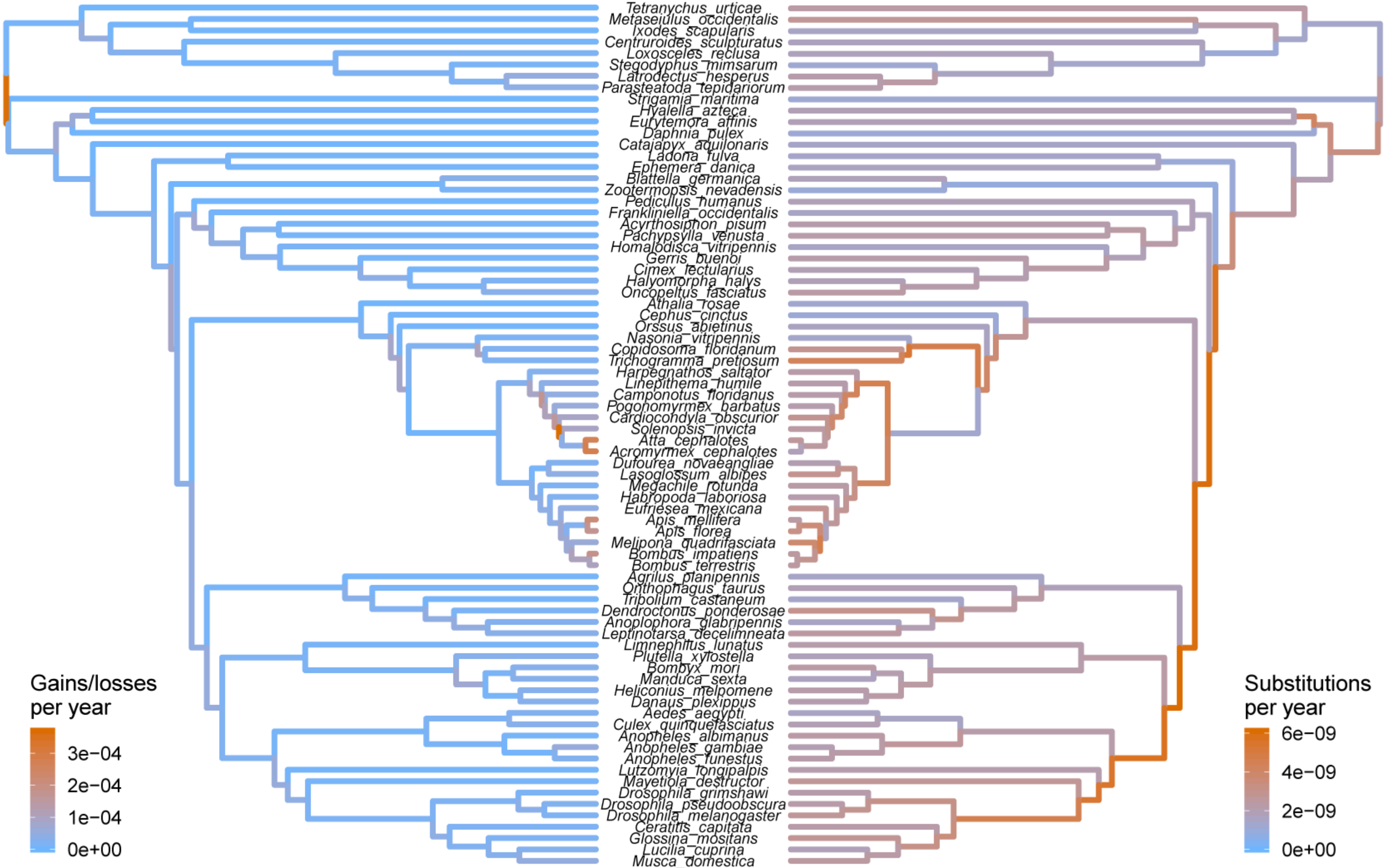
The branches of the ML tree can be scaled by time to infer lineage specific gain/loss rates

Accelerated rates in leafcutter ants

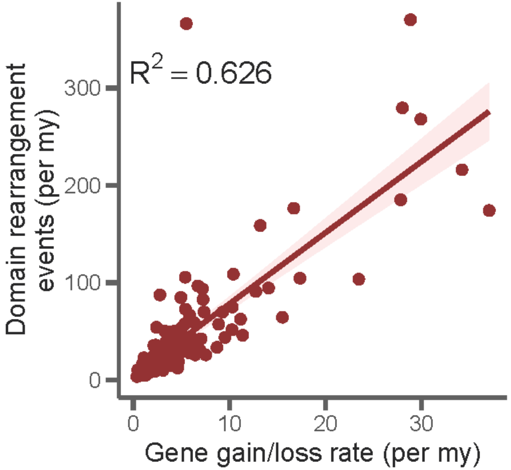




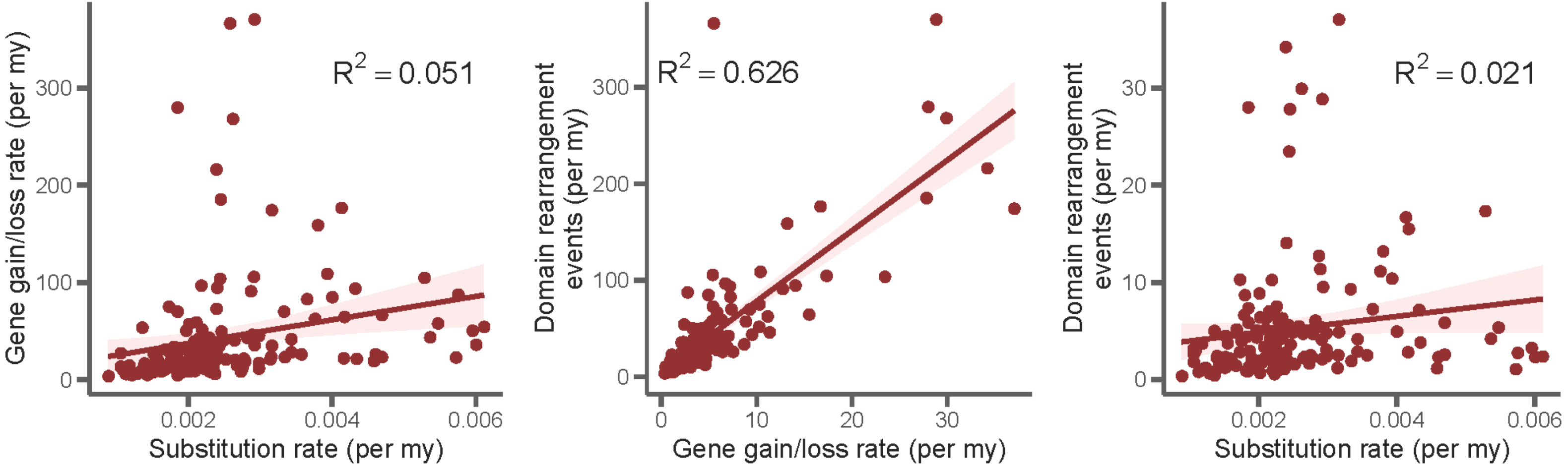
Elias Dohmen

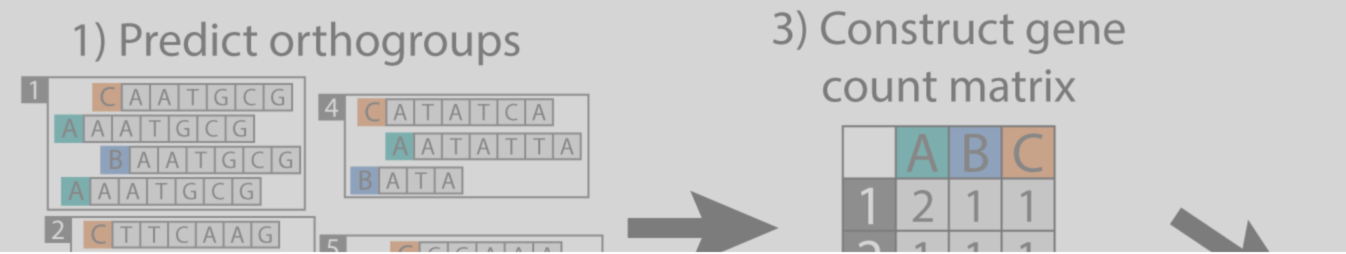

Tetranychus_urticae
Metaseiulus_occidentalis
Ixodes_scapularis
Centruroides_sculpturatus
Loxosceles_reclusa
Stegodyphus_mimsarum
Latrodectus_hesperus
Parasteatoda_tepidariorum
Strigamia_maritima
Hyalella_azteca
Eurytemora_affinis
Daphnia_pulex
Catajapyx_aquilonaris
Ladona_fulva
Ephemera_danica
Blattella_germanica
Zootermopsis_nevadensis
Pediculus_humanus
Frankliniella_occidentalis
Acyrthosiphon_pisum
Pachypsylla_venusta
Homalodisca_vitripennis
Gerris_buenoi
Cimex_lectularius
Halyomorpha_halys
Oncopeltus_fasciatus
Athalia_rosae
Cephus_cinctus
Orssus_abietinus
Nasonia_vitripennis
Copidosoma_floridanum
Trichogramma_pretiosum
Harpegnathos_saltator
Linepithema_humile
Camponotus_floridanus
Pogonomyrmex_barbatus
Cardiocondyla_obscurior
Solenopsis_invicta
Atta_cephalotes
Acromyrmex_cephalotes
Dufourea_novaeangliae
Lasioglossum_albipes
Megachile_rotunda
Habropoda_laboriosa
Eufriesea_mexicana
Apis_mellifera
Apis_florea
Melipona_quadrifasciata
Bombus_impatiens
Bombus_terrestris
Agrilus_planipennis
Onthophagus_taurus
Tribolium_castaneum
Dendroctonus_ponderosae
Anoplophora_glabripennis
Leptinotarsa_decelimneata
Limnephilus_lunatus
Plutella_xylostella
Bombyx_mori
Manduca_sexta
Heliconius_melpomene
Danaus_plexippus
Aedes_aegypti
Culex_quinquefasciatus
Anopheles_albimanus
Anopheles_gambiae
Anopheles_funestus
Lutzomyia_longipalpis
Mayetiola_destructor
Drosophila_grimshawi
Drosophila_pseudoobscura
Drosophila_melanogaster
Ceratitis_capitata
Glossina_morsitans
Lucilia_cuprina
Musca_domestica
Gains/losses per year
3e−04
2e−04
1e−04
0e+00
Substitutions per year
6e−09
4e−09
2e−09
0e+00

Gene gain and loss rates are correlated with protein domain rearrangements



Neither are correlated with substitution rate



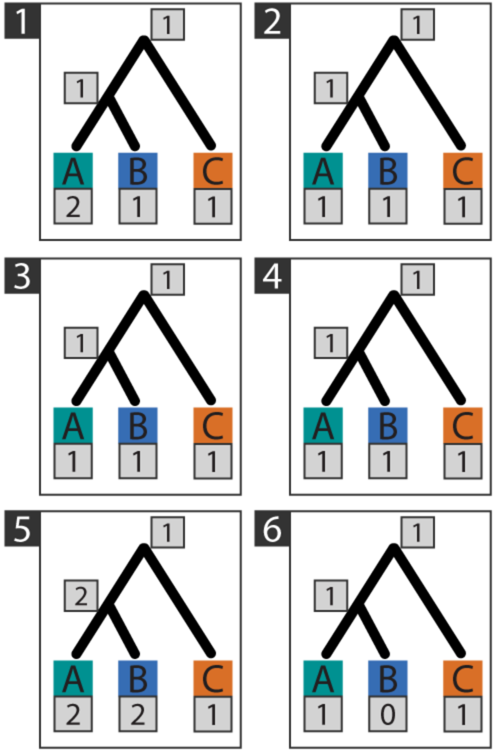

1) Predict orthogroups

3) Construct gene count matrix

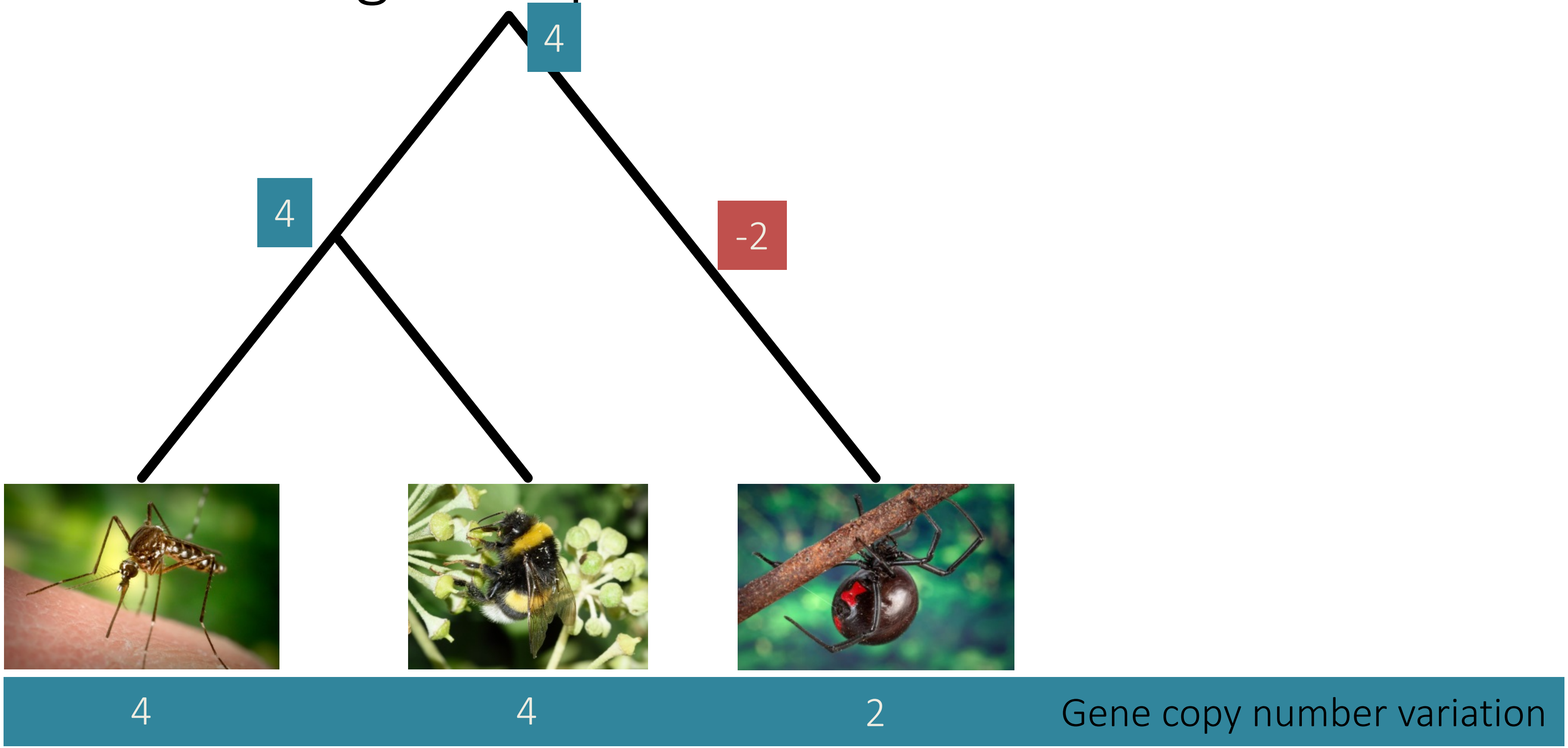
With ancestral gene counts we can:

1. Infer rates of gene gain/loss
2. Count gene gains and losses and check for rapid changes on every lineage
3. Estimate gene counts in extinct ancestors

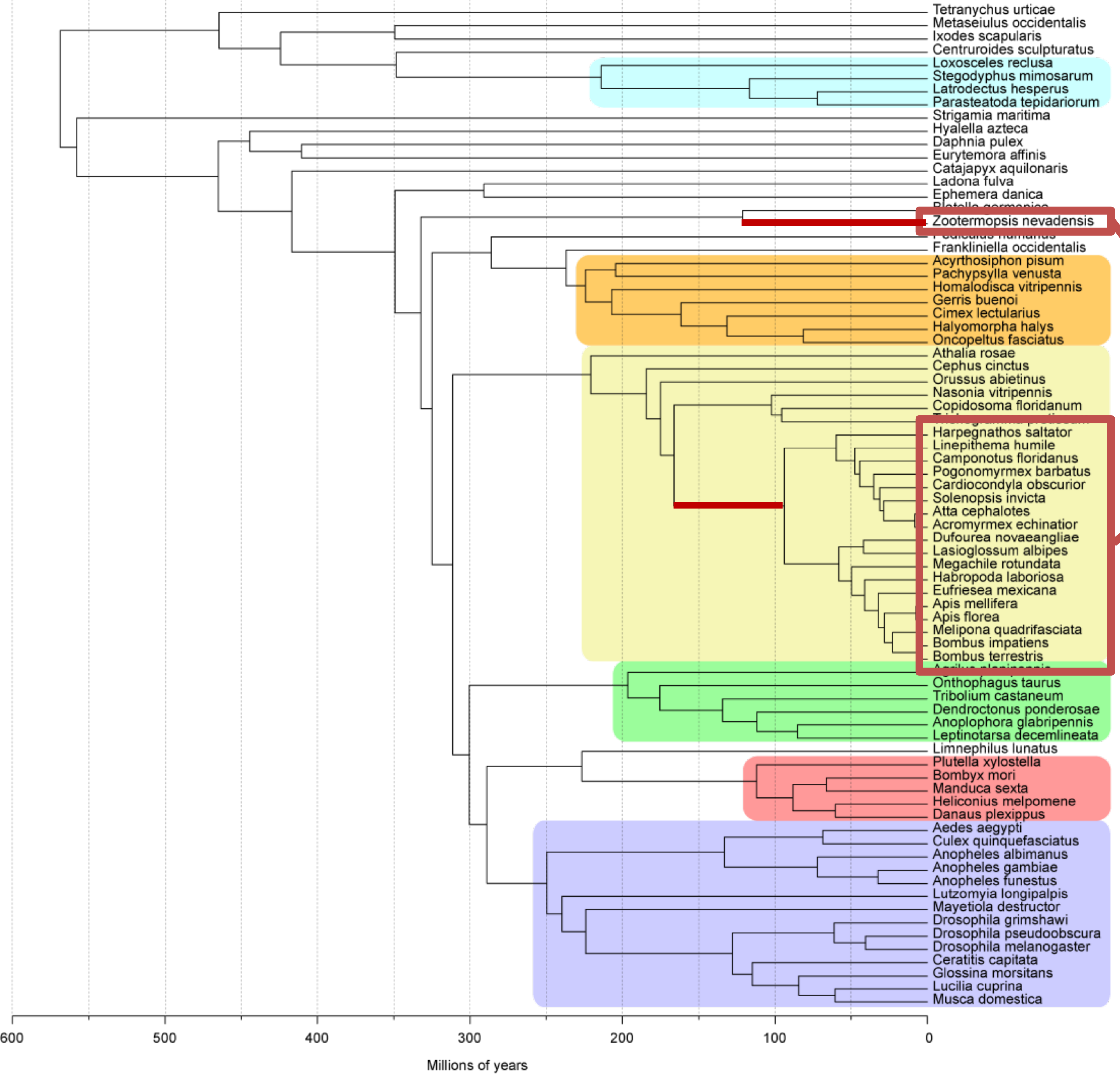
4) Infer ancestral gene counts



What specific gene family changes are interesting or important?



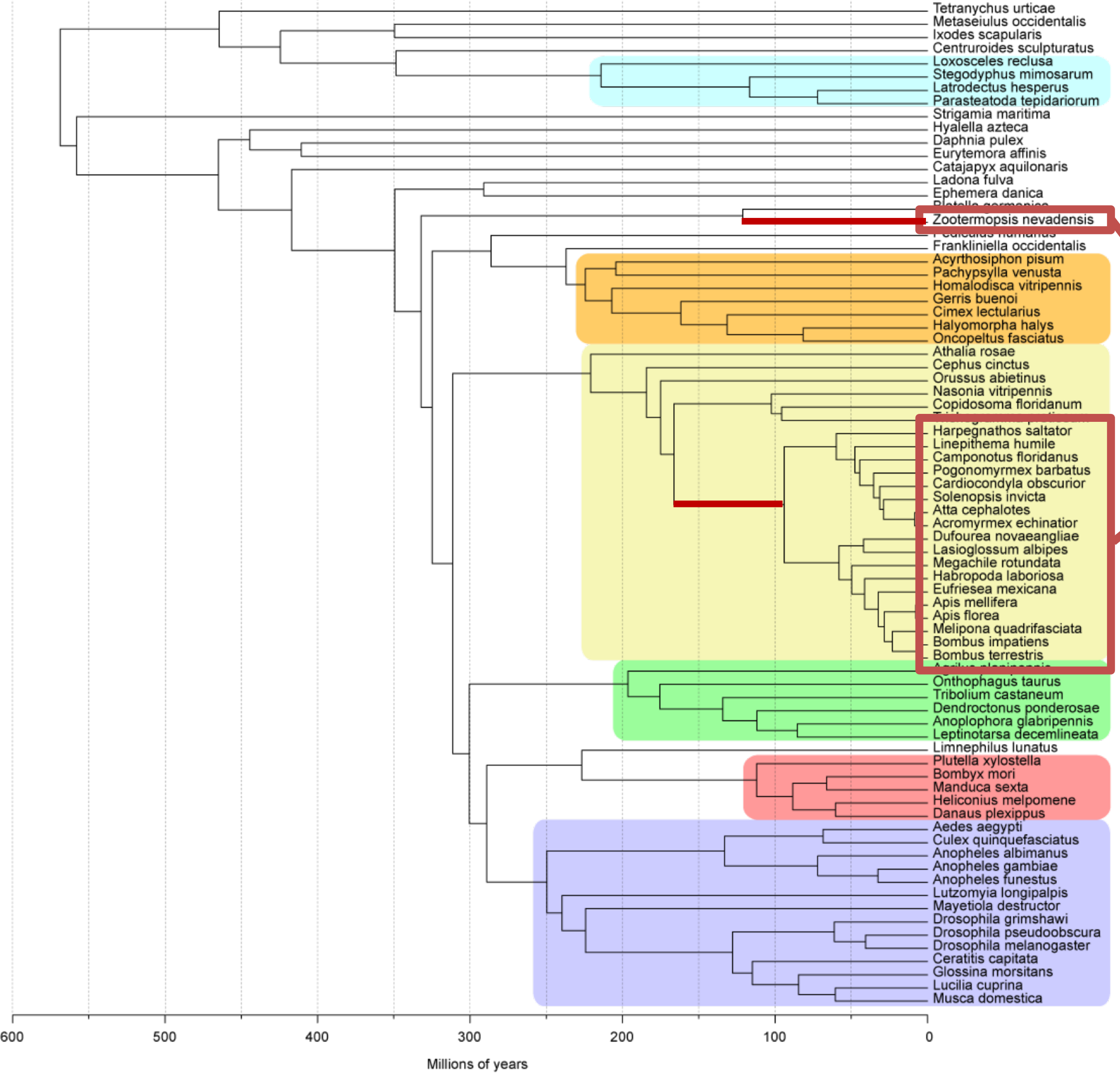
Common gene family changes among eusocial insects



Termites

Bees & ants

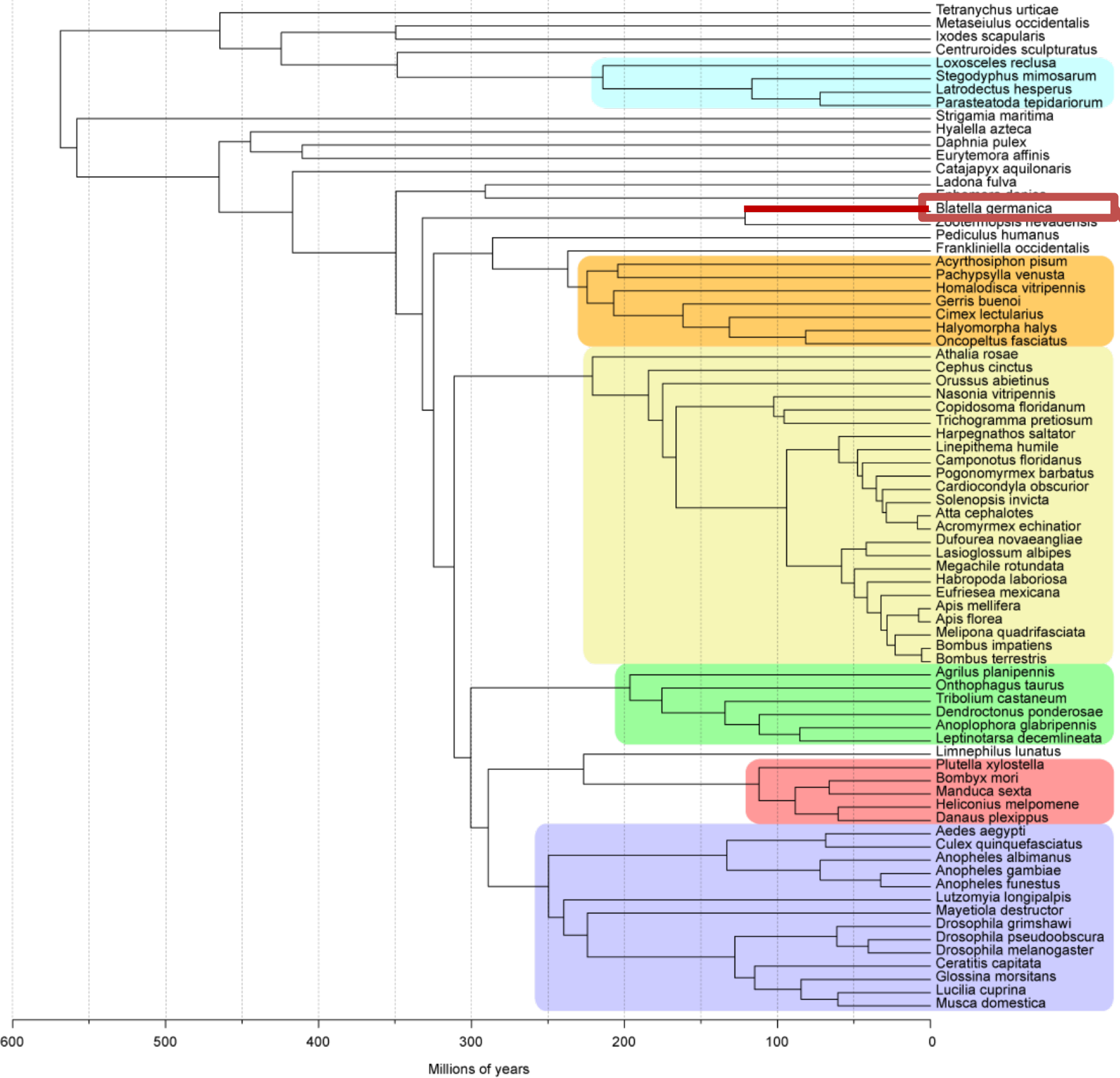
Common gene family changes among eusocial insects



Termites

Bees & ants

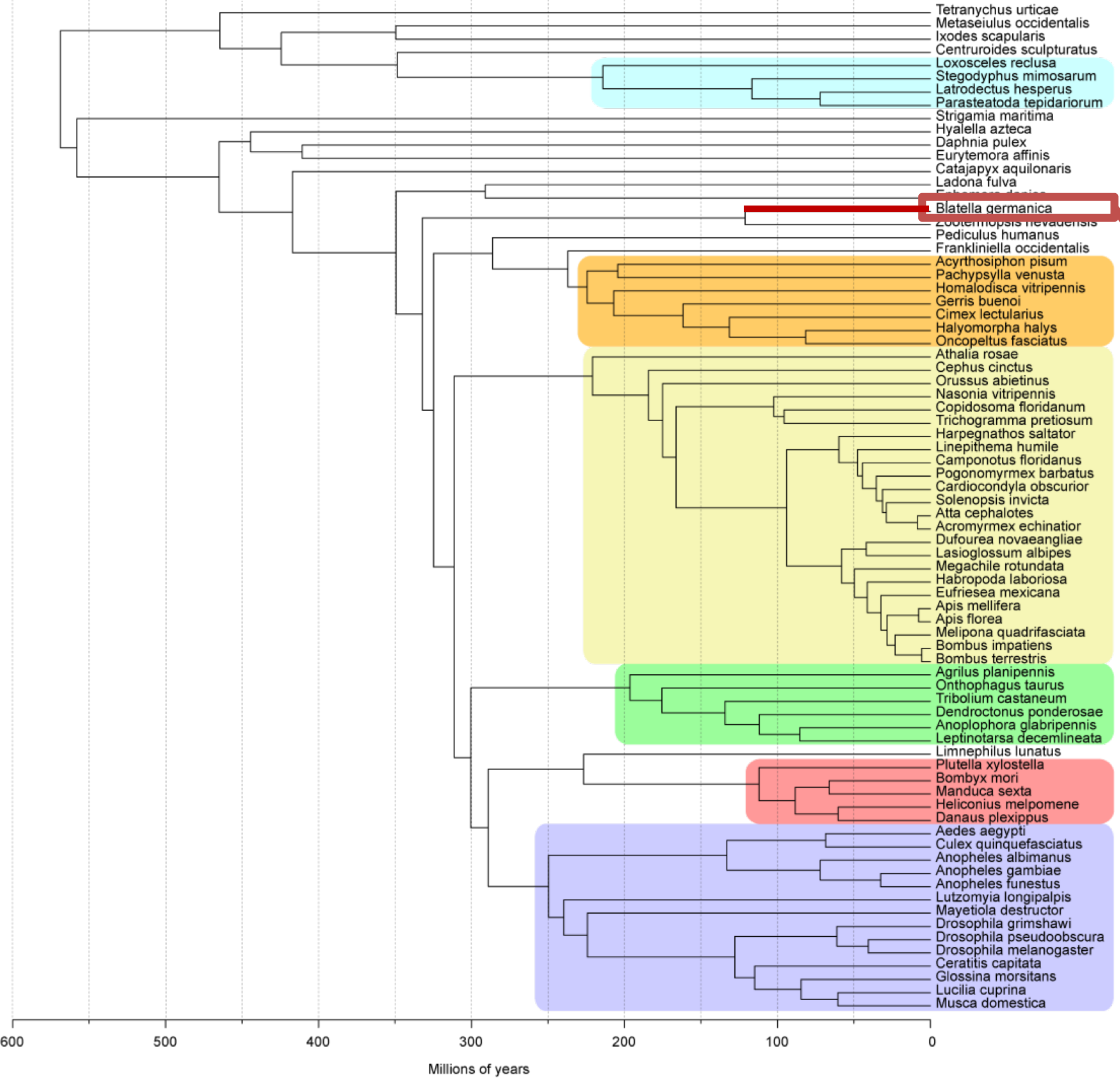
- 41 enriched functional terms in BOTH groups
- Olfactory reception and odorant binding



Largest number of gene family changes

German cockroach

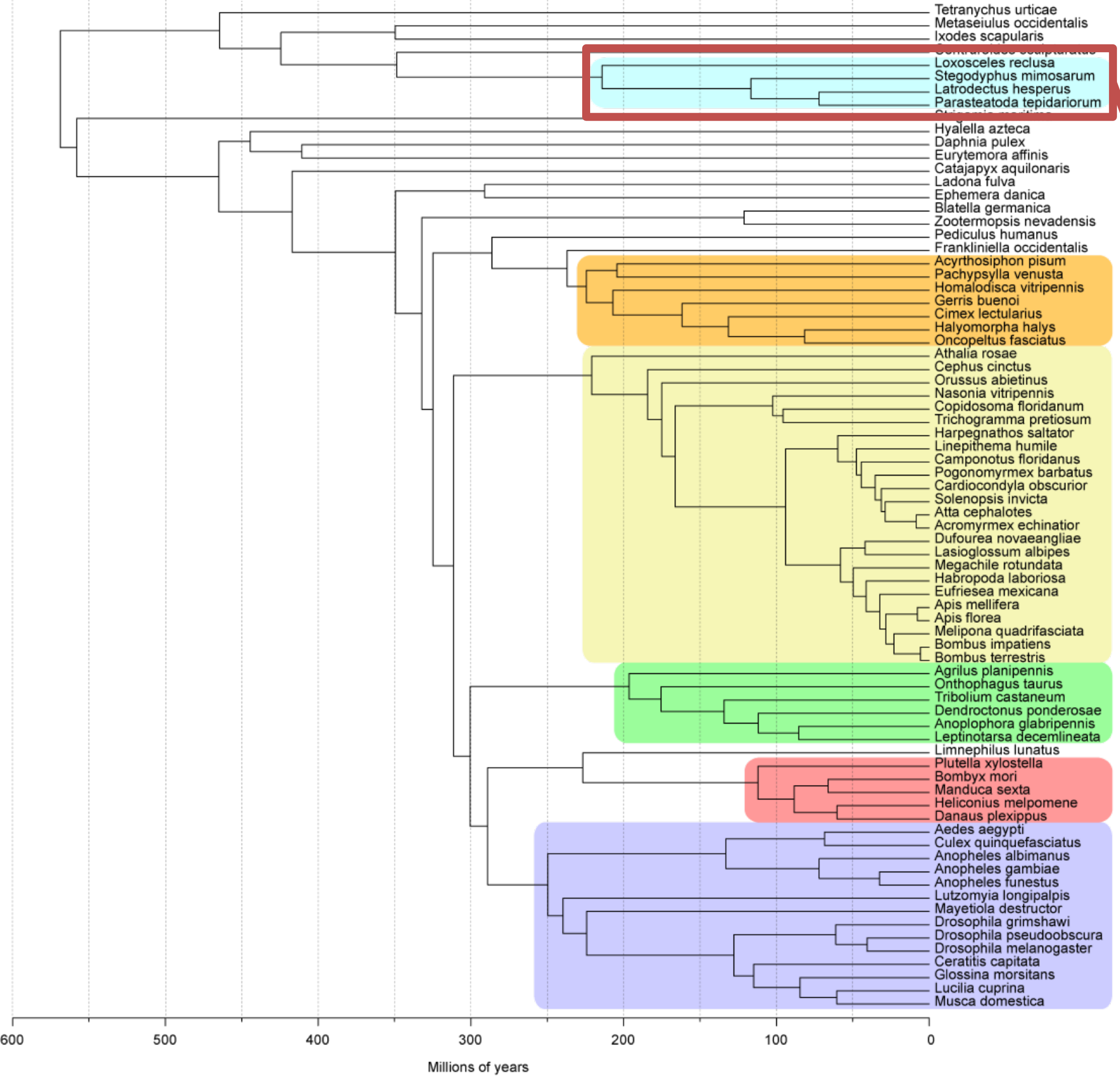
Largest number of gene family changes



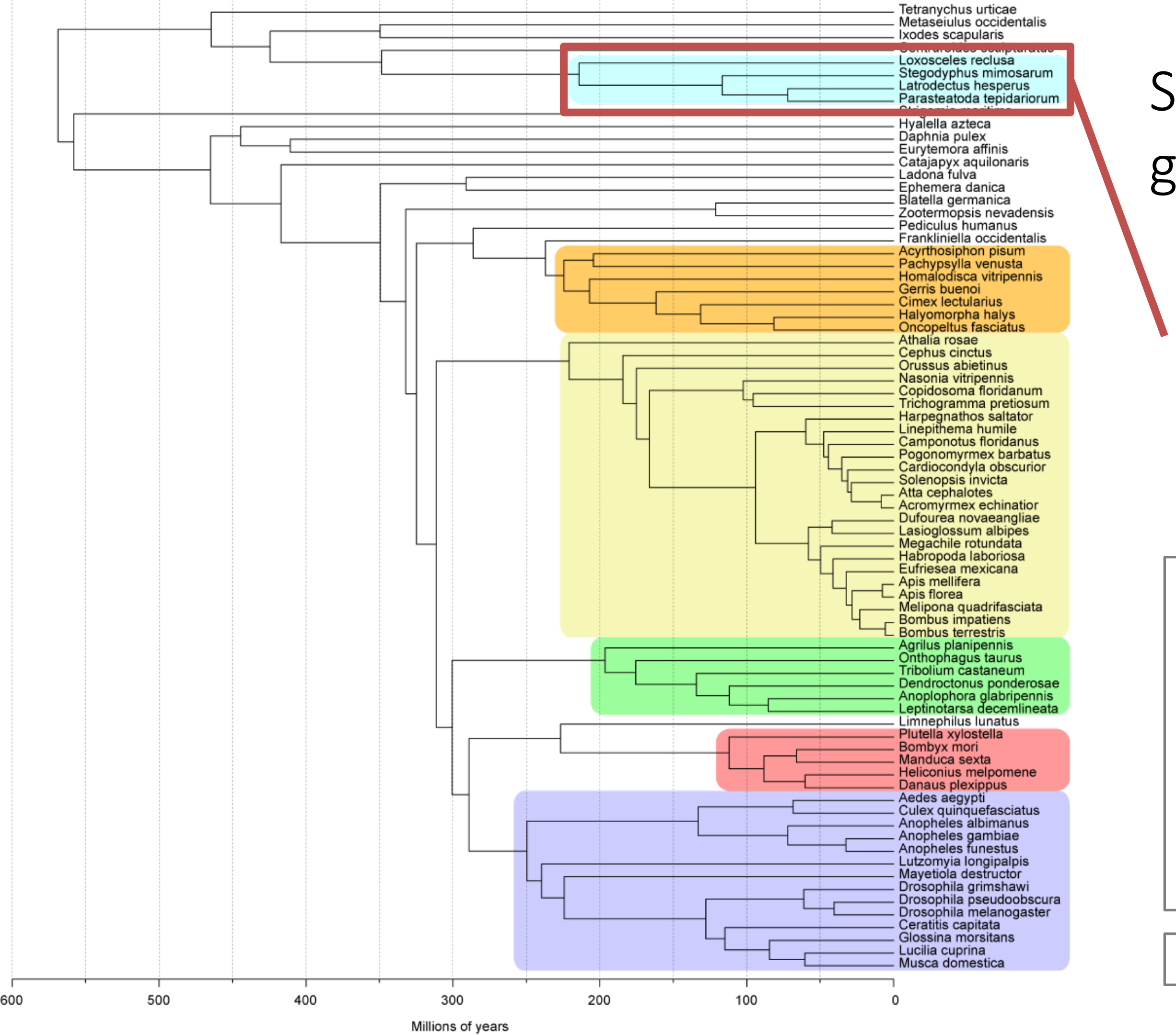
German cockroach

- Most RAPID gene family changes
- Major expansion of chemosensory genes
- Most protein domain rearrangements

Spider silk and venom gene families



Araneae

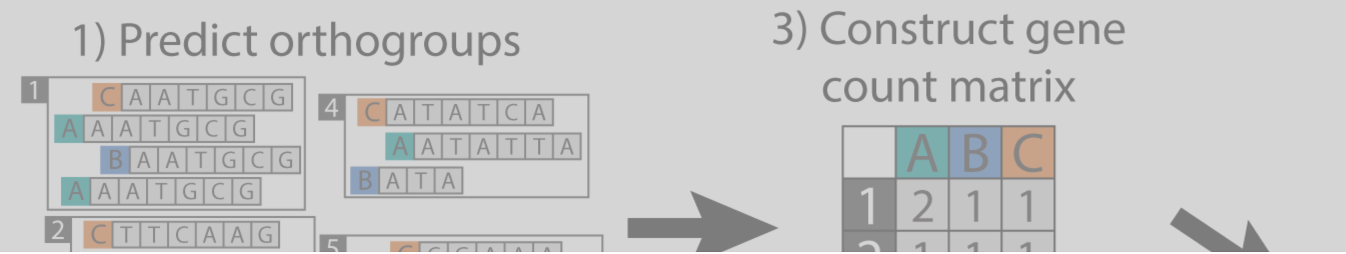


Spider silk and venom gene families

Araneae

- 10 rapidly expanding gene families within Araneae related for silk or venom
- High rate of protein domain emergences, including some related to venom

Jessica Garb

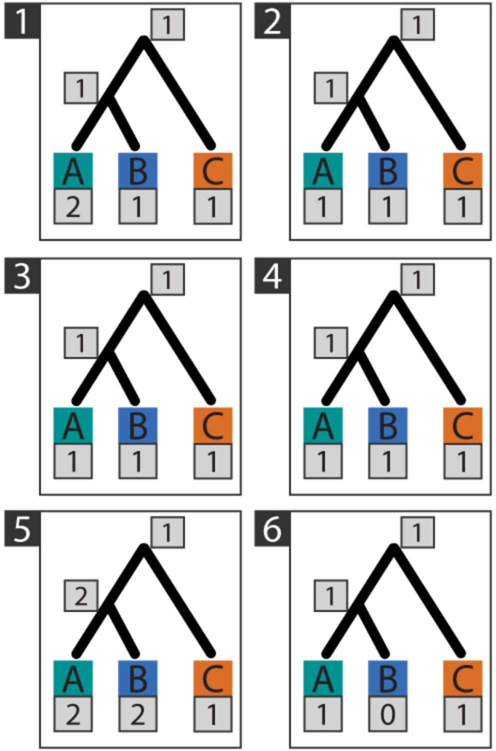

1) Predict orthogroups

3) Construct gene count matrix

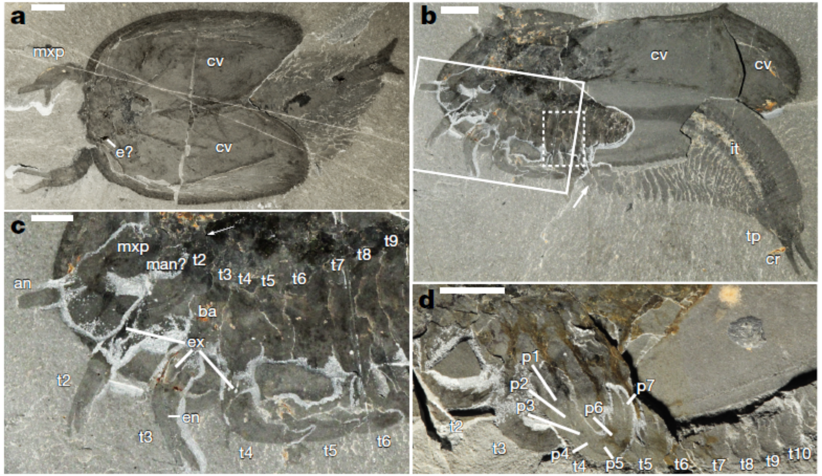
With ancestral gene counts we can:

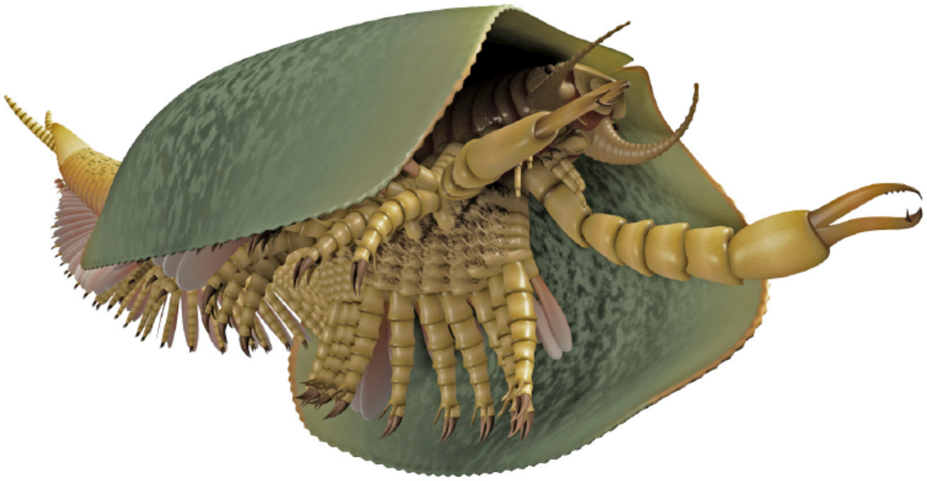
1. Infer rates of gene gain/loss
2. Count gene gains and losses and check for rapid changes on every lineage
3. Estimate gene counts in extinct ancestors

4) Infer ancestral gene counts



What did the ancestral insect (LICA) look like?



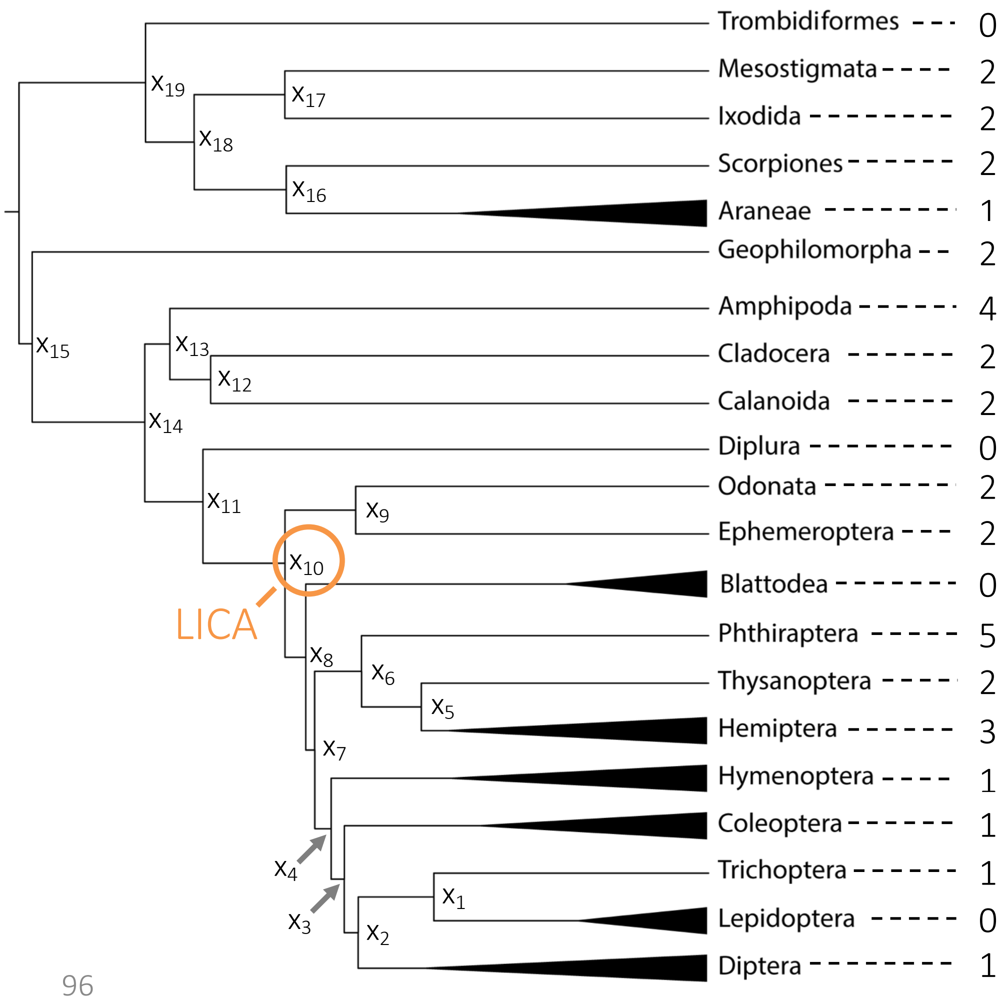


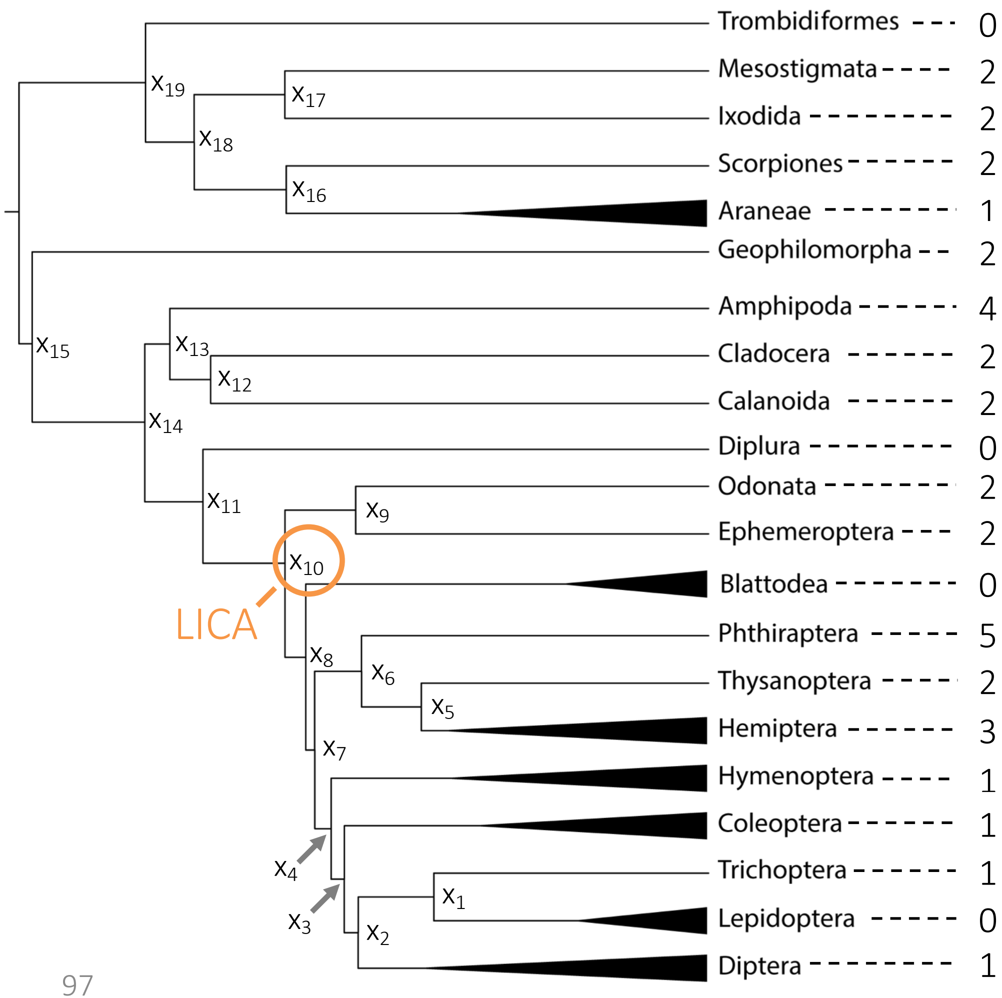
Burgess Shale fossils illustrate the origin of the mandibulate body plan

Cédric Aria[1,2]† & Jean-Bernard Caron[1,2,3]

NATURE | VOL 545 | 4 MAY 2017

How can we infer characteristics of the genome of LICA?

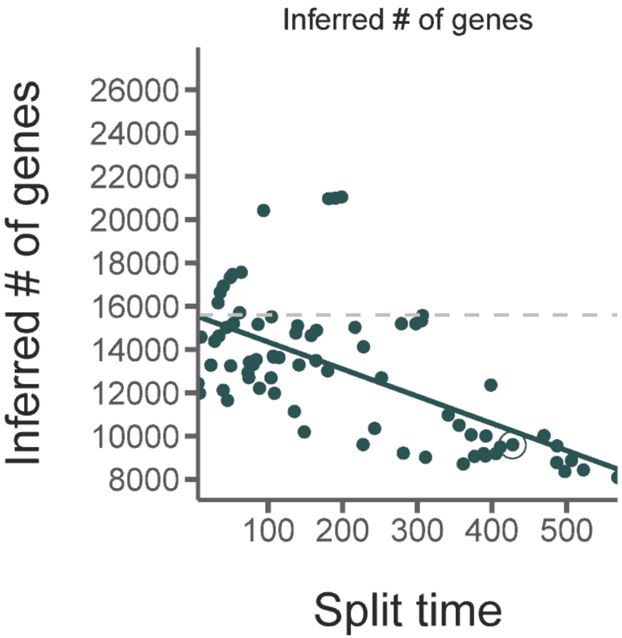




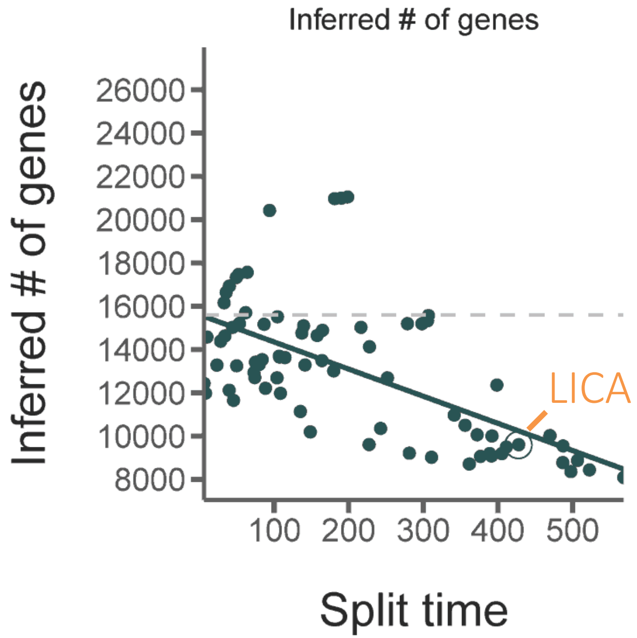
How can we infer characteristics of the genome of LICA?

How many genes were present in the LICA genome?

Ancestral genome sizes are underestimated due to extinctions

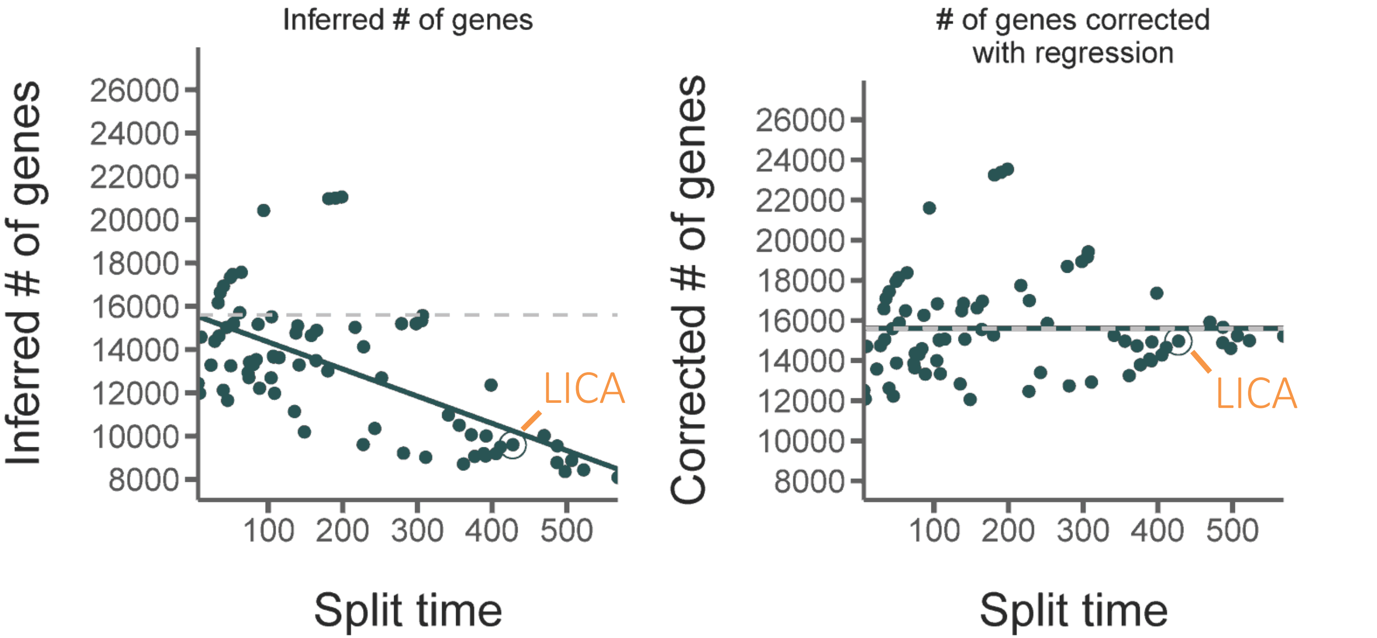


Ancestral genome sizes are underestimated due to extinctions



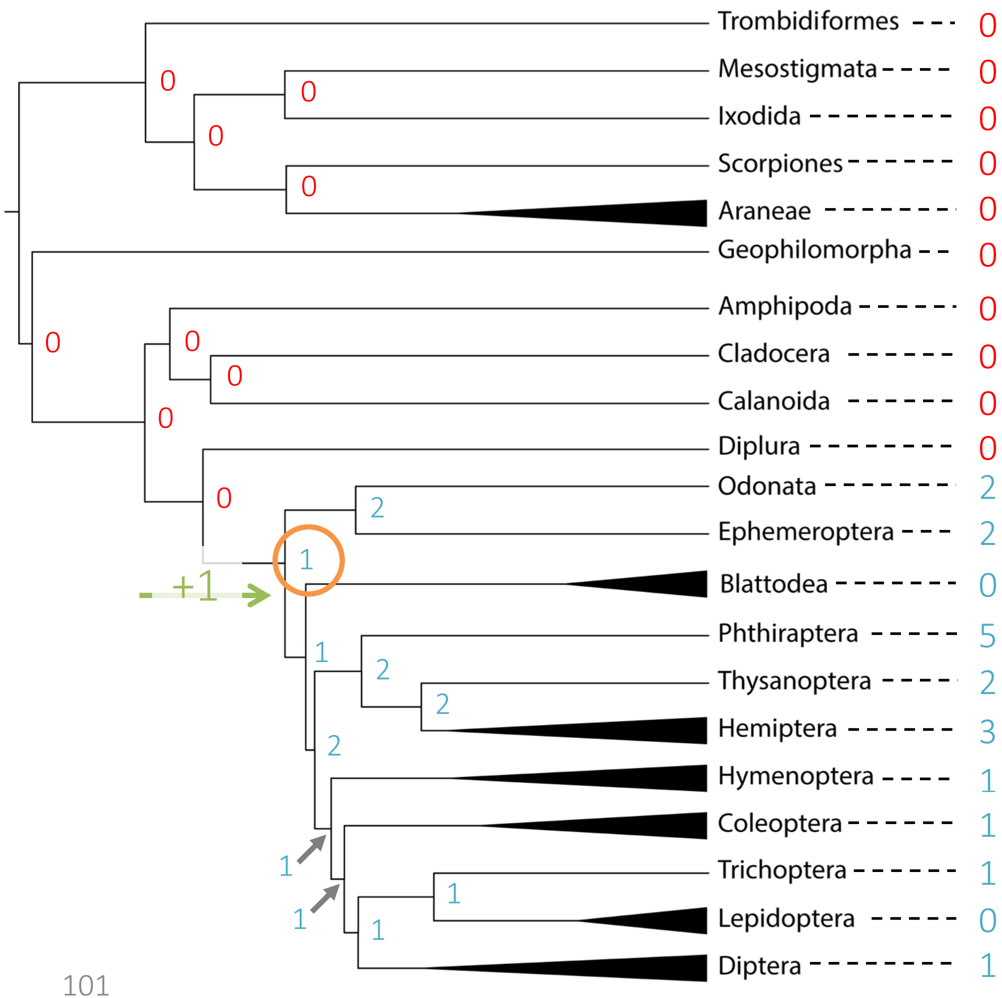
Estimated: 9,601 genes

Ancestral genome sizes are underestimated due to extinctions



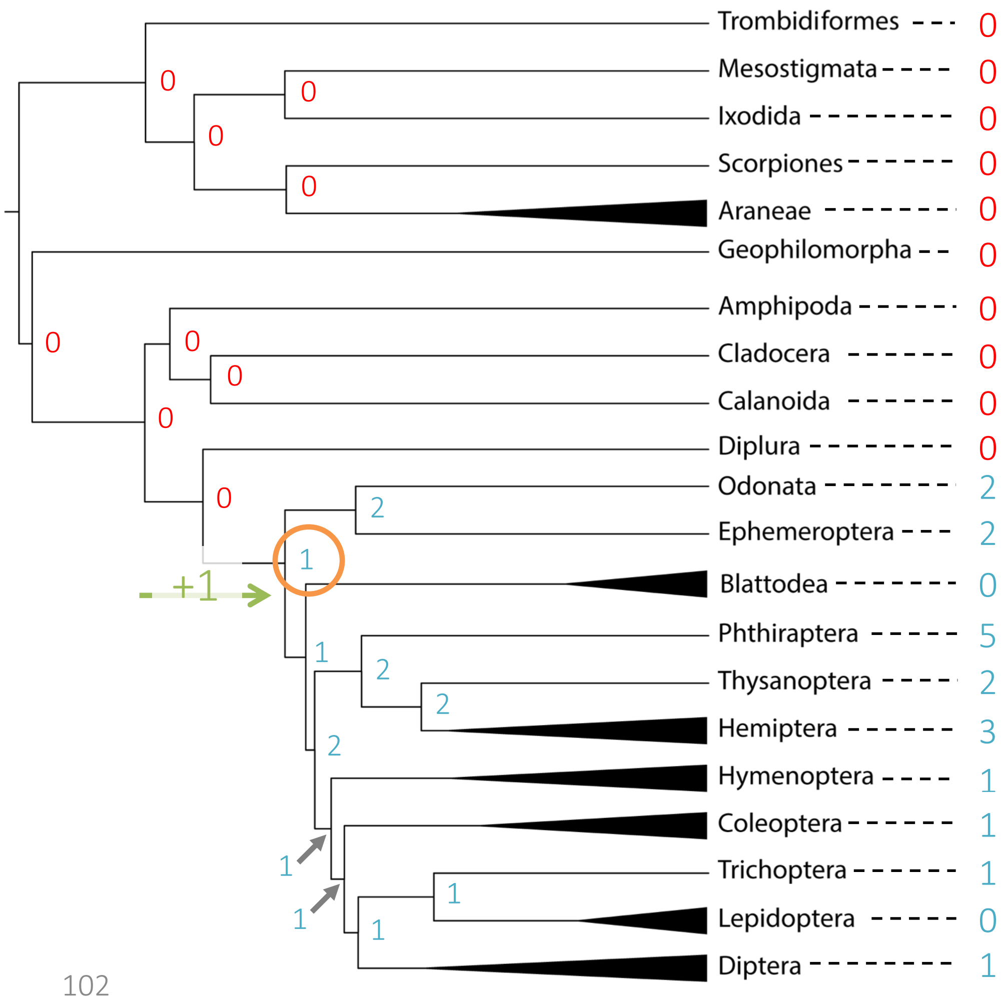
Estimated: 9,601 genes

Corrected: 14,965 genes



How can we infer characteristics of the genome of LICA?

Which families were 'born' during the transition to insects?



How can we infer characteristics of the genome of LICA?

Which families were 'born' during the transition to insects?

147 emergent insect families

Emergent insect families correspond to insect lifestyle adaptations

Changes in exoskeleton development

7 chitin and cuticle production families

Emergent insect families correspond to insect lifestyle adaptations

Changes in exoskeleton development

7 chitin and cuticle production families

Ability to sense in a terrestrial environment

1 visual learning and behavior family
2 odorant binding families
5 families involved in neural activity

Emergent insect families correspond to insect lifestyle adaptations

Changes in exoskeleton development

7 chitin and cuticle production families

Ability to sense in a terrestrial environment

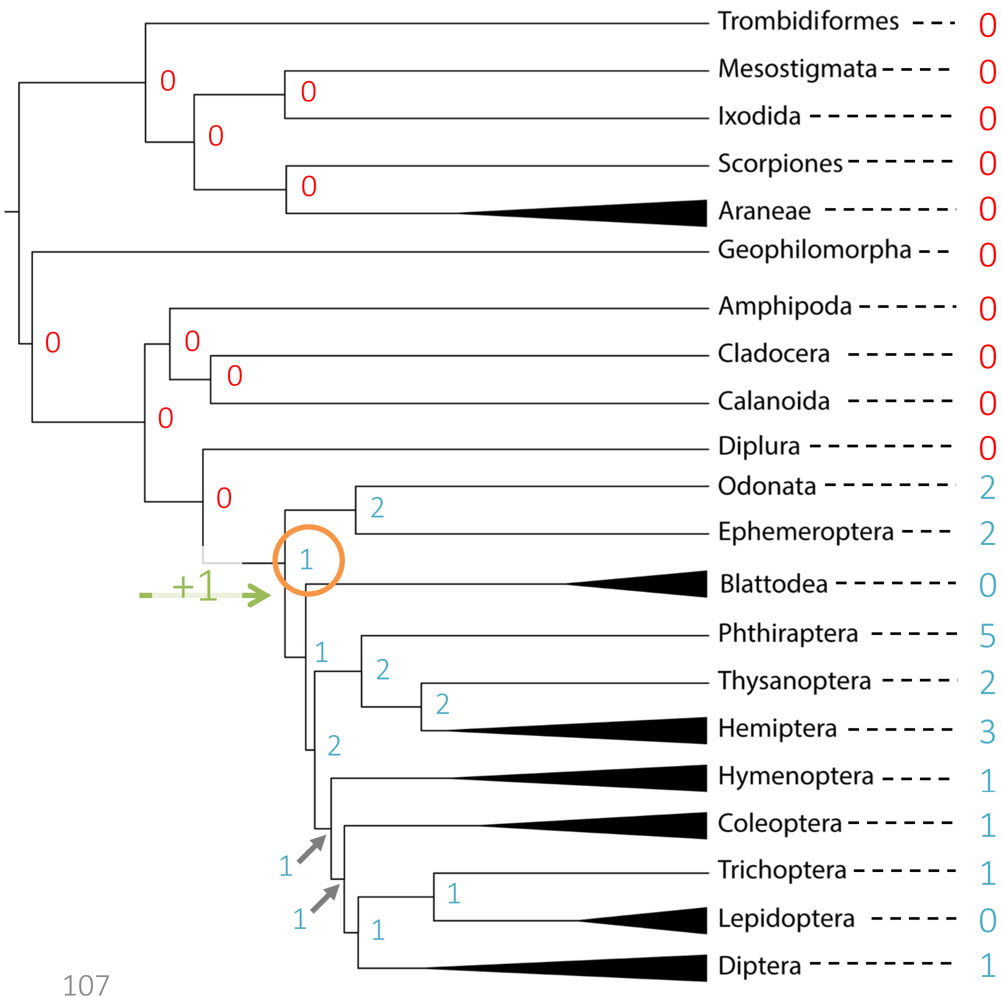
1 visual learning and behavior family
2 odorant binding families
5 families involved in neural activity

Unique development

1 larval behavior family
4 imaginal disk development families

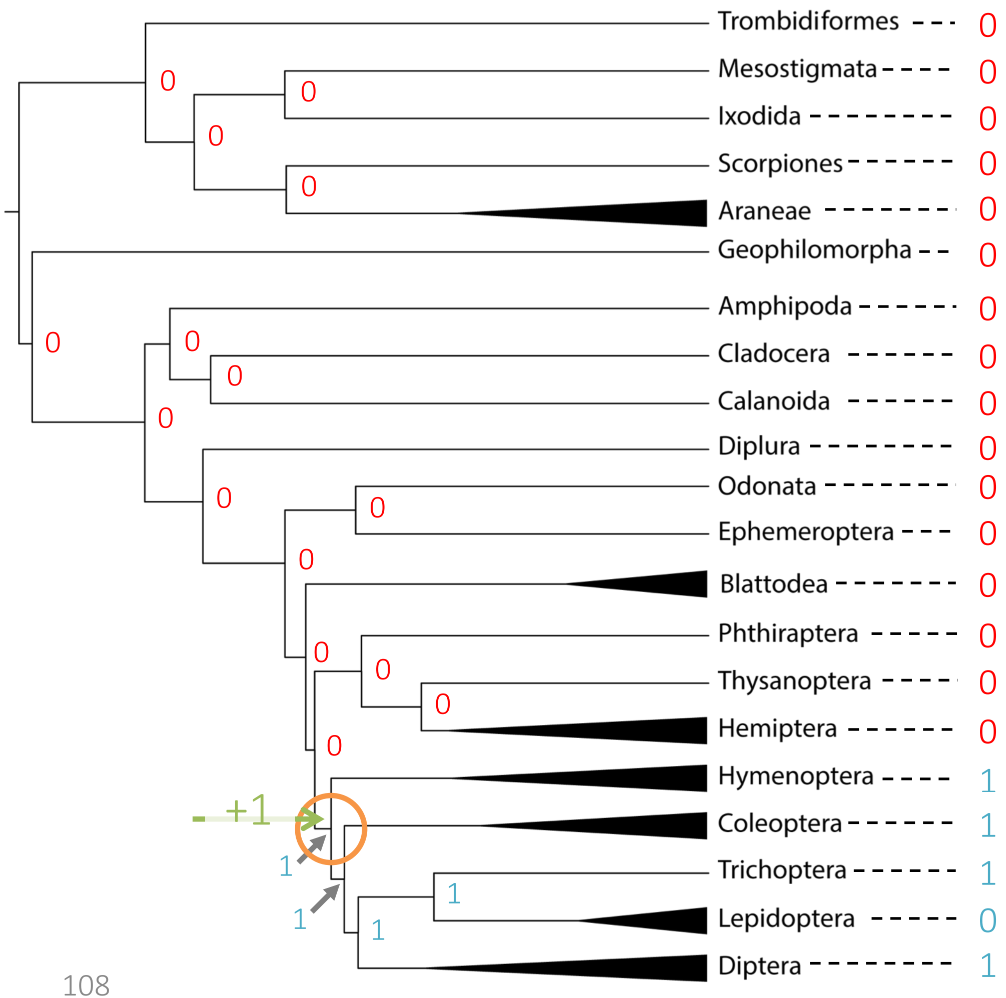
Emergent insect families correspond to insect lifestyle adaptations

Changes in exoskeleton development	7 chitin and cuticle production families
Ability to sense in a terrestrial environment	1 visual learning and behavior family 2 odorant binding families 5 families involved in neural activity
Unique development	1 larval behavior family 4 imaginal disk development families
Flight	3 wing morphogenesis families



How can we infer characteristics of the genome of LICA?

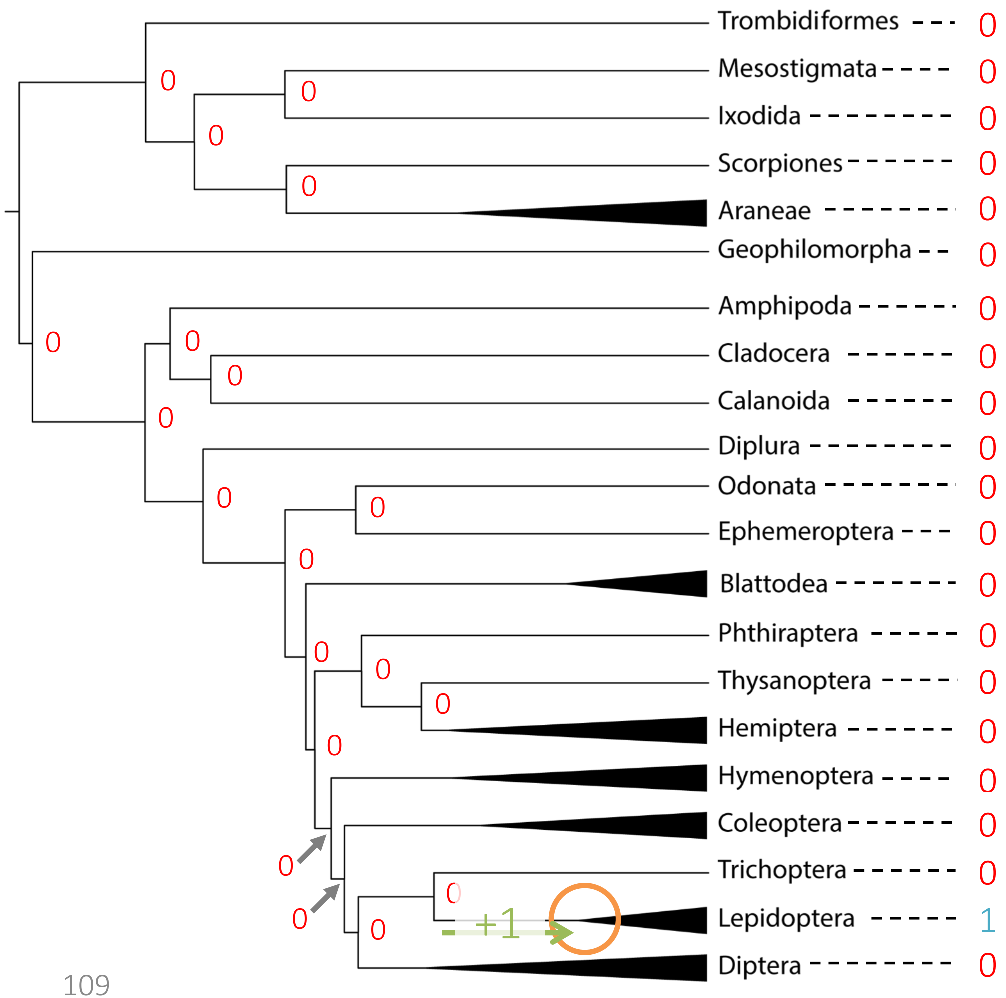
Which families were ‘born’ during the transition to insects?



How can we infer characteristics of the genome of LICA?

Which families were ‘born’ during the ~~transition to insects?~~ transition to Holometabola?

Only 10 emergent Holometabola gene families

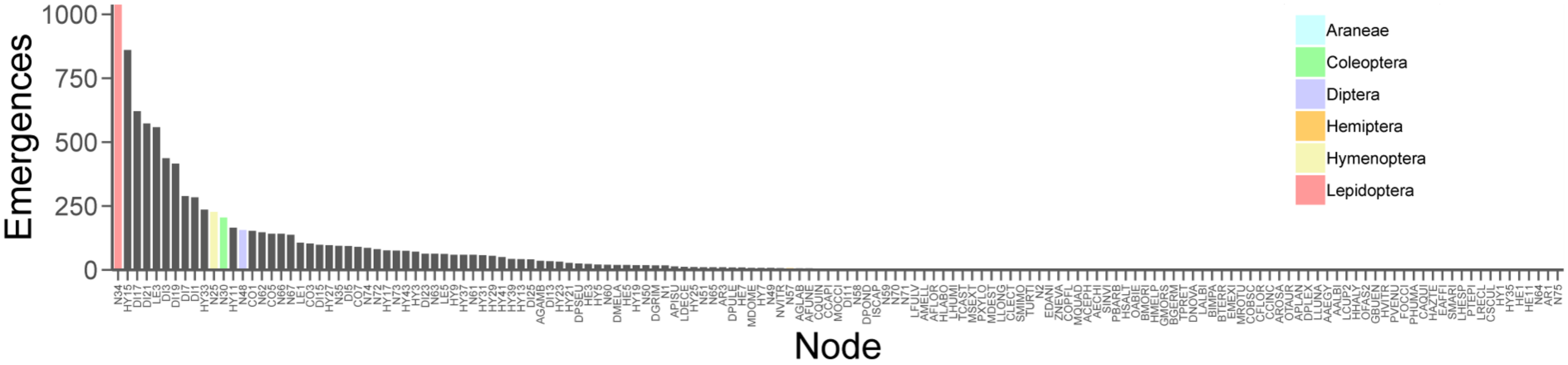


How can we infer characteristics of the genome of LICA?

Which families were 'born' during the
~~transition to insects?~~
~~transition to Holometabola?~~
transition to Lepidoptera?

1,038 emergent Lepidopteran gene families

Lepidoptera has the most emergent gene families

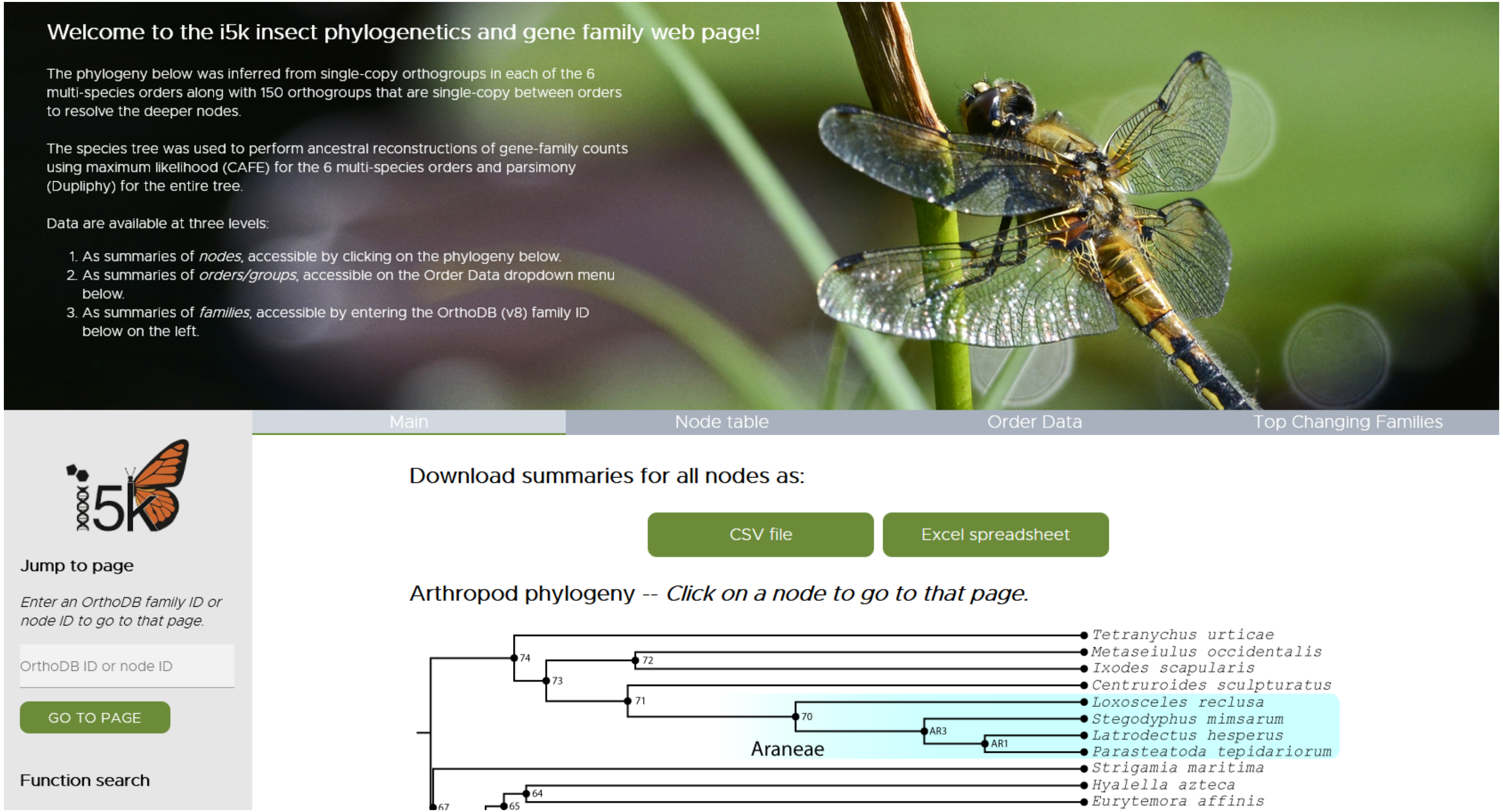


Today's topics

1. Determining the Arthropod phylogeny
2. Reconstructing ancestral gene counts
3. Using the i5k gene family web site

All data has been made available in our online tool

https://i5k.gitlab.io/ArthroFam/



Welcome to the i5k insect phylogenetics and gene family web page!

The phylogeny below was inferred from single-copy orthogroups in each of the 6 multi-species orders along with 150 orthogroups that are single-copy between orders to resolve the deeper nodes.

The species tree was used to perform ancestral reconstructions of gene-family counts using maximum likelihood (CAFE) for the 6 multi-species orders and parsimony (Dupliphy) for the entire tree.

Data are available at three levels:

1. As summaries of *nodes*, accessible by clicking on the phylogeny below.
2. As summaries of *orders/groups*, accessible on the Order Data dropdown menu below.
3. As summaries of *families*, accessible by entering the OrthoDB (v8) family ID below on the left.

Jump to page

Enter an OrthoDB family ID or node ID to go to that page.

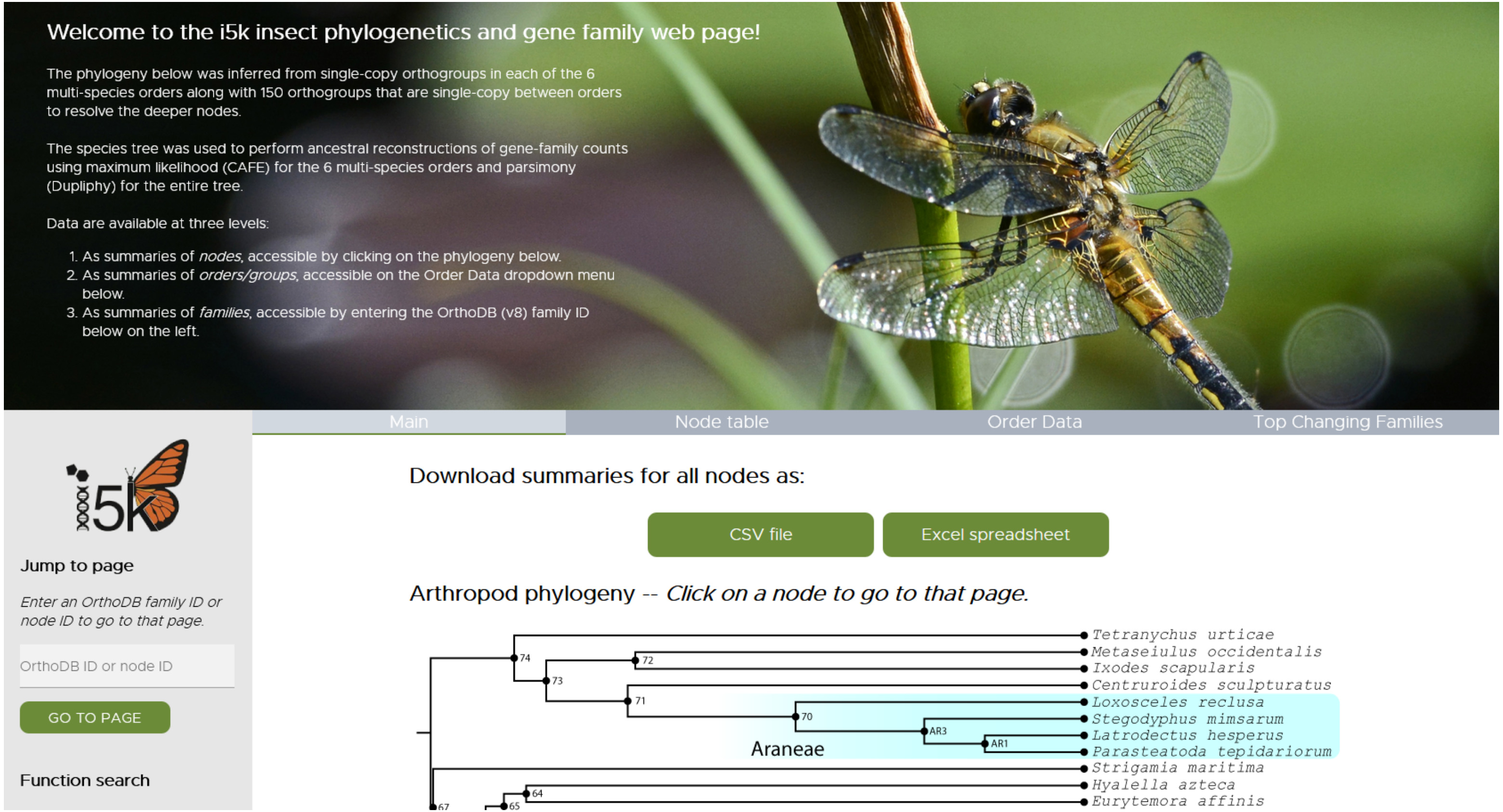
Function search

Download summaries for all nodes as:

Arthropod phylogeny -- *Click on a node to go to that page.*

Working search functions temporarily available at:

https://cgi.soic.indiana.edu/~grthomas/i5k-web/main.html



Welcome to the i5k insect phylogenetics and gene family web page!

The phylogeny below was inferred from single-copy orthogroups in each of the 6 multi-species orders along with 150 orthogroups that are single-copy between orders to resolve the deeper nodes.

The species tree was used to perform ancestral reconstructions of gene-family counts using maximum likelihood (CAFE) for the 6 multi-species orders and parsimony (Dupliphy) for the entire tree.

Data are available at three levels:

1. As summaries of *nodes*, accessible by clicking on the phylogeny below.
2. As summaries of *orders/groups*, accessible on the Order Data dropdown menu below.
3. As summaries of *families*, accessible by entering the OrthoDB (v8) family ID below on the left.

Jump to page

Enter an OrthoDB family ID or node ID to go to that page.

Function search

Download summaries for all nodes as:

Arthropod phylogeny -- *Click on a node to go to that page.*

Demo

Acknowledgements

- Matthew Hahn
- Stephen Richards
- Rob Waterhouse
- Jessica Garb
- Elias Dohmen
- Ariel Chipman

The i5k community

The Hahn lab + Clara Boothby

i5k website:
http://i5k.github.io/

Gene family website:
https://i5k.gitlab.io/ArthroFam/