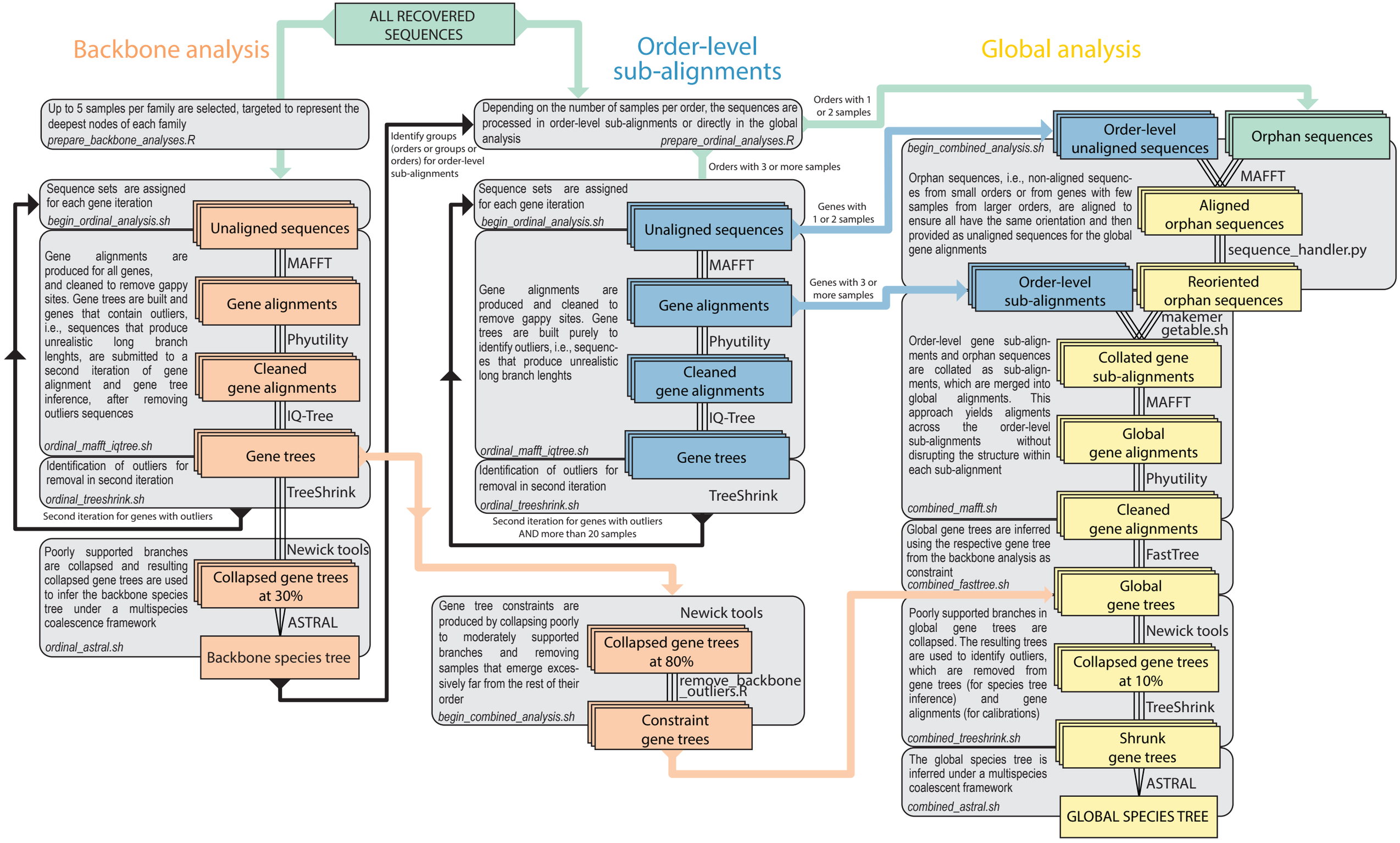
ALL RECOVERED SEQUENCES
Backbone analysis
Order-level sub-alignments
Global analysis
Up to 5 samples per family are selected, targeted to represent the deepest nodes of each family
prepare_backbone_analyses.R
Identify groups (orders or groups or orders) for order-level sub-alignments
Depending on the number of samples per order, the sequences are processed in order-level sub-alignments or directly in the global analysis
prepare_ordinal_analyses.R
Orders with 1 or 2 samples
Orders with 3 or more samples
Sequence sets are assigned for each gene iteration
begin_ordinal_analysis.sh
Unaligned sequences
MAFFT
Gene alignments
Phyutility
Cleaned gene alignments
IQ-Tree
Gene trees
ordinal_mafft_iqtree.sh
Gene alignments are produced for all genes, and cleaned to remove gappy sites. Gene trees are built and genes that contain outliers, i.e., sequences that produce unrealistic long branch lenghts, are submitted to a second iteration of gene alignment and gene tree inference, after removing outliers sequences
Identification of outliers for removal in second iteration
ordinal_treeshrink.sh
TreeShrink
Second iteration for genes with outliers
Poorly supported branches are collapsed and resulting collapsed gene trees are used to infer the backbone species tree under a multispecies coalescence framework
Newick tools
Collapsed gene trees at 30%
ASTRAL
ordinal_astral.sh
Backbone species tree
Sequence sets are assigned for each gene iteration
begin_ordinal_analysis.sh
Unaligned sequences
MAFFT
Gene alignments
Phyutility
Cleaned gene alignments
IQ-Tree
Gene trees
ordinal_mafft_iqtree.sh
Gene alignments are produced and cleaned to remove gappy sites. Gene trees are built purely to identify outliers, i.e., sequences that produce unrealistic long branch lenghts
Identification of outliers for removal in second iteration
ordinal_treeshrink.sh
TreeShrink
Second iteration for genes with outliers AND more than 20 samples
Genes with 1 or 2 samples
Genes with 3 or more samples
Gene tree constraints are produced by collapsing poorly to moderately supported branches and removing samples that emerge excessively far from the rest of their order
begin_combined_analysis.sh
Newick tools
Collapsed gene trees at 80%
remove_backbone_outliers.R
Constraint gene trees
begin_combined_analysis.sh
Order-level unaligned sequences
Orphan sequences
MAFFT
Orphan sequences, i.e., non-aligned sequences from small orders or from genes with few samples from larger orders, are aligned to ensure all have the same orientation and then provided as unaligned sequences for the global gene alignments
Aligned orphan sequences
sequence_handler.py
Order-level sub-alignments
Reoriented orphan sequences
makemer
getable.sh
Order-level gene sub-alignments and orphan sequences are collated as sub-alignments, which are merged into global alignments. This approach yields aligments across the order-level sub-alignments without disrupting the structure within each sub-alignment
Collated gene sub-alignments
MAFFT
Global gene alignments
Phyutility
Cleaned gene alignments
combined_mafft.sh
Global gene trees are inferred using the respective gene tree from the backbone analysis as constraint
FastTree
Global gene trees
combined_fasttree.sh
Poorly supported branches in global gene trees are collapsed. The resulting trees are used to identify outliers, which are removed from gene trees (for species tree inference) and gene alignments (for calibrations)
Newick tools
Collapsed gene trees at 10%
TreeShrink
Shrunk gene trees
combined_treeshrink.sh
The global species tree is inferred under a multispecies coalescent framework
ASTRAL
GLOBAL SPECIES TREE
combined_astral.sh