--wicker_pld (80) --wicker_minLen (80) --wicker_no_internals (no) --wicker_no_ltrs (no) Associated information Input or dependency Output --ltr_divergence PhyLTR function --remove_GC_from_modeltest -reroot <node> I auto --Itrharvest --minlenltr (100) --maxlenltr (1000) --mindistltr (1000) --maxdistltr (15000) --bootstrap_reps --convert_to_ultrametric -ultrametric -orfhits -transcribed line options, set defaults, and load status file (for resuming)

When many processes (PhyLTR functions or --auto_outgroup --rmhomoflank --flank_evalue external programs) complete, (all-by-all, LTRs) (all-by-all, internal regions) --similar (85.0) --vic (60) paths to directories or files are stored in a text file PhyLTR.output/status along with a key.
When PhyLTR starts up the keys and paths --flank_ple
--flank_plencutoff
--maxiterate_small_clusters г-----, --classify_dfam (yes) --min_clust_size (7) --Itrdivergence (turns on modeltest()) --mintsd (4) --maxtsd (20) --xdrop (5) --mat (2) --classify_repbase (yes)
--keep_conflicting_classifications (no)
--keep_no_classification (no) command line flags Assigns classifications based on Dfam and Repbase result: --nosmalls (no) lare read into a dictionary paths. --maxiterate_small_clusters --maxiterate_small_clusters --maxiterate_medium_clusters Before starting most of the main processes, (Dfam trumps Repbase)
Dfam: CT Superfamily --maxiterate_medium_clusters --maxiterate_medium_clusters --mafft_smallAln_maxclustsize Settings the paths dictionary is checked for the --repbase_tblastx_evalue (1e-5) --nhmmer_reporting_evalue (10) --mafft_smallAln_maxclustsize --mafft_smallAln_maxclustsize --mafft_mediumAln_maxclustsize --soloLTRminPid (80.0) --soloLTRminLen (80.0) --mafft_largeAln_maxclustsize --min_clust_size (7) --min clust size (7) --remove_GC_trom_modeltest --ins (-3) --del (-3) (PhyLTR/LTRdigest_HMMs/hmms) --mafft_largeAln_maxclustsize --ltr_divergence '------' --geneconv (g1,g2,g3) --geneconv (g1,g2,g3) --circos --Itrdivergence (no) --soloLTRmaxEvalue (1e-3) ·-----Perform gene conversion , _ - - - - - - · · or LTR divergence analysis? ·----' Itrharvest() Itrdigest() classify_by_homology() - do not remove homoflanks AnnotateORFs() PhyLTR shortClassif() align_ltrs() ltr_divergence() geneconvLTRs() Clustering align whole elements - do not include outgroup V-------Settings Exploratory visualizations --mcl (no) aligner()

checks if
alignment has If --remove_GC_from_modeltest_aln is on LASER

LTR-R diversification

phangorn
ape

changing transposition and deletion rates over time transposition---include_monosites HKY85 if no model test already been done derived elements / AutoAlign() FastTree2 summarizeClusters() , 'p-----analyses in R BIC, -s 11, no I, GTR+G (44 models) N-----false positives TE pHMMs (Pfam) **↓ ← − − − − −** ideogramFromLengths.py Category of (all-by-all, whole elements) Aldous plots process | TreeShape.R <u>geneconvLTRs() output</u> SoloLTRsearch() output capture PAUP* block `L____' <solo>/<clustMethod>_<clustSettings>.SoloLTRsummary.tsv for divergence estimation shortClassif() output <geneconv>/LTRs/ElementsWithEvidenceOfGeneConversion <geneconv>/LTRs/<GCsettings>/LTR_retrotransposon<j>_LTRs.fasta.aln.tab <gffs>/ByClassification/<inFasta>.LTR-RTs.<sf>.gff PhyLTR Python 3 Initialization Output <clusts>/Wicker_<custSettings>.summary.tab (cluster sizes by sf) (not overwritten if resuming) <u>geneconvClusters() output</u> align_ltrs() output AnnotateOrfs() output <clusts>/Wicker_<custSettings>.membership.tab <out>/status <out>/log.txt <clusts>/<sf>/wicker_groups_<sf> (wicker clusters by element name) <geneconv>/ElementsWithEvidenceOfGeneConversion <orfs>/internals.gff (MISNAMED-A FASTA file) <align>/LTRs/LTR_retrotransposon<j>_LTRs.fasta.aln <clusts>/<sf>/internals.gff <geneconv>/<sf>_summary (location of summary files) <align>/LTRs/LTR_retrotransposon<j>_LTRs.fasta.aln.trimmed Itr_divergence() output <orfs>/internals.gff.orfs <clusts>/<sf>/internals.fasta <geneconv>/<sf>/<sf>_<GCsettings>.summary <orfs>/internals.gff.orfs.gff <clusts>/<sf>/internals.fasta.< blastdbFiles > <geneconv>/<sf>/<GCsettings>/clust<i>_elements.fasta.aln.tab <orfs>/FullWithORFs_gt_<orfLen>.gff <|trdivergence>/LTR_divergences.tab
<|trdivergence>/PAUP/nexus/<sf>_<i>_LTR_retrotransposon<j>_divergence_<model>.nex <clusts>/<sf>/Internals_selfBlastn.tab <orfs>/FullWithORFs_gt_<orfLen>.ORFs.fasta <u>ltrharvest() output</u> <clusts>/<sf>/LTRs_<sf>.gff <gffs>/<inFasta>.LTRdigest.withORFs_gt_<orfLen>.gff <clusts>/<sf>/LTRs_<sf>.fasta <ltrdivergence>/PAUP/divergences/divergence.<model>.LTR_retrotransposon<j> <suffix>/<inFasta> (a copy) <clusts>/<sf>/LTRs_<sf>.fasta.< blastdbFiles > modeltest() output <suffix>/<inFasta>.<suffixArrayFiles> <clusts>/<sf>/LTRs_selfBlastn.tab <harvest>/<inFasta>.ltrharvest.out.gff (sorted) <modeltest>/<GCremoved>/<sf>/1_iters/Summary.txt (PAUP blocks for divergence estimation) <digest>/<inFasta>.LTRdigest_tabout.csv <modeltest>/<GCremoved>/<sf>/1_iters/<sf>/cluster_<i>/1_iters/<sf>_I<custSettings>_<i>.tree <modeltest>/<GCremoved>/<sf>/1_iters/<sf>/cluster_<i>/1_iters/<sf>_I<custSettings>_<i>.tree.jModelTest2.out <clusts>/MCL_<custSettings>.summary.tab (cluster sizes by sf) classify by homology() output <clusts>/MCL_<custSettings>.membership.tab AutoAlign() output <clusts>/<sf>/<sf>.fasta <dfam>/<inFasta>.nhmmer_DfamHits.table <clusts>/<sf>/i<sf>.fasta.<blastdbFiles> <align>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/elements.gff <dfam>/<inFasta>.LTR_retrotransposon_DfamBestHits.tab <u>AutoAlign() output</u> <u>ltrdigest() output</u> <clusts>/<sf>/out.LTR_retrotransposon_<sf>.mcl.<settings> (cluster assignments) <repbase>/<inFasta>.tblastx Repbase.tab <align>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/elements.fasta <clusts>/<sf>/LTR_retrotransposon.AllByAllblastn.abc <align>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/elements.fasta.aln <repbase>/<inFasta>.LTR_retrotransposon_RepbaseBestHits.tab <align>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/elements.gff <clusts>/<sf>/LTR_retrotransposon_<sf>.mci <digest>/<inFasta>.LTRdigest.gff <gffs>/LTRdigest_LTR_retrotranspsons.gff <align>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/elements.fasta.aln.trimal <clusts>/<sf>/LTR_retrotransposon_<sf>.tab <align>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/elements.fasta <digest>/<inFasta>.LTRdigest_3ltr.fas <gffs>/LTRharvest_LTR_retrotranspsons.gff <align>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/elements.fasta.aln <clusts>/<sf>/<inFasta>.LTRdigest_LTR_retrotransposon_<sf>_AllByAll.blastn.tab <digest>/<inFasta>.LTRdigest_5ltr.fas <gffs>/RemoveNonLTRretrotransposons.log <align>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/elements.fasta.aln.trimal <digest>/<inFasta>.LTRdigest_complete.fas summarizeClusters() output <digest>/<inFasta>.LTRdigest_ppt.fas <gffs>/<sf>.gff (ERRORS IN THESÉ, NEED TO DELETE) <digest>/<inFasta>.LTRdigest_pdom_<FastaFiles> DEF: <sf>.<clustMethod>.<clustSettings> = <clustering> <gffs>/<inFasta>.LTRdigest.withDfam.gff <digest>/<inFasta>.LTRdigest_conditions.csv <gffs>/<inFasta>.LTRdigest.withDfam.withRepbase.gff <digest>/<inFasta>.LTRdigest_tabout.csv <gffs>/ByClassification/<clustering>/<clustering>.cluster_<i>.gff <gffs>/<inFasta>.LTRdigestClassifiedNoFP.gff (final) Text file(s) generated os>/<clustMethod>/circos.<sf>.cluster_<i>.geneconv_<GCsettings>/circos>/<clustMethod>/circos.<sf>.cluster_<i>.geneconv_<GCsettings>/circos>/<clustMethod>/plots.elements/<sf>.cluster_<i>.geneconv_<GCsettings>/<clustMethod>/plots.elements/<sf>.cluster_<i>.geneconv_<GCsettings>/<clustMethod>/plots.elements/<sf>.cluster_<i>.geneconv_<GCsettings>/<clustMethod>/plots.elements/<sf>.cluster_<i>.geneconv_<GCsettings>/<clustMethod>/plots.elements/<sf>.cluster_<i>.geneconv_<GCsettings>/<clustMethod>/plots.elements/<sf>.cluster_<i>.geneconv_<GCsettings>/<clustMethod>/plots.elements/<sf>.cluster_<i>.geneconv_<GCsettings>/<clustMethod>/plots.elements/<sf>.cluster_<i>.geneconv_<GCsettings>/<clustMethod>/plots.elements/<sf>.cluster_<i>.geneconv_<GCsettings>/<clustMethod>/plots.elements/<sf>.cluster_<i>.geneconv_<GCsettings>/<clustMethod>/plots.elements/<sf>.cluster_<i>.geneconv_<GCsettings>/<clustMethod>/plots.elements/<sf>.cluster_<i>.geneconv_<GCsettings>/<clustMethod>/plots.elements/<sf>.cluster_<i>.geneconv_<GCsettings>/<clustMethod>/plots.elements/<sf>.cluster_<i>.geneconv_<GCsettings>/<clustMethod>/plots.elements/<sf>.cluster_<i>.geneconv_<GCsettings</td>.geneconv_</t PhyLTR install dir = PhyLTR PhyLTR output dir = out <circos>/<clustMethod>/<sf>.cluster_<i>.geneconv_<GCsettings>.links.track <circos>/<clustMethod>/<sf>.cluster_<i>>.geneconv_<GCsettings>.links_untransposed.track <circos>/<clustMethod>/<sf>.cluster_<i>.gff phylo() + bootstrap() output <circos>/<clustMethod>/<sf>.cluster_<i>.gff.heatmap.track <circos>/<clustMethod>/<sf>.cluster_<i>.gff.tile.track <align>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/seqboot.conf <circos>/<clustMethod>/<sf>.cluster_<i>.LTR_highlights.track <align>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/outfile (SEQBOOT multi-phylip) <circos>/<clustMethod>/<sf>.cluster_<i>.seg.track <phylo>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/_cluster_<i>.bootstrapped.newick <circos>/<clustMethod>/circos.<sf>.cluster_<i>.geneconv_<GCsettings(s)>/data/<trackFiles> <phylo>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/<sf>_cluster_<i>>.bootstrapped.pathd8_ultrametric_outgroup_<k>.newick <circos>/<clustMethod>/circos.<sf>.cluster_<i>.geneconv_<GCsettings>/etc/bands.conf <phylo>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/<m>/alnReplicate_<m>.fasta <circos>/<clustMethod>/circos.<sf>.cluster_<i>.geneconv_<GCsettings>/etc/break.conf <phylo>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/<m>/tree.newick **Abbreviations** <circos>/<clustMethod>/circos.<sf>.cluster_<i>.geneconv_<GCsettings>/etc/circos.conf <phylo>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/allReplicates.newick <circos>/<clustMethod>/circos.<sf>.cluster_<i>.geneconv_<GCsettings>/etc/circos.conf.heatmaps <phylo>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/elements.fasta.aln.trimal.main_tree <circos>/<clustMethod>/circos.<sf>.cluster_<i>.geneconv_<GCsettings>/etc/conforig <phylo>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/elements.fasta.aln.trimal.main_tree.bootstrapped <circos>/<clustMethod>/circos.<sf>.cluster_<i>.geneconv_<GCsettings>/etc/ideogram.conf <phylo>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/elements.fasta.aln.trimal.main_tree.bootstrapped.pathd8_ultrametric <circos>/<clustMethod>/circos.<sf>.cluster_<i>.geneconv_<GCsettings>/etc/ideogram.label.conf <circos>/<clustMethod>/circos.<sf>.cluster_<i>.geneconv_<GCsettings>/etc/ideogram.position.conf <phylo>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/pathd8.in (PATHd8 configuration file) <out> = main PhyLTR output directory (default is PhyLTR.output/) <circos>/<clustMethod>/circos.<sf>.cluster_<i>.geneconv_<GCsettings>/etc/image.conf <circos>/<clustMethod>/circos.<sf>.cluster_<i>.geneconv_<GCsettings>/etc/image.generic.conf <suffix> = <out>/suffixerator/ AutoAlign() output <circos>/<clustMethod>/circos.<sf>.cluster_<i>.geneconv_<GCsettings>/etc/ticks.conf <ltrharvest> = <out>/LTRharvest/ <circos>/<clustMethod>/circos.<sf>.cluster_<i>.geneconv_<GCsettings>.justelements/etc/bands.conf <Itrdigest> = <out>/LTRdigest/ <circos>/<clustMethod>/circos.<sf>.cluster <i>.geneconv <GCsettings>.justelements/etc/break.conf <align>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/Outgroups <orfs> = <out>/AnnotateORFs/ <align>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/elements.gff <circos>/<clustMethod>/circos.<sf>.cluster_<i>.geneconv_<GCsettings>.justelements/etc/circos.conf <gffs> = <out>/GFF_output/ <circos>/<clustMethod>/circos.<sf>.cluster_<i>.geneconv_<GCsettings>.justelements/etc/circos.conf.heatmaps <align>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/elements.fasta <dfam> = <out>/DfamClassification/ <circos>/<clustMethod>/circos.<sf>.cluster <i>.geneconv <GCsettings>.justelements/etc/conforig <align>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/blast_all_by_all_flanks <repbase> = <out>/RepbaseClassification/ <align>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/repeat_regions.gff <circos>/<clustMethod>/circos.<sf>.cluster_<i>.geneconv_<GCsettings>.justelements/etc/ideogram.conf <clusts> = <out>/<clustMethod>/<clustSettings>/Clusters/ <align>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/repeat_regions.flanks.gff
<align>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/repeat_regions.flanks.fasta
<align>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/repeat_regions.flanks.fasta.trimal <circos>/<clustMethod>/circos.<sf>.cluster <i>.geneconv <GCsettings>.justelements/etc/ideogram.label.conf <align> = <out>/<clustMethod>/<clustSettings>/Alignments/ <circos>/<clustMethod>/circos.<sf>.cluster_<i>.geneconv_<GCsettings>.justelements/etc/ideogram.position.conf <geneconv> = <out>/<clustMethod>/<clustSettings>/GENECONV/
<sololtr> = <out>/<clustMethod>/<clustSettings>/SoloLTRsearch/ <circos>/<clustMethod>/circos.<sf>.cluster_<i>.geneconv_<GCsettings>.justelements/etc/image.conf <circos>/<clustMethod>/circos.<sf>.cluster_<i>.geneconv_<GCsettings>.justelements/etc/image.generic.conf <align>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/repeat_regions.flanks.fasta.trimal.phylip <align>/<part>/<GCremoved>/<flanks>/<outgroup>/<sf>/cluster_<i>/repeat_regions.flanks.fasta.<blastdbFiles> <circos> = <out>/Circos/ <circos>/<clustMethod>/circos.<sf>.cluster_<i>.geneconv_<GCsettings>.justelements/etc/ticks.conf <modeltest> = <out>/<clustMethod>/<clustSettings>/Modeltest/ <LTRdivergence> = <out>/<clustMethod>/<clustSettings>/LTR_divergence/ <phylo> = <out>/<clustMethod>/<clustSettings>/Trees/

<sf> = each superfamily (e.g. Gypsy or Pao)

<k> = LTR RT ID when used as an outgroup

<clustMethod> = WickerFam or MCL
<part> = WholeElements or InternalRegions

<inFasta> = input FASTA filename <custSettings> = a string specifying cluster settings (e.g. l6 for MCL)

<orfLen> = minimum ORF length given by --min_orf_len (default 300 bp)

flanks> = NoPairsWithHomlogousFlanks or AllElements

<GCremoved> = NoGCFiltering or GeneconversionDisallowed

<outgroup> = WithOutgroup or NoOutgroup

<GCsettings> = any combination of g0, g1, g2
<*Files> = multiple files, where * is an explanation

<model> = nucleotide substitution model (e.g. HKY85)

<i> = cluster number <j> = LTR RT ID

<m> = bootstrap replicate #