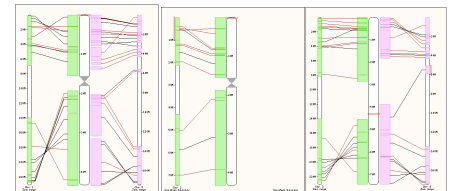


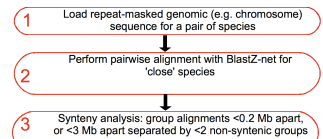


GRAMENE

Ensembl Compara Gene Tree Pipeline¹

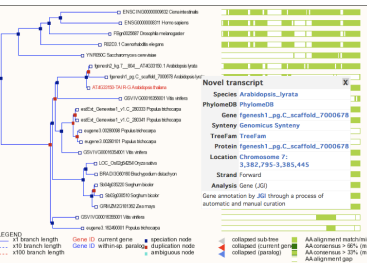
Synthetic genes and regions were calculated using gene orthologues predicted from Gramene Protein Complex analysis. JACOBIanet was used to identify colinear genes. To increase sensitivity, chaining was created on the basis of gene order, excluding positions of non-orthologous genes, rather than using gene coordinates. Chains were required to have at least five colinear genes with no more than ten intervening genes between neighbors. Resulting gene pairs were classified as 'synthetic/colinear'. Next, we searched for additional synteries among non-colinear genes to account for small-scale rearrangements and assembly errors. This also allowed members of tandemly duplicated clusters to be identified. Because of the small size of the genome, we were able to identify all synteries. The remaining 1000 genes were classified as non-colinear anchors were identified and mapped to the insensitive positions in species2. If the non-colinear ortholog in species2 was located within five genes of either anchor, then the non-colinear pair of genes was classified as 'synthetic/anchor'. Most of these occurred within already defined colinear chains. The coordinates of colinear chains were used to define syntetic regions.

Whole Genome Alignment
(BlastZ-net^{7,8} Pairwise Alignment)

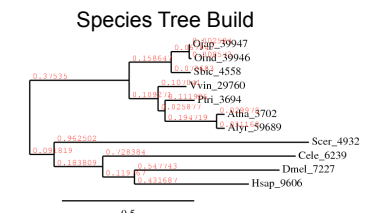
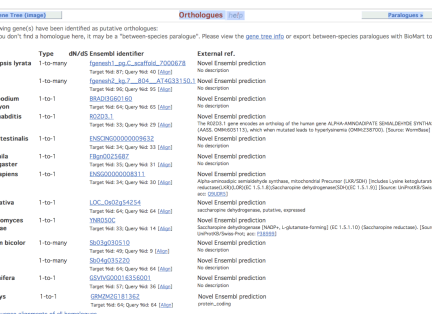


BlastZ-net alignments are provided for closely related pairs of species. The alignments are the results of post-processing the raw BlastZ results. In the first step, original blocks are chained according to their location in both genomes. The netting process chooses for the reference species

Gene Tree & Schematic Alignment



Species Tree Build



1532 trees with members from 11 species were identified, 419 were excluded from the subsequent tree building process due to the lack of orthologous representative from at least one species (i.e. all the members from that species originated from duplication events instead of speciation events). For each of the 1125 qualified trees, the multiple cDNA alignment was reconstructed from multiple protein alignment by using a comparison database with the aligned genes removed. The 1125 unrooted multiple cDNA alignments were then concatenated into one large multiple cDNA alignment and stored in phylip format for PHML-Basemr² analysis. The basemr was run with this concatenated multiple alignment with a guisting topology tree. REV model. The resulting species tree was calculated from 2579787 nucleotide sites, and has branch lengths representing the neutral nucleotide substitution rates for each species.

Gramene Compara Analysis Summary

five model metazoan genomes
Drosophila melanogaster, *Anopheles gambiae*, *Drosophila obscura*, *Drosophila dentissima*, *Drosophila melanogaster*

six monacot genomes

four dicot genomes

In graminere compans database with the shared gaps removed. The 1113 ungrouped multiple cDNA alignments were then concatenated into one large size multiple cDNA alignment and stored in phylip format for PAML-baseml² analysis. The baseml was run with this concatenated multiple alignment with a guiding topology tree, REV model. The resulting species tree was calculated from 2579787 nucleotide sites, and has branch-lengths representing the neutral nucleotide substitution rates for each species.

1. Vellie AL, Sweatt J, Ureta-Vidal A, Durbin R, Heng L, Birney E. EnsemblCompara GeneTrees: Analysis of complete, duplication aware phylogenetic trees in vertebrates. *Genome Research* 2004; 19: 227-33.
2. Husek J. *seq_hsearch*: clustering sequences for sparse graphs. http://bioinformatics.pict.dtc.utoronto.ca/~husek/seq_hsearch/
3. Wallace IM, O'Sullivan C, Higgins DG, Nottingham C. M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acid Research* 2009; Mar 23;34(5):1650-1652.
4. Li H et al., TreeDist2 <http://bioinformatics.pict.dtc.utoronto.ca/~husek/treeDist2/>
5. Yang Z. *PhyML: a program package for phylogenetic analysis by maximum likelihood* *Comput Appl Biosci* 1997; Oct; 13(5):555-556.
6. Yang Z et al 2004. "Duplications: a tool for mining sequence duplications and gene duplications" *Bioinformatics* 20(18):3943-6
7. Schwartz S et al. "Human-Mouse Alignments with BLASTZ" *Genome Res* 13(1):107-125
8. Kent WJ et al. "Evolution's caudron: Duplication, deletion, and rearrangement in the mouse and human genomes" *Proc Natl Acad Sci U S A* 2003; 100(20):11484-9