

Comparative genomics: Insights into conifer evolution

Amanda De La Torre

Pär K. Ingvarsson

Outline

- Introduction
 - Definitions of common terms
 - Methods: programs and resources
- Comparative genomics in Conifers
- Current work under the ProCoGen project

INTRODUCTION

What is comparative genomics?

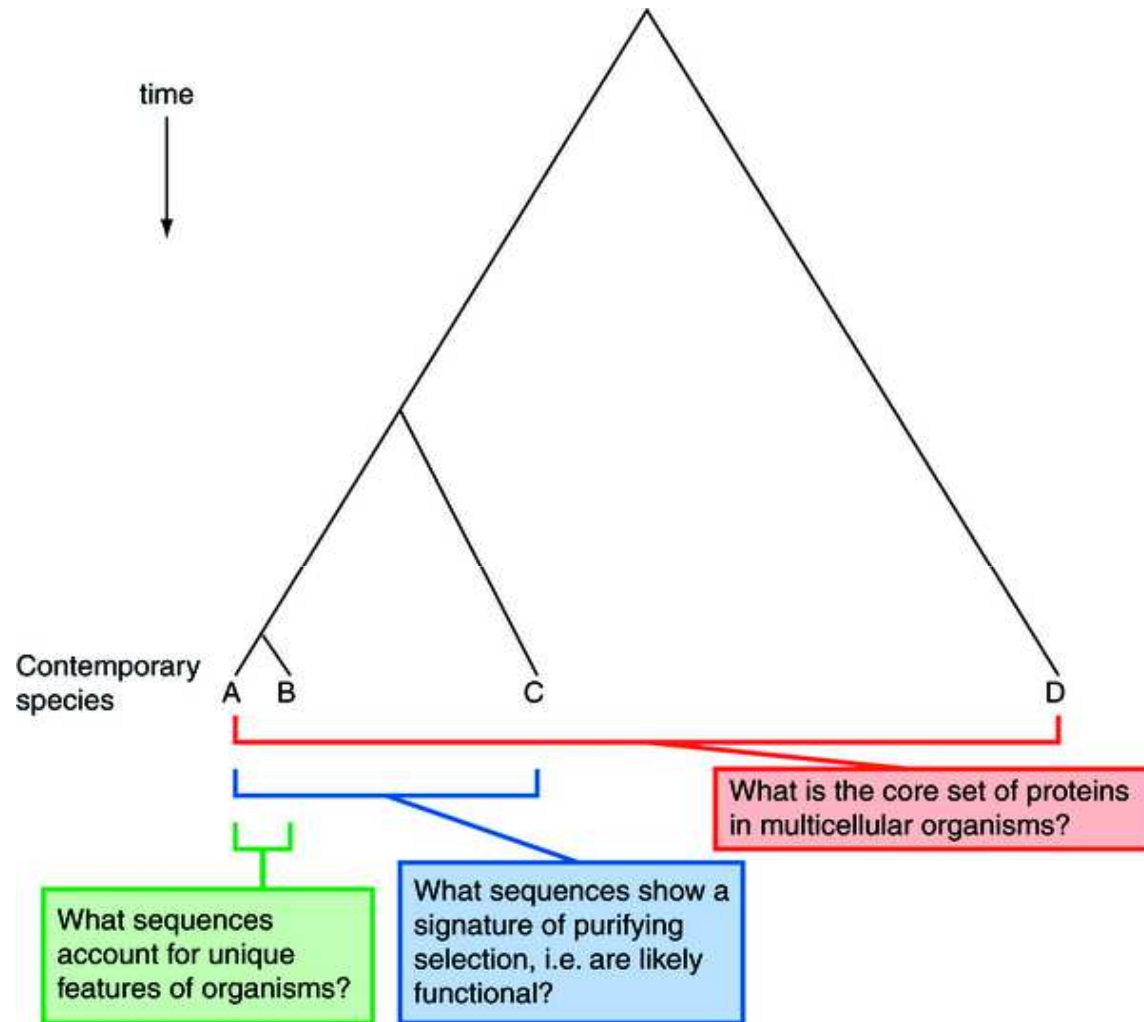
All modern genomes have arisen from common ancestral genomes, therefore the information gained in one organism can be important in other even distantly related organisms. Comparative genomics uses sequence comparisons within and between species to deduce the evolutionary history and the function (if any) of particular DNA sequences.

Principles

- Changes in DNA and protein sequences are subject to drift, selection and mutation.
- Rates of evolutionary change vary both between functional categories of sequences and regionally within genomes.
- DNA sequences encoding and regulating the expression of essential proteins and RNAs will be conserved; whereas those that encode or control the expression of proteins responsible for differences between species will be divergent.

Comparative genomics definition

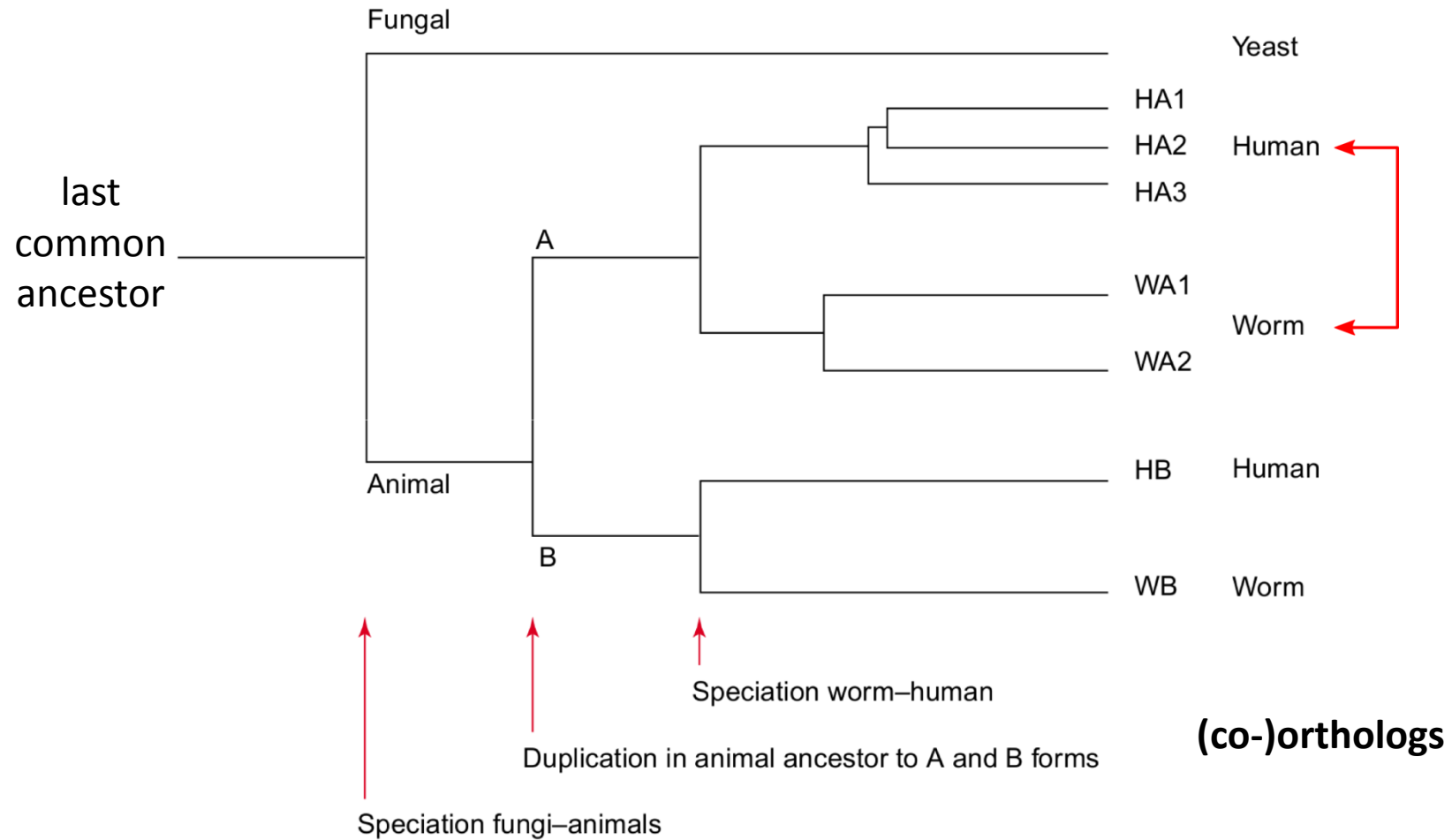
“The combination of genomic data and comparative/evolutionary biology to address questions of genome structure, evolution and function”



Definition of common terms

- Homology- Shared ancestral common origin
- Orthologs- Homologous genes evolved through speciation (common ancestor)
- Paralogs- Homologous genes evolved through duplication within the same genome.

Homology



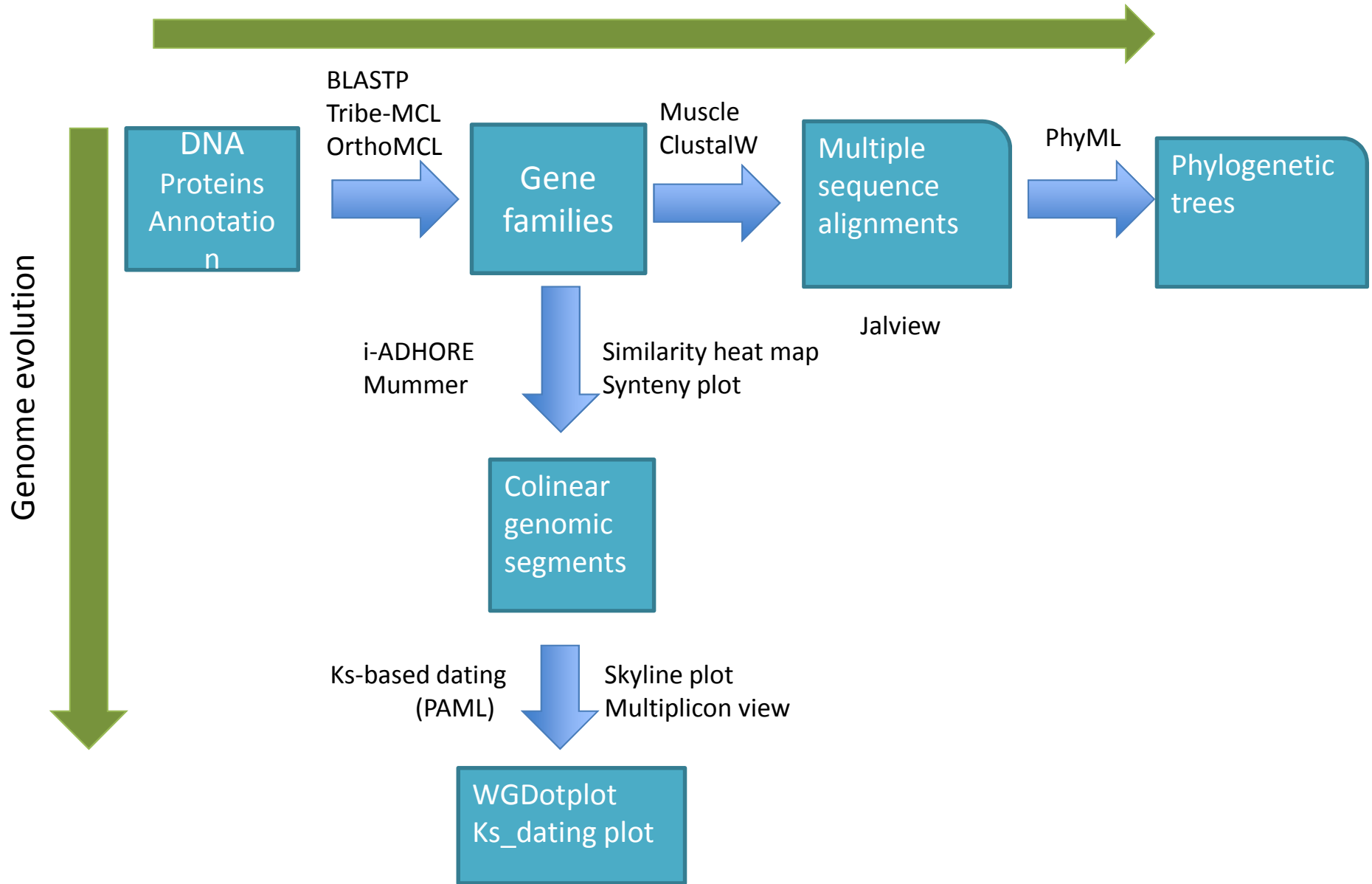
Definition of common terms

- Synteny- Conservation of gene content
- Colinearity- Conservation of gene content and order
- Macro-colinearity: Marker-based (Maps)
- Micro-colinearity: DNA based or gene based

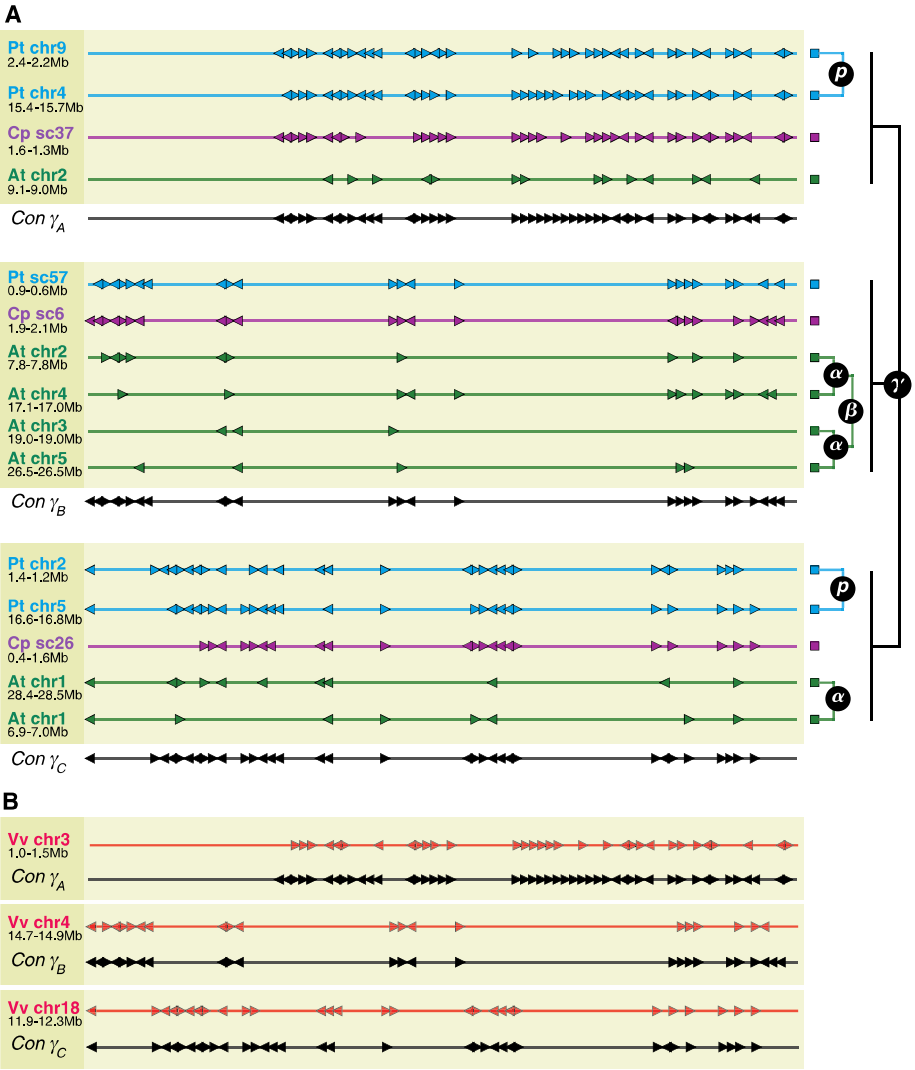
What can we learn from comparisons between species?

- Genome conservation: transfer knowledge from model organisms to non-model organisms
- Genome variation: understand how genomes change over time in order to identify evolutionary processes and constraints
- Detection of functional elements: Coding elements (exons) and non-coding sequences

Gene family evolution



Example of colinearity in plant genomes



Programs

Detection of orthologous sequences

- BLAST <http://blast.ncbi.nlm.nih.gov/Blast.cgi>
- Tribe-MCL <http://www.bioperl.org/wiki/TribeMCL>
- OrthoMCL <http://orthomcl.org/orthomcl/>

Alignment, synteny and colinearity

- ClustalW <http://www.clustal.org>
- Muscle <http://www.drive5.com/muscle/>
- Mummer <http://mummer.sourceforge.net>
- i-Adhore <http://bioinformatics.psb.ugent.be/software/details/i--ADHoRe>

Phylogenetic-Genome evolution

- PAML <http://abacus.gene.ucl.ac.uk/software/paml.html>
- Mega <http://www.megasoftware.net>
- PhyML <http://www.atgc-montpellier.fr/phyml/>

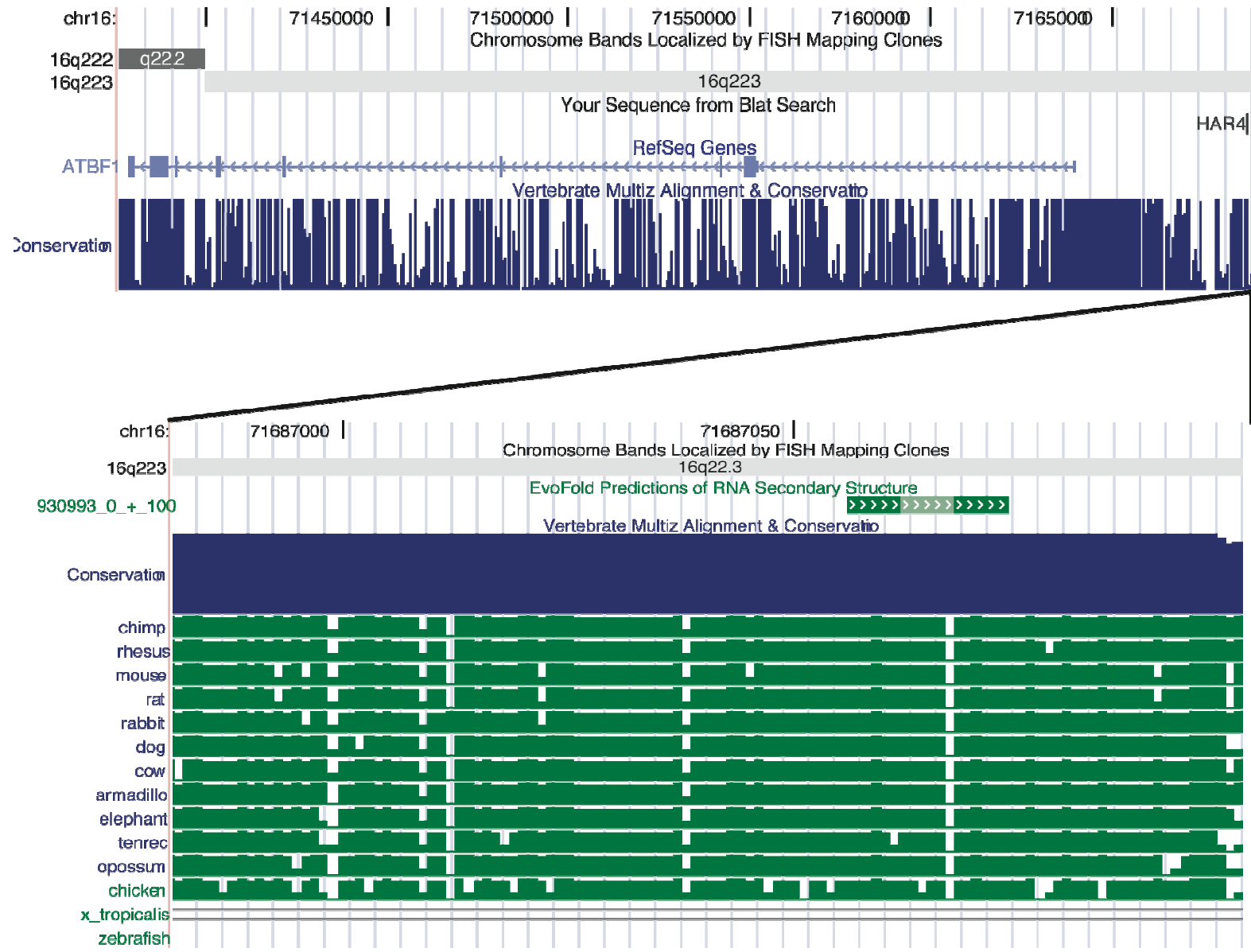
Resources

Comparative genomics

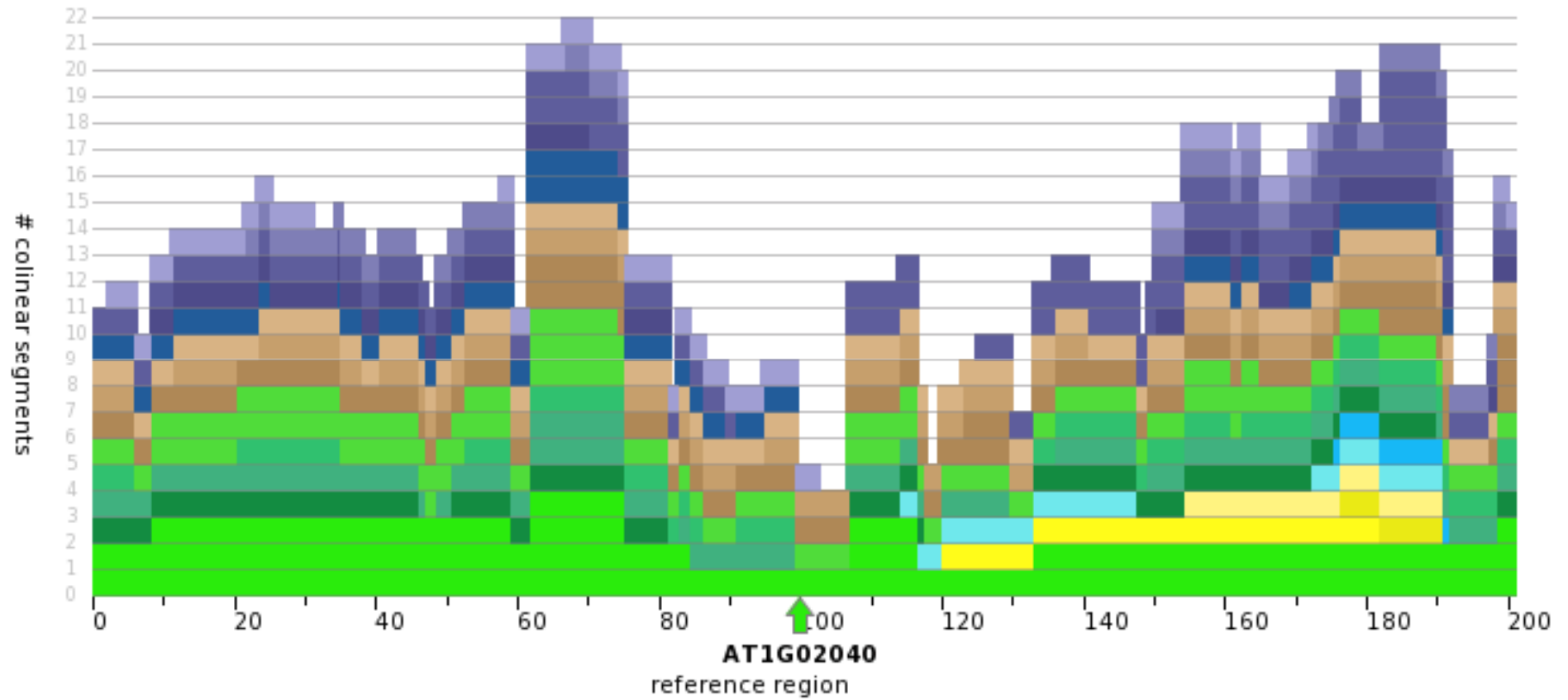
- PLAZA <http://www.bioinformatics.psb.ugent.be/plaza>

Annotation

- ORCAE <http://www.bioinformatics.psb.ugent.be/orcae>
- Ensembl <http://www.ensembl.org>
- UCSC <http://www.genome.ucsc.edu>
- NCBI <http://www.ncbi.nlm.nih.gov>
- Phytozome <http://www.phytozome.net>



Example of colinearity in plant genomes



COMPARATIVE GENOMICS IN CONIFERS

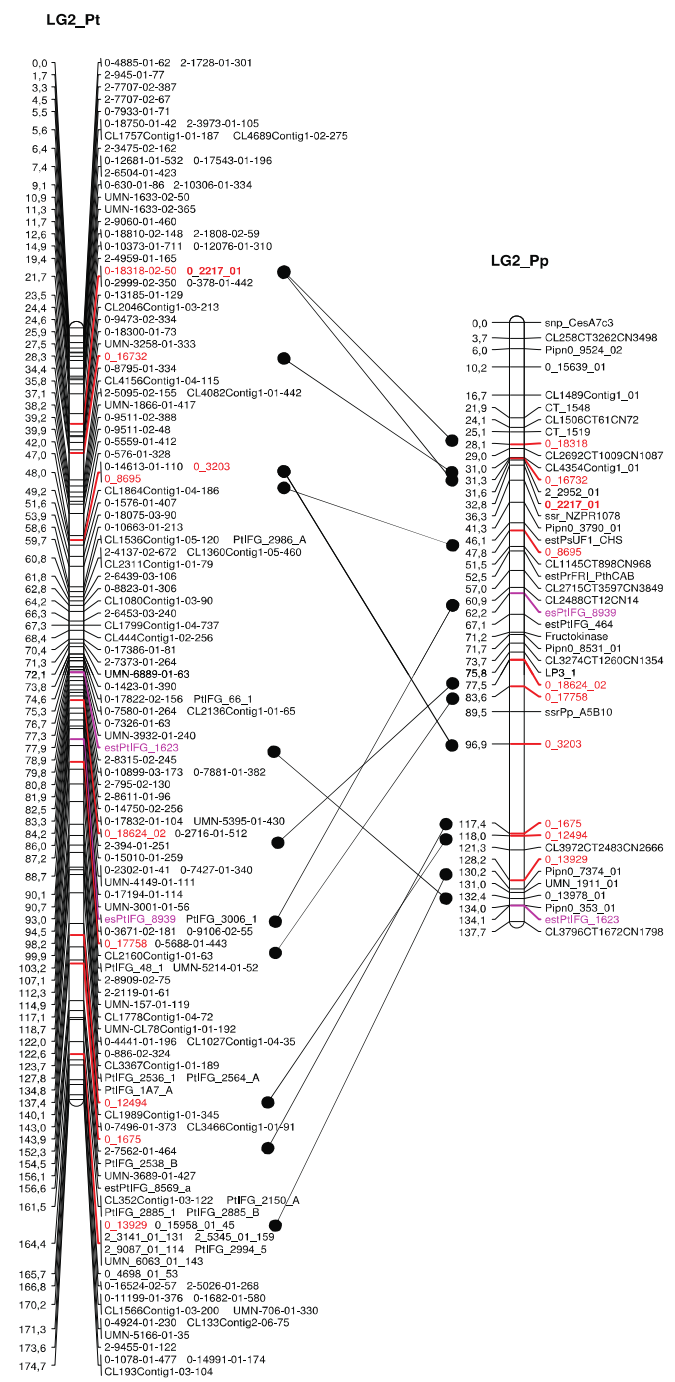
Comparative genomics in conifer species (Comparative mapping)

Species	Marker	Reference
<i>Pinus taeda</i> , <i>P. radiata</i>	RFLP	Davey et al. 1999
<i>Pinus sylvestris</i> , <i>P.taeda</i>	EST	Komulainen et al. 2003
<i>Pinus taeda</i> , <i>P.pinaster</i>	QTL	Chagne et al. 2003
<i>Pinus taeda</i> , <i>Pseudotsuga menziesii</i>	ESTP, RFLP	Krutovsky et al. 2004
<i>Picea glauca</i> , <i>P.mariana x rubens</i> , <i>P. abies</i>	AFLP, SSR, ESTPs	Pelgas et al. 2006
<i>Pinus taeda</i> , <i>Pinus elliotti</i> , <i>P. caribea</i>	SSR	Shepherd et al. 2008
<i>Pseudotsuga menziesii</i>	AFLP	Mansfield et al. 2008
<i>Pseudotsuga menziesii</i>	RFLP, SNP	Eckert et al. 2009
<i>Picea</i> , <i>Pinus</i> , <i>Pseudotsuga</i>	239 COS markers	Liewlaksaneeyahawin et al. 2009

Comparative genomics in conifer species (Comparative mapping)

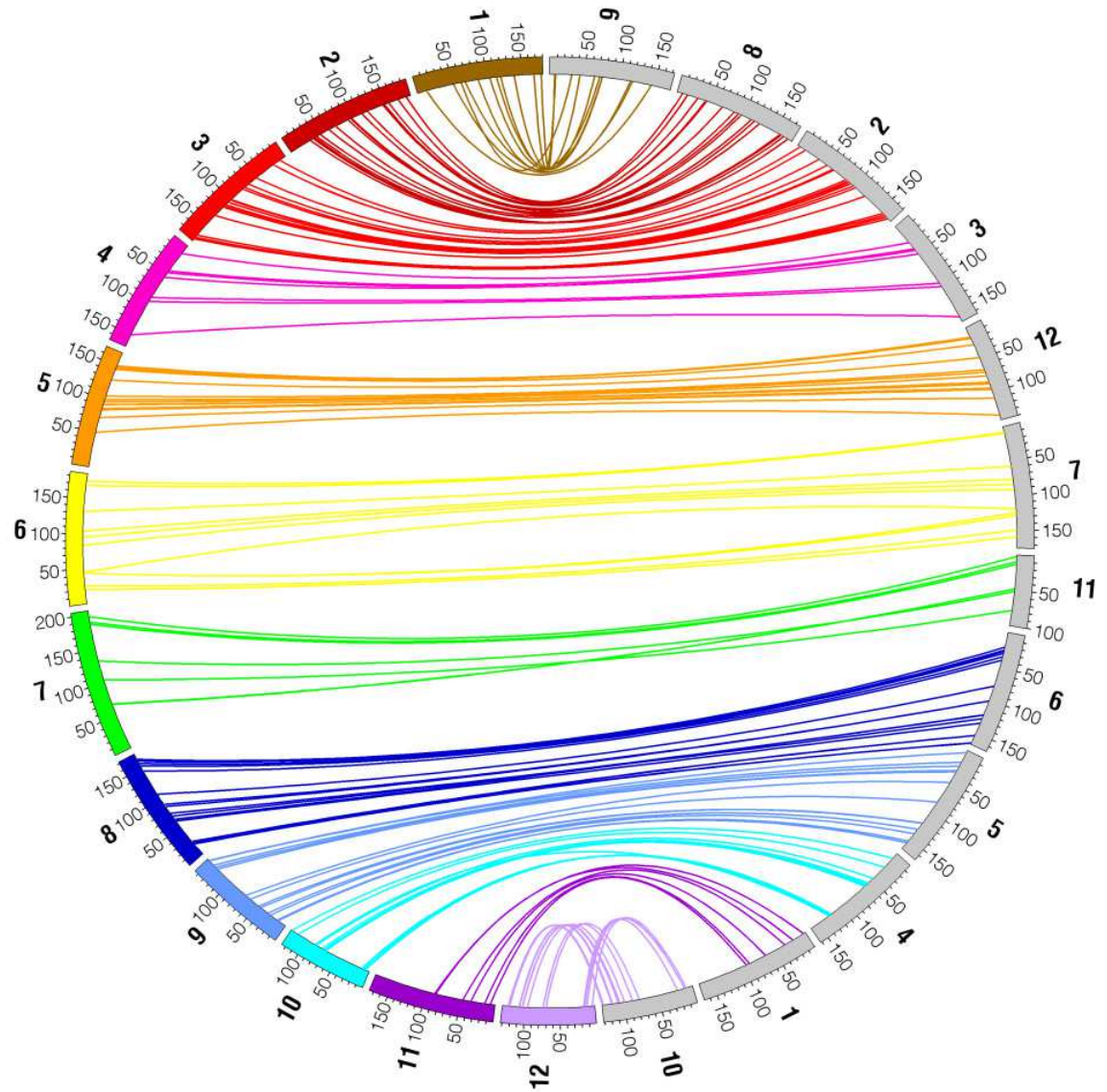
Species	Marker	Reference
<i>Pinus taeda</i> , <i>P.pinaster</i>	EST, COS	Chancerel et al. 2011
<i>Pinus taeda</i>	2466 consensus high-density genetic map using SSR, ESTP, isozymes, RAPD, RFLP, SNPs	Martinez-Garcia et al. 2013
<i>Picea glauca</i> , <i>Pinus taeda</i> , <i>Picea mariana</i> , <i>Pinus pinaster</i>	2770 loci, 1801 genes	Pavy et al. 2012

Comparative mapping *Pinus taeda* and *Pinus pinaster*



Chancellor et al. 2011

Syntenic regions between *Picea glauca* and *Pinus taeda*



Comparative genomics in conifers

(Genome and gene family evolution)

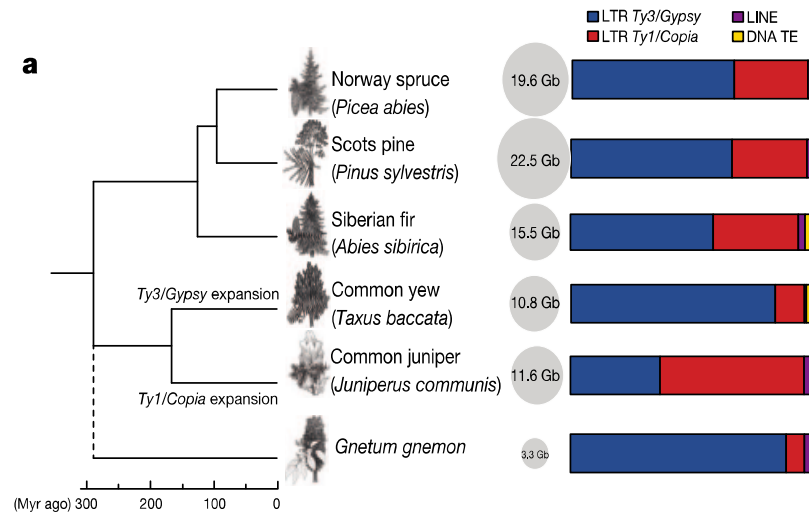
Species		Reference
<i>Pinus spp, Picea spp, Abies fraseri, Tsuga canadensis, Ginkgo biloba</i>	Evolution of intron sizes	Morse et al. 2009
<i>Pinus taeda, several angiosperms</i>	Evolution of repetitive elements	Kovach et al. 2010
<i>Picea glauca, Pinus taeda, Picea mariana, Pinus pinaster</i>	Conservation of gene space Gene family evolution	Pavy et al. 2012
<i>Picea sitchensis, Pinus taeda, Arabidopsis thaliana, Populus trichocarpa</i>	Substitution rates between angiosperms and gymnosperms	Buschiazzo et al. 2012

Comparative genomics in conifers

(Genome and gene family evolution)

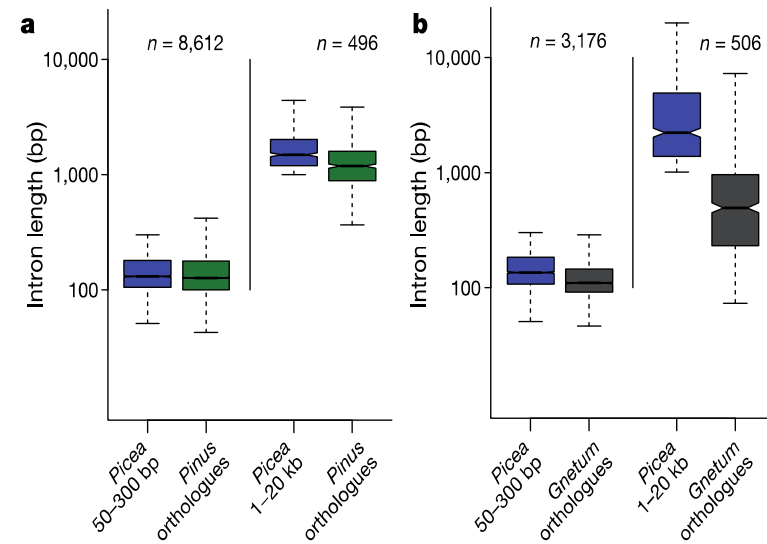
Species		Reference
489 conifer species	Evolutionary history of conifers using phylogenetic analysis based on nuclear and chloroplast genes	Leslie et al. 2012
<i>Cupressaceae spp.</i>	Evolutionary history of Cupressaceae	Pittermann et al. 2012
<i>Picea abies, Pinus sylvestris, Taxus baccata, Juniperus communis, Abies sibirica, Gnetum gnemon</i>	Gene space, gene family evolution and evolution of intron sizes	Nystedt et al. 2013
<i>Picea glauca, Medicago truncatula</i>	Substitution rates and gene family evolution	Pavy et al. 2013

Intron evolution_example

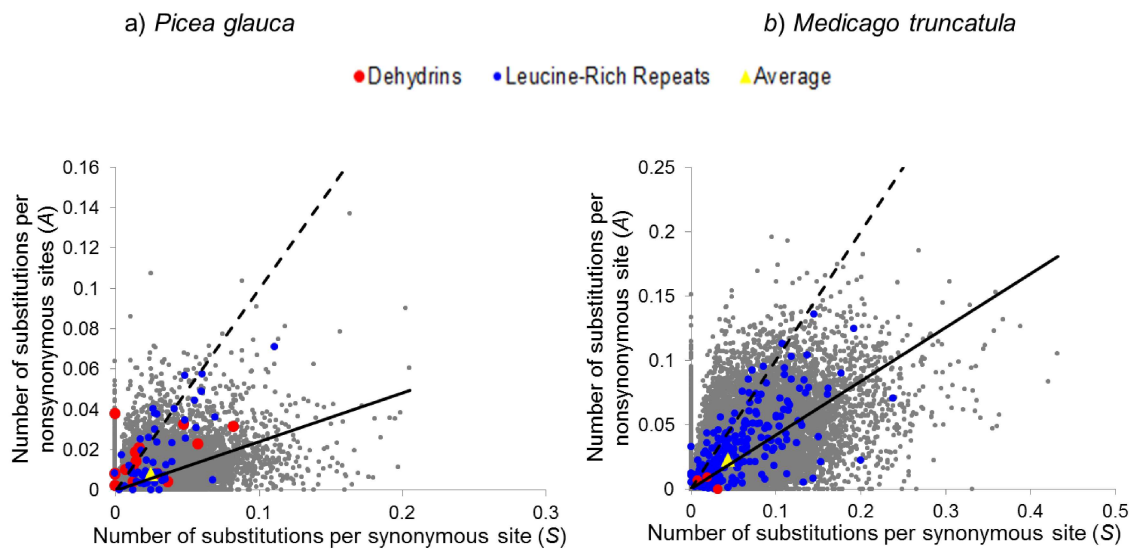


Gymnosperm genomes are mainly composed by repetitive (transposable) elements

Intron size is conserved among gymnosperm species

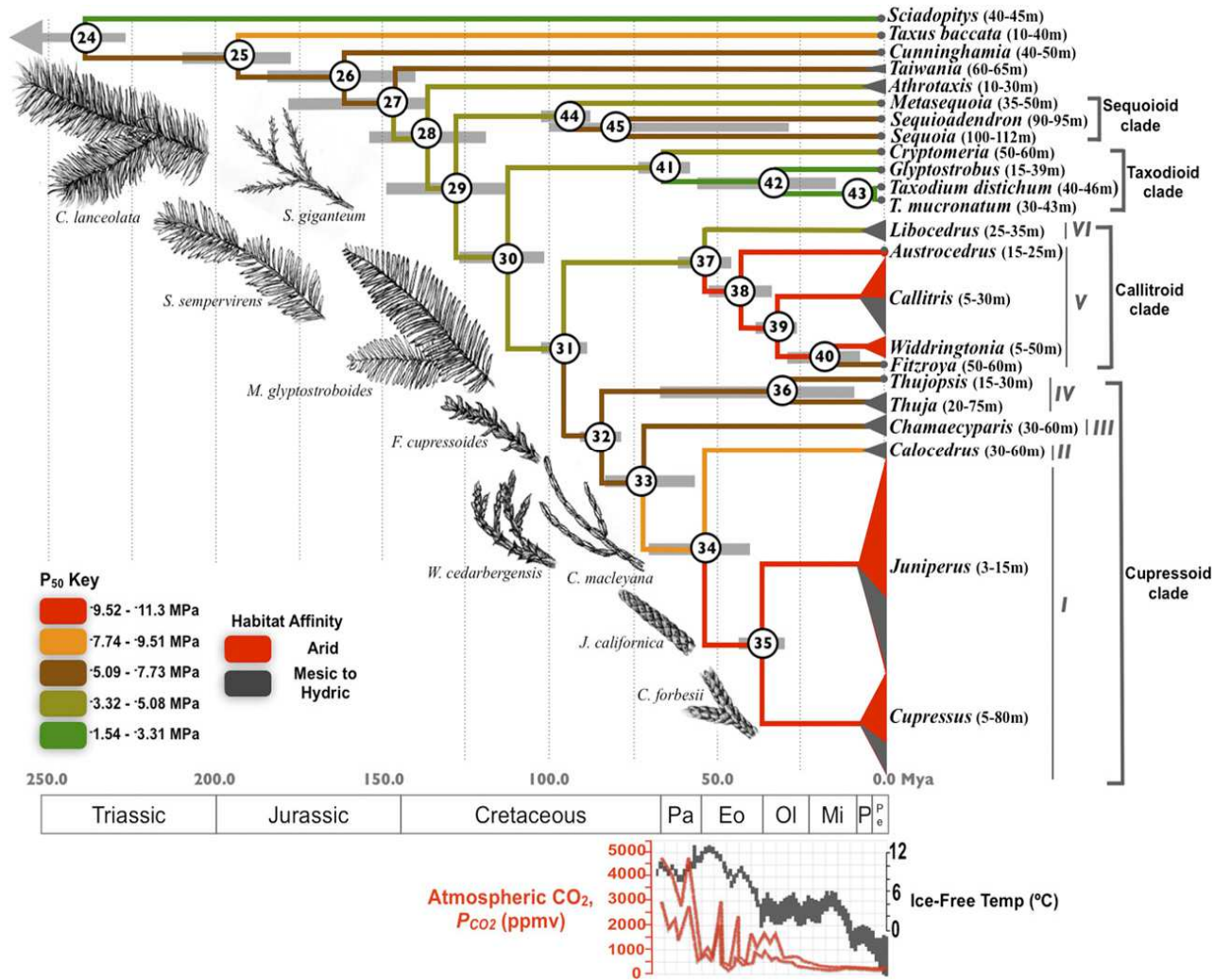


Gene family evolution_example



Dehydrins and LRR gene families have an excess of nonsynonymous SNPs and highest A/S ratios

Evolutionary history_example



Conifer resources

- Conifer Comparative Genomics Project <http://dendrome.ucdavis.edu/ccgp>
- Comparative re-sequencing in Pinaceae <http://dendrome.ucdavis.edu/crsp>
- Comparative Mapping Database <http://dendrome.ucdavis.edu/cmap/index.php?linkto=CMapHome>

Conifer resources

- *Norway spruce*

CONGENIE <http://congenie.org>

- *White spruce*

SmartForests <http://www.smartforests.ca>

Arborea <http://www.arborea.ulaval.ca>

COMPARATIVE GENOMICS CONIFERS (PROCOGEN PROJECT)

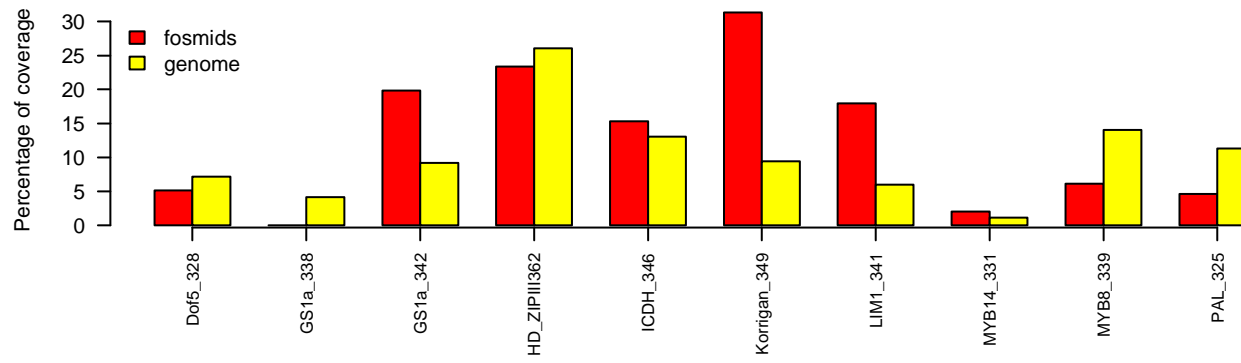
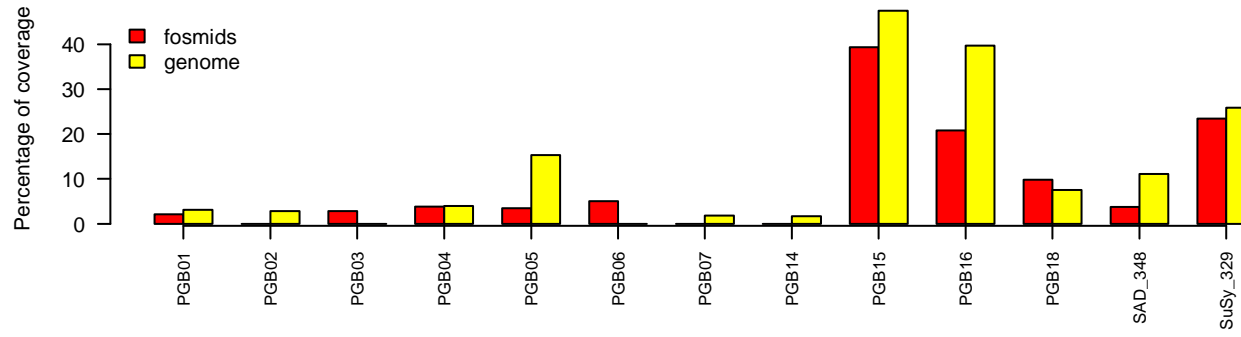
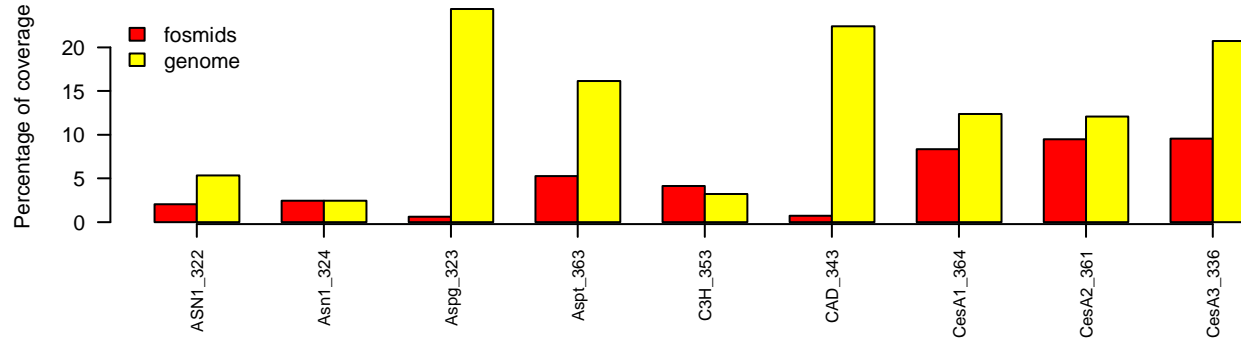
Project1: Micro-synteny of *Picea abies* and *Picea glauca*

- Diploid whole genome shotgun data (*P.abies* 1.0 assembly, 19.6 Gb); and 450 fosmid-pools (100-6000 fosmids) from *P.abies* were compared to 38 bacterial artificial chromosomes (BACs) from *P.glauca*.
- Data analysis was done using customized shell and perl scripts. Alignments were made using Nucmer and displayed with mummerplot.

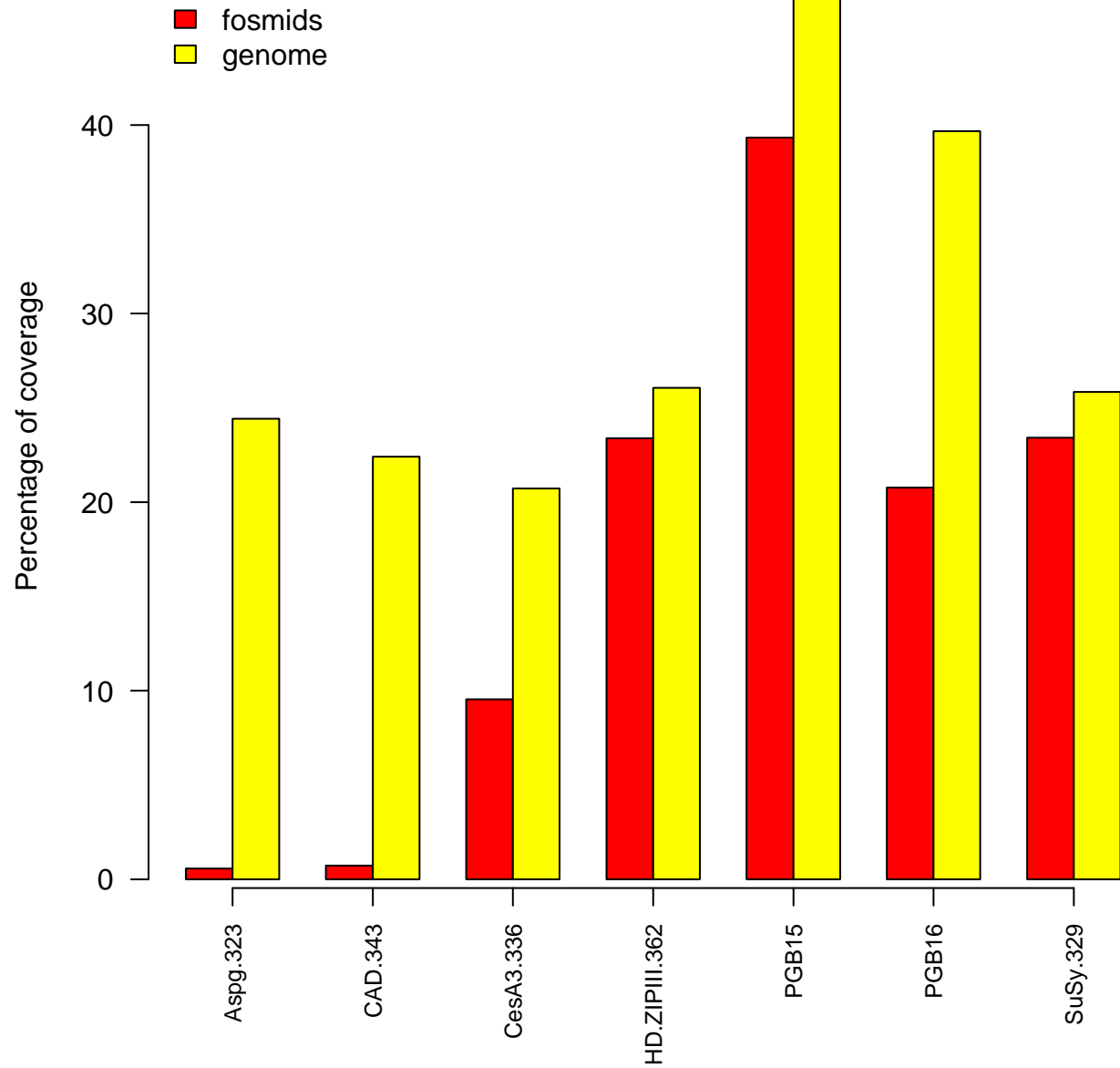
Preliminary results

- We found that the number of BACs with at least one match to *P.abies* varied from 28 (fosmid-pools) to 30 (whole-genome) from a total of 38. The percentage of coverage in the BAC scaffolds by one or several fosmid-pools varied from 0.57 to 47.48%.
- Significant synteny was found in twelve BACs, in which >12% of the BAC sequence was contained in one or several *P.abies* contigs. Ten of these BACs contain important genes in wood formation or nitrogen metabolism in both *P.glauca* and *P.abies*, such as Asparaginase, Cellulose synthase, LIM transcription factor, sucrose synthase, between others.

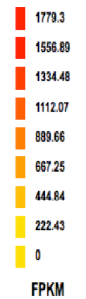
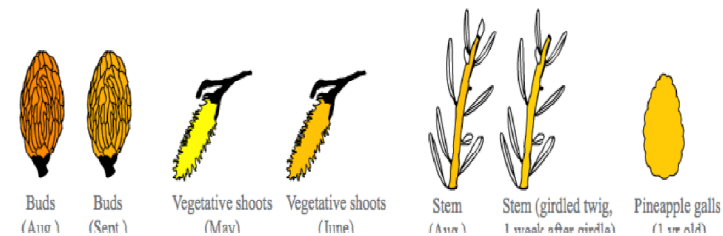
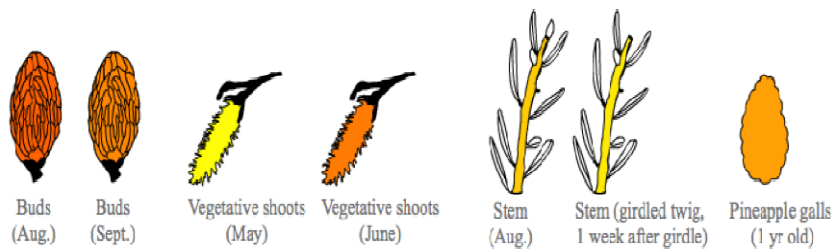
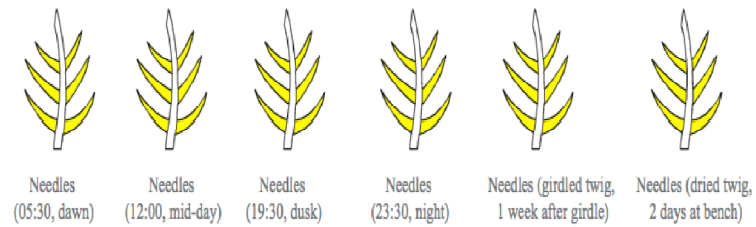
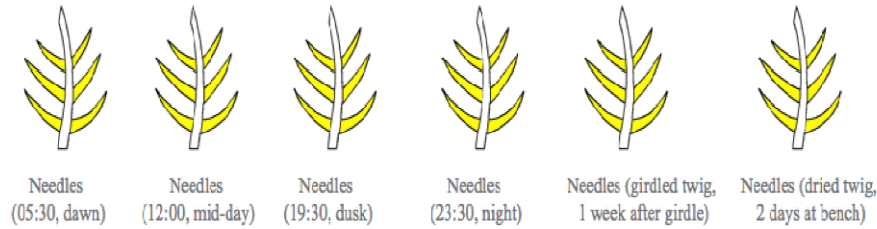
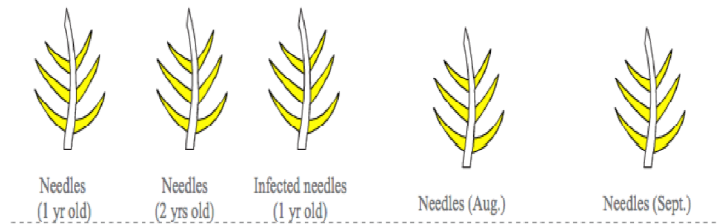
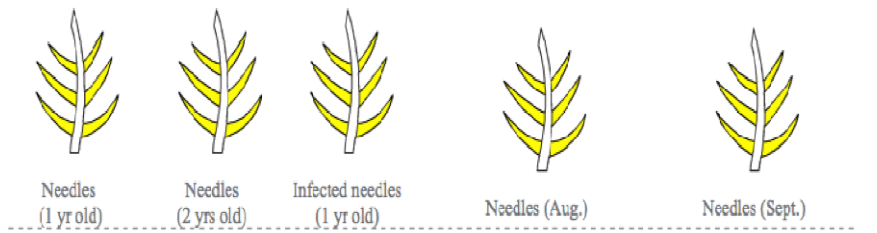
White vs Norway spruce comparisons



BACs with higher synteny



Aspg: L-asparaginase, CAD: Cinnamyl alcohol dehydrogenase, Ces: Cellulose synthase family protein
HD: Homeodomain-leucine zipper transcription factor, SuSy: Sucrose synthase



Gene MA_10429177 (xxx)

Gene MA_123269 (Sucrose synthase)

Conclusions

- Conifers are well suited for comparative analyses.
- Most of the studies were focus on comparative mapping/ genome evolution (evolution of repetitive elements, intron sizes)/gene family evolution and phylogenetic studies.
- Development on new sequencing technologies and whole-genome sequencing of conifers provides a great opportunity for the study of conifer evolution.

Thank you