Phylostratigraphic approach in evolutionary analysis: comparison of methods

Tatiana Martusheva[3], Zakhar Mustafin[1,2], Sergey Lashin[1,2,3]

1 – Kurchatov Genomics Center
2 – Institute of Cytology and Genetics
3 – Novosibirsk State University
Novosibirsk, Russia

Novosibirsk, 2020

Phylostratigraphic tree:

	Clade-specific genes (Number)	(%)	Age (Ma)
Arabidopsis thaliana	1084	(4.0%)	0
Arabidopsis	216	(0.8%)	13
Camelineae	172	(0.6%)	23
Brassicaceae	900	(3.3%)	31
Brassicales	32	(0.1%)	71
Malvids	146	(0.5%)	96
Rosids	218	(0.8%)	104
Core eudicotyledons	510	(1.9%)	125
Magnoliophyta (flowering)	971	(3.6%)	170
Spermatophyta (seed bearing)	946	(3.5%)	329
Tracheophyta (vascular)	475	(1.7%)	496
Embryophyta (land)	3840	(14.1%)	547
Viridiplantae (green plants)	1031	(3.8%)	968
Eukaryota	8229	(30.3%)	1628
Cellular organisms	8433	(31.0%)	2520
Total: 27203			

Cenozoic · Mesozoic · Phanerozoic · Precambrian

Z. W. Arendsee, L. Li, and E. S. Wurtele, "Coming of age: orphan genes in plants," Trends Plant Sci., vol. 19, no. 11, pp. 698–708, Nov. 2014, doi: 10.1016/J.TPLANTS.2014.07.003.

Phylostratigraphic analysis helps to define the time of the gene emergence. It is done through evaluation of the orthologic gene distribution in the genomes of the organisms, related to various taxonomic groups.

Phylostratigraphic tree is grown/constructed on the basis of gene homology and gene age.

Phylostratum shows the level of the distant phylogenetic node in the whole taxonomic tree, including at least one species from the list of species, which has orthologs of the analyzed gene.

To determine the selection pressure on a gene in phylostratigraphy **Ka/Ks** is also calculated

Ka - the number of non-synonymous substitutions per non-synonymous site.
Ks - the number of synonymous substitutions per synonymous site for each orthologous gene pair.

In this study the results of various gene age calculating methods are compared:

Cellular Organisms	1
Eukaryota	2
Viridiplantae	3
Embryophyta	4
Tracheophyta	5
Magnoliophyta	6
eudicotyledons	7
core eudicotyledons	8
rosids	9
Brassicales	10
Arabidopsis	11
A.thaliana	12

Package myTAI

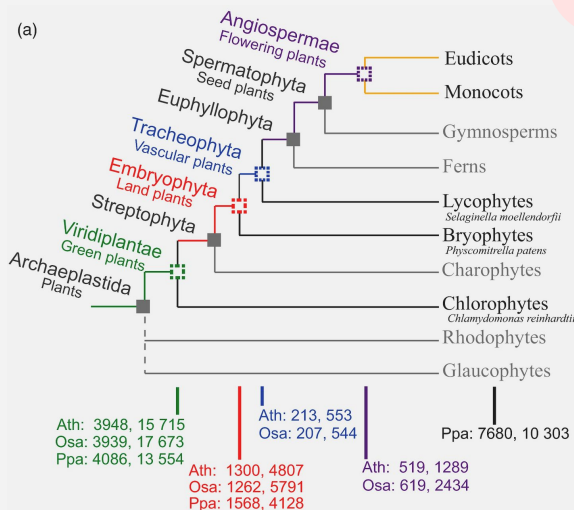TAI calculation - transcriptome age index:
- Orthologous gene search by BLAST

Orthoscape (a cytoscape application)

PAI calculation - phylostratigraphic age index:
- Orthologous gene search on the base of KEGG

Cellular Organisms	0
Eukaryota	1
Viridiplantae	2
Streptophyta	3
Embryophyta	4
Tracheophyta	5
Spermatophyta	6
Magnoliophyta	7
eudicotyledons	8
Gunneridae	9
Pentapetalae	10
rosids	11
malvids	12
Brassicales	13
Brassicaceae	14
Camelineae	15
Arabidopsis	16
Arabidopsis thaliana (thale cress)	17

M. Quint, H.-G. Drost, A. Gabel, K. K. Ullrich, M. Bönn, and I. Grosse, "A transcriptomic hourglass in plant embryogenesis," Nature, vol. 490, no. 7418, pp. 98–101, Sep. 2012, doi: 10.1038/nature11394.

(a)

Angiospermae
Flowering plants

Spermatophyta
Seed plants

Euphyllophyta

Tracheophyta
Vascular plants

Embryophyta
Land plants

Streptophyta

Viridiplantae
Green plants

Archaeplastida
Plants

Eudicots
Monocots
Gymnosperms
Ferns
Lycophytes
Selaginella moellendorfii
Bryophytes
Physcomitrella patens
Charophytes
Chlorophytes
Chlamydomonas reinhardtii
Rhodophytes
Glaucophytes

Ath: 3948, 15 715
Osa: 3939, 17 673
Ppa: 4086, 13 554

Ath: 1300, 4807
Osa: 1262, 5791
Ppa: 1568, 4128

Ath: 213, 553
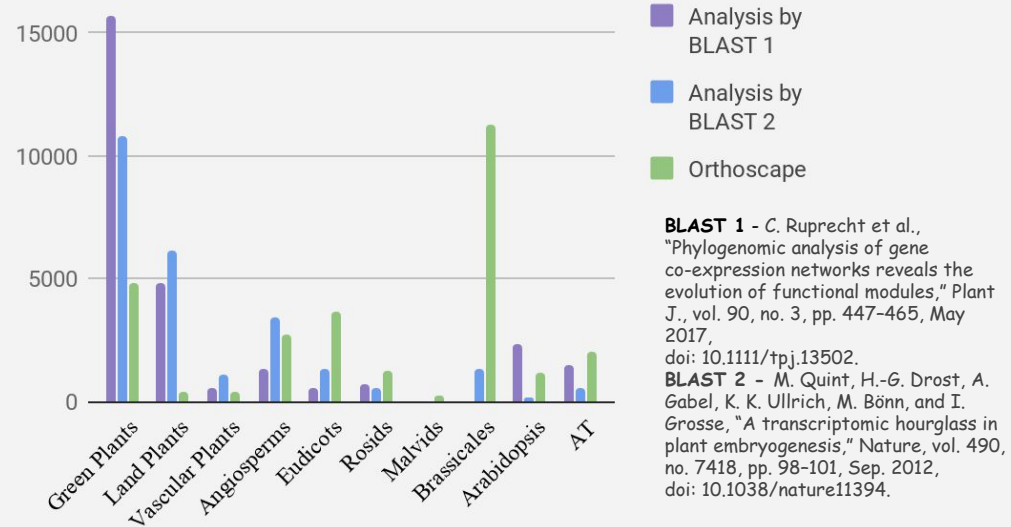Osa: 207, 544

Ath: 519, 1289
Osa: 619, 2434

Ppa: 7680, 10 303

C. Ruprecht et al., "Phylogenomic analysis of gene co-expression networks reveals the evolution of functional modules," Plant J., vol. 90, no. 3, pp. 447–465, May 2017, doi: 10.1111/tpj.13502.

Z. S. Mustafin, S. A. Lashin, Y. G. Matushkin, K. V. Gunbin, and D. A. Afonnikov, "Orthoscape: a cytoscape application for grouping and visualization KEGG based gene networks by taxonomy and homology principles," BMC Bioinformatics, vol. 18, no. S1, pp. 1–9, 2017, doi: 10.1186/s12859-016-1427-5.

The comparison of calculation results phylostratigraphy age genes:

Phylostratum		Genes in phylostratum Analysis by BLAST 1	Genes in phylostratum Analysis by BLAST 2	Orthoscape
Green Plants	1	15708	10781	4789
Land Plants	2	4807	6160	399
Vascular Plants	3	553	1068	426
Angiosperms	4	1289	3443	2708
Eudicots	5	516	1299	3625
Rosids	6	697	527	1246
Malvids	7	22	15	246
Brassicales	8	22	1318	11269
Arabidopsis	9	2318	160	1195
AT	10	1484	522	1987

Legend:
- Analysis by BLAST 1
- Analysis by BLAST 2
- Orthoscape

BLAST 1 - C. Ruprecht et al., "Phylogenomic analysis of gene co-expression networks reveals the evolution of functional modules," Plant J., vol. 90, no. 3, pp. 447–465, May 2017, doi: 10.1111/tpj.13502.
BLAST 2 - M. Quint, H.-G. Drost, A. Gabel, K. K. Ullrich, M. Bönn, and I. Grosse, "A transcriptomic hourglass in plant embryogenesis," Nature, vol. 490, no. 7418, pp. 98–101, Sep. 2012, doi: 10.1038/nature11394.

Comparison of the results of the studies using BLAST.
The researchers chose different phylostratigraphic trees, where they pointed the key periods in organism evolution.
- Correlation coefficient is 0.74
- Choice of the key taxons on the phylostratigraphic tree doesn't influence much on the gene age, when been analyzed by the same method.

Comparison of the myTai and Orthoscape results :
- Correlation coefficient is 0.27
- The results are completely different when using these methods.
- Orthoscape found considerably more young genes, while myTai found more ancient genes.