



Mobilo ģenētisko elementu izplatība priežu gēnu rajonos

Angelika Voronova, Martha Rendon, Par Ingvarsson, Dainis Ruņģis

Latvijas Valsts mežzinātnes institūts “Silava”; Zviedrijas Lauksaimniecības Zinātņu universitāte



Valsts izglītības
attīstības aģentūra

NACIONĀLAIS
ATTĪSTĪBAS
PLĀNS 2020



EIROPAS SAVIENĪBA
Eiropas Reģionālās
attīstības fonds

IEGULDĪJUMS TAVĀ NĀKOTNĒ

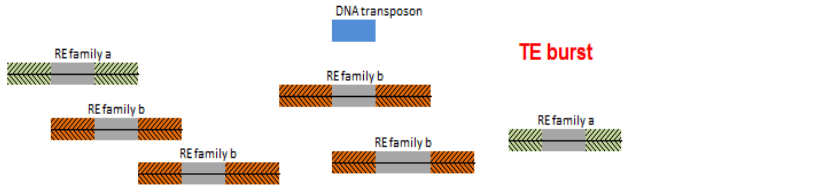


In a regular state mobile elements are inactivated by methylation; regions of TE form of a heterochromatin state of DNA

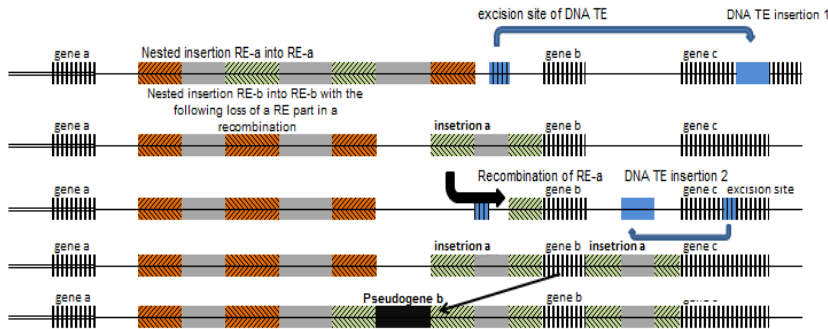


Unfavourable conditions/ Stress

In a stress conditions some of the non-coding regions of DNA could be relaxed and transcribed, some of TEs from those regions could transpose or replicate



in the next generations new genotypes form that could affect phenotype/ or not



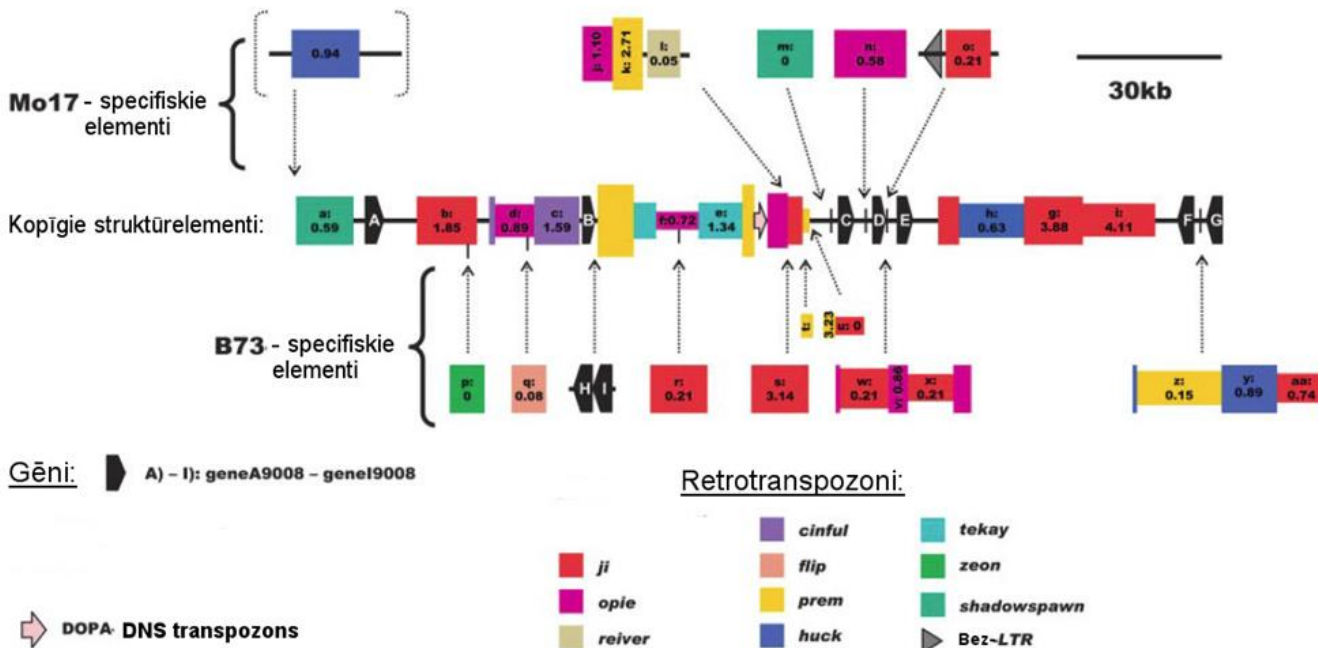
Natural selection

Only genotypes with preferences or silent mutations will be passed to the next generations.

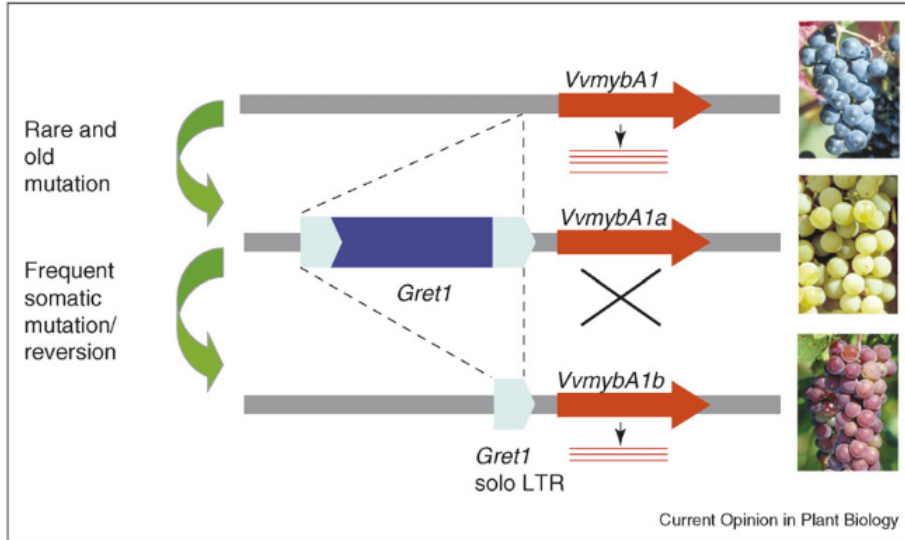
Modern genomes of higher plants contain thousands of mobile elements in the silent regions of DNA; all kinds of nested repeats and pieces of different TE types in the gene UTR and introns, that may have a regulative effect on genes; pseudogenes that could be reshuffled by new insertions forming new genes.



PostDoc
Latvia



Kukurūzas (*Zea mays*) 2S hromosomas fragmenta salīdzinošā shēma starp divām kukurūzas inbrīdām līnijām Mo17 un B73 (modificēts no Brunner *et al.* 2005).



Kobayashi S, Goto-Yamamoto N, Hirochika H: Retrotransposon- induced mutations in grape skin color. *Science*, 2004,304:982.

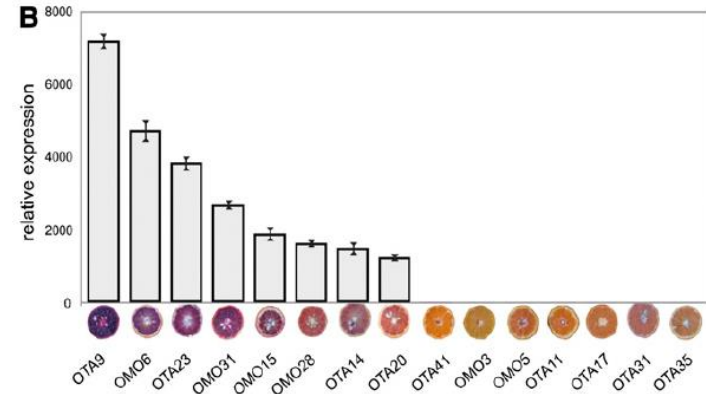


Figure 4. Expression Analysis of *Ruby*.

(A) Expression of *Ruby* in different tissues of blond Cadenera and blood Moro oranges. Albedo is the spongy white layer of the orange peel. The flavedo is the peripheral surface of the pericarp.

(B) Expression of *Ruby* in diploid hybrids obtained crossing clementine (*C. clementina* cv Oroval) with *C. sinensis* cv Moro (OMO series) or *C. sinensis* cv Tarocco (OTA series). Hybrids lacking anthocyanins (OMO41, 3, and 5; OTA11, 17, 31, and 35) appear slightly different colors due to varying carotenoid levels in the fruit flesh. Error bars show SE of the mean.

Retrotransposons Control Fruit-Specific, Cold-Dependent Accumulation of Anthocyanins in Blood Oranges (Butelli *et.al.* 2012), *The Plant Cell*, 2012 July 2012, 24 (7)

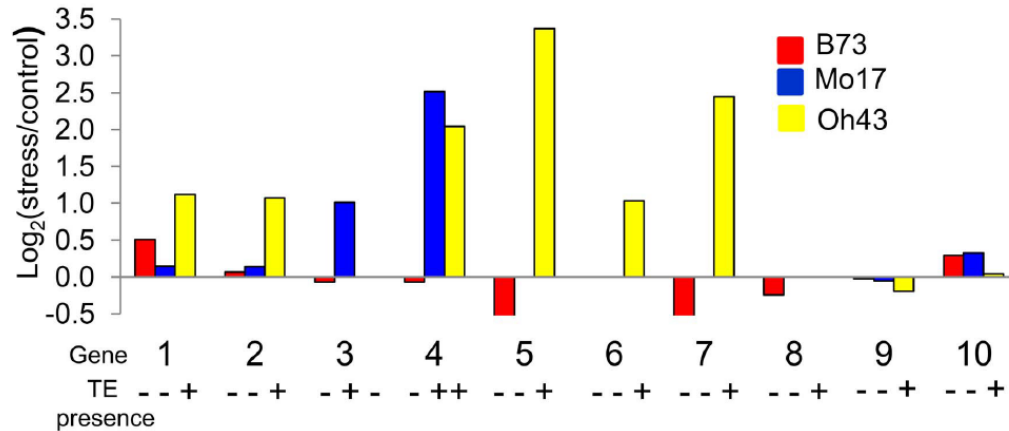


Fig. 3. Stress-induced up-regulation of gene expression correlates with the variation in TE presence. (A) Proportion of genes up-regulated in B73 that are also up-regulated in Mo17 and Oh43 is shown for all TE families under the stress condition with highest enrichment for the TE family. (B) The relative expression levels in stress compared to control treatments (\log_2 ratio) is shown for B73, Mo17, and Oh43 for each of the 10 expressed genes that are polymorphic for insertions of TEs. The presence/absence of the TE for each genotype-inbred combination is shown by '+' and '-' symbols. The genes are as follows: 1-GRMZM2G102447; 2-GRMZM2G108057; 3-GRMZM2G071206; 4-GRMZM2G108149; 5-GRMZM2G400718; 6-GRMZM2G347899; 7-GRMZM2G517127; 8-GRMZM2G378770; 9-GRMZM2G177923; 10- GRMZM2G504524. All genes with TE insertion polymorphism are listed in S8 Table.

doi:10.1371/journal.pgen.1004915.g003

Makarevitch *et al.*, 2015

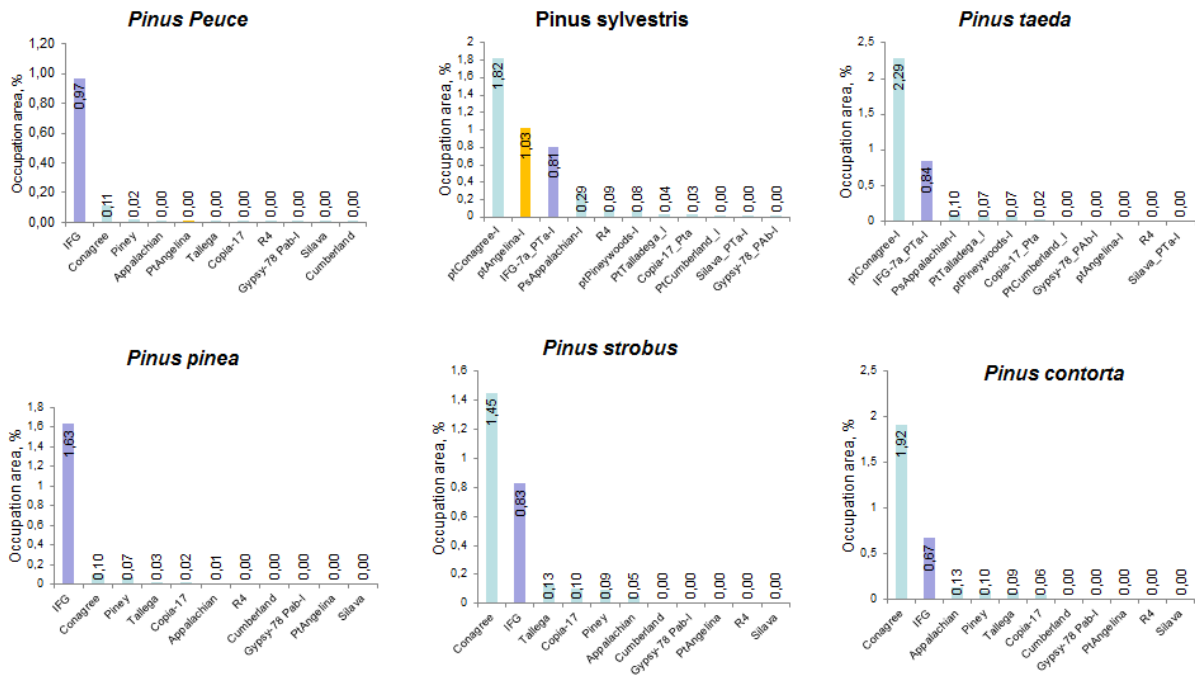


Figure 1. Occupation area (%) of RE families relative to species average genome size.

It was assumed that each estimated copy represents full-length element. Estimation of the copy number of eleven REs was performed using Real-time PCR absolute quantification with Maxima SYBR Green/ROX qPCR Master Mix (*Thermo Scientific*) reagents and StepOne software v2.2.2 (*Applied Biosystems*). Plasmids with cloned RE sequences were used for standard curves (6 dilutions 1:10; 3 replicates), for plasmid with a known insert sequence, molecular weight was calculated using the Sequence Manipulation Suite: DNA Molecular Weight (Stothard, 2000). Plasmid copy number was calculated using the formula: copy nb. = (amount, ng) * Avogadro nb. ($6.022 \cdot 10^{23}$) / $1 \cdot 10^9$ * (mol weight, Da). Copy number of each RE was calculated relative to the amount of DNA analysed and the genome size (2C) of the various species.



Pētījuma mērķis ir identificēt parastās priedes (*Pinus sylvestris* L.) gēnus saskaitītus ar transponējamo elementu izraisītajām strukturālajām izmaiņām

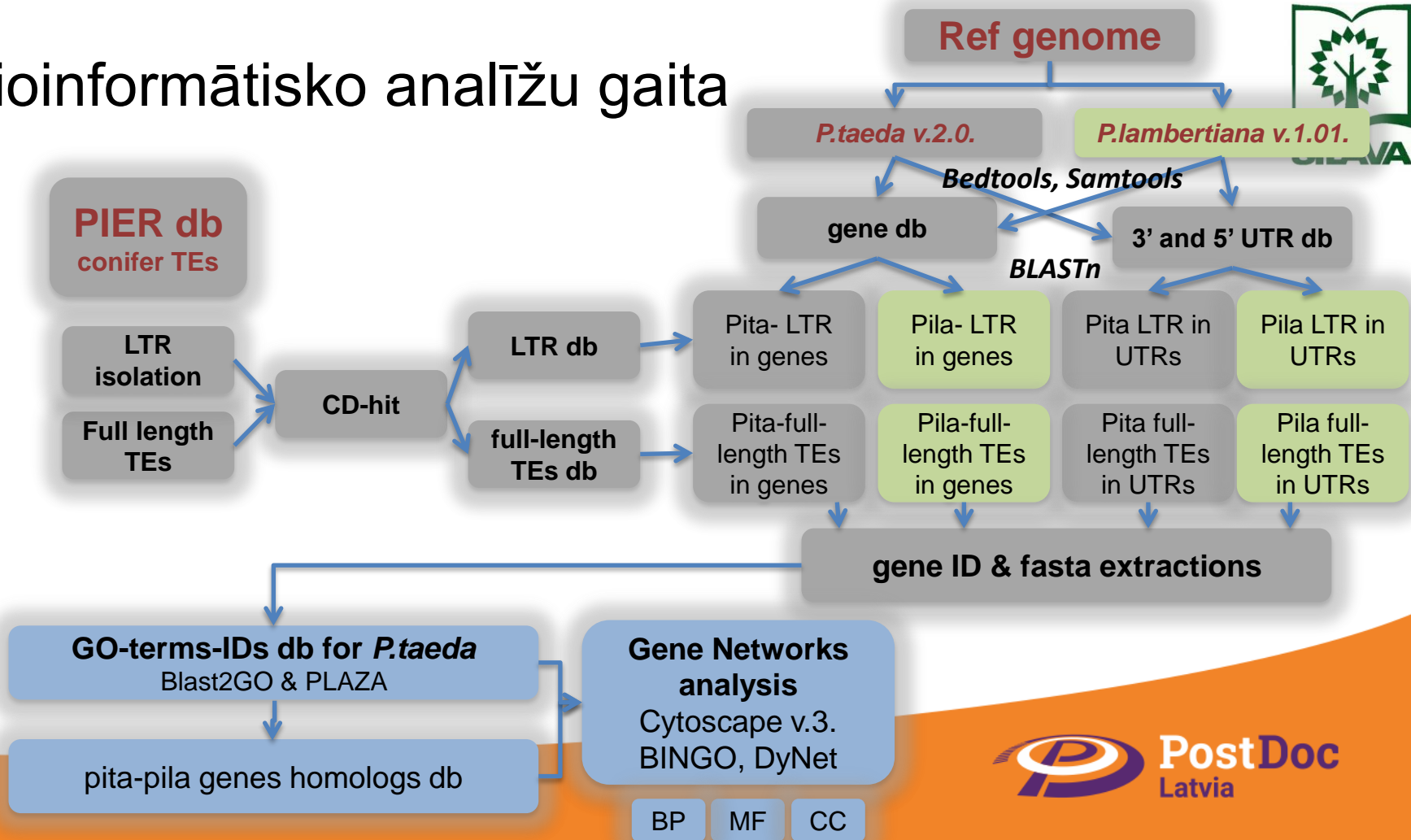
Pirmais pētījuma posms- bioinformātiskā pieejamo priežu genomu analīze



Pinus taeda/ P. lambertiana/ P.sylvestris

- *Pinus taeda* v.2.01 (6,58 GB; 36 730 gēnu)
- *Pinus lambertiana* v.1.0.(HQ-8 779;LQ- 71 167)
- *Pinus sylvestris* unannotated scaff (474 gēnu- *P.taeda* homologu, 12 737 eksonu , no tiem 2021 aizņēma visu kontingu)
- TE atkārtotamība
- NGS sekvenču kvalitāte
- Transponējamo elementu datubāzes kvalitāte, automātiskā anotēšana
- Genomu versiju atšķirības
- GO anotācija jaunākai genoma versijai vēl nav pieejama

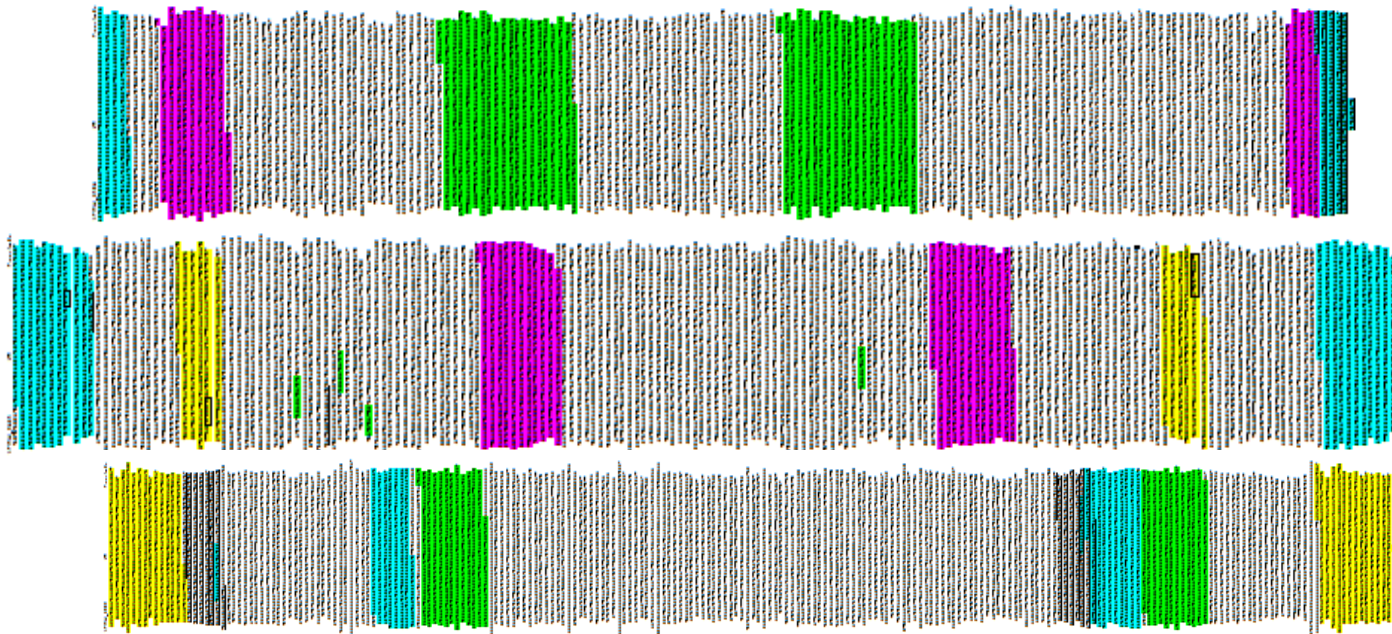
Bioinformātisko analīžu gaita



PIER v.2.0 skujkoku transponējamo elementu datubāze

Transponējamo elementu datubāze (19 700-- 15 622 garumā no 257 -35042 bp) Total CPU time 126419.09

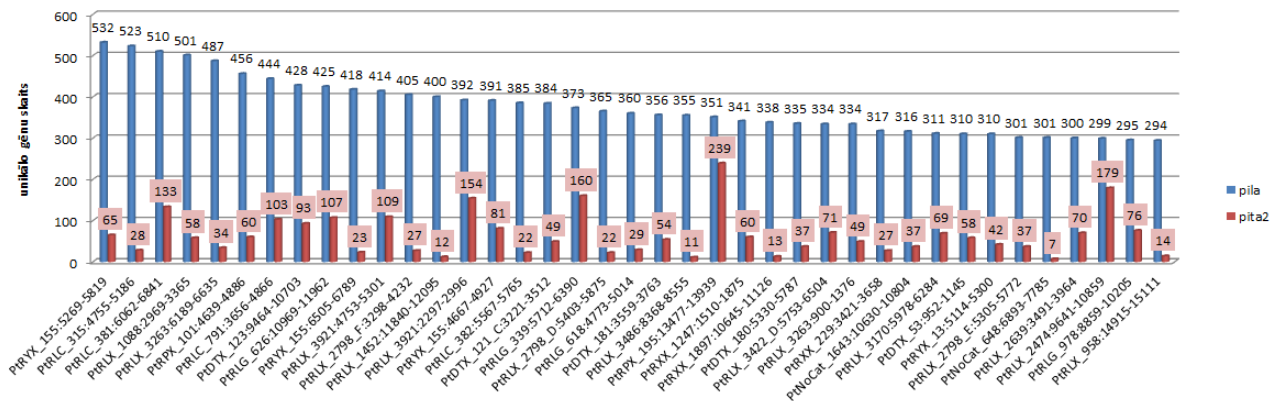
Garu terminālo atkārtojumu (LTR) datubāze (24 591-- 9 659) Total CPU time 1515.90



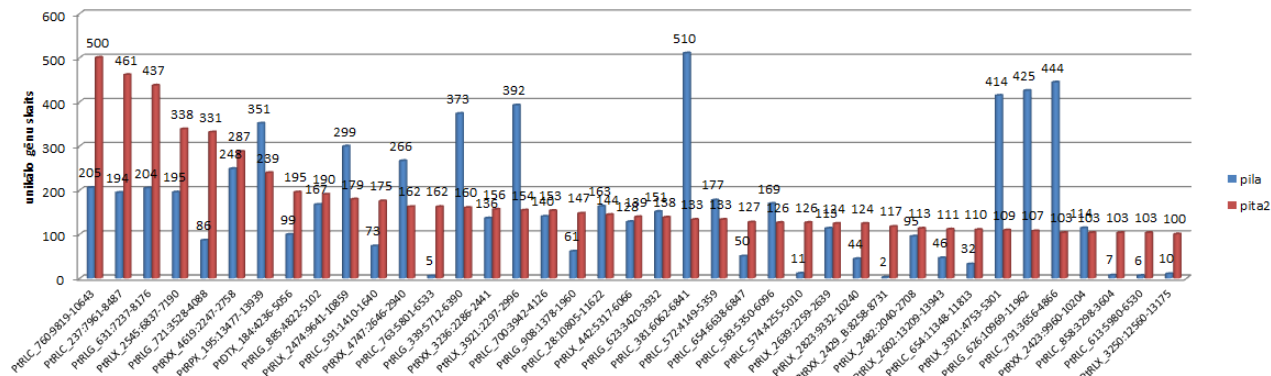
Mobilo elementu izplatība gēnu rajonos, %

TE class/superfamily	Pilna izmēra TE gēnos	LTR gēnos	pilns RE gēna UTR	RE ≥50% from UTR	LTR in UTR
DNS transpozoni	17	4	37	4	4
Retrotranspozoni (RE)	66	89	34	95	90
Bez klasifikācijas	14	6	29	1	6
LTR Retrotranspozoni	84	57	-	90	56
<i>Copia</i>	13	24	-	28	20
<i>Gypsy</i>	17	16	-	23	16
bez-LTR RE	8	7	43	10	7
Nezināmi LTR RE	54	-	43	40	36
Nezināmi RE	8	37	14	2	38

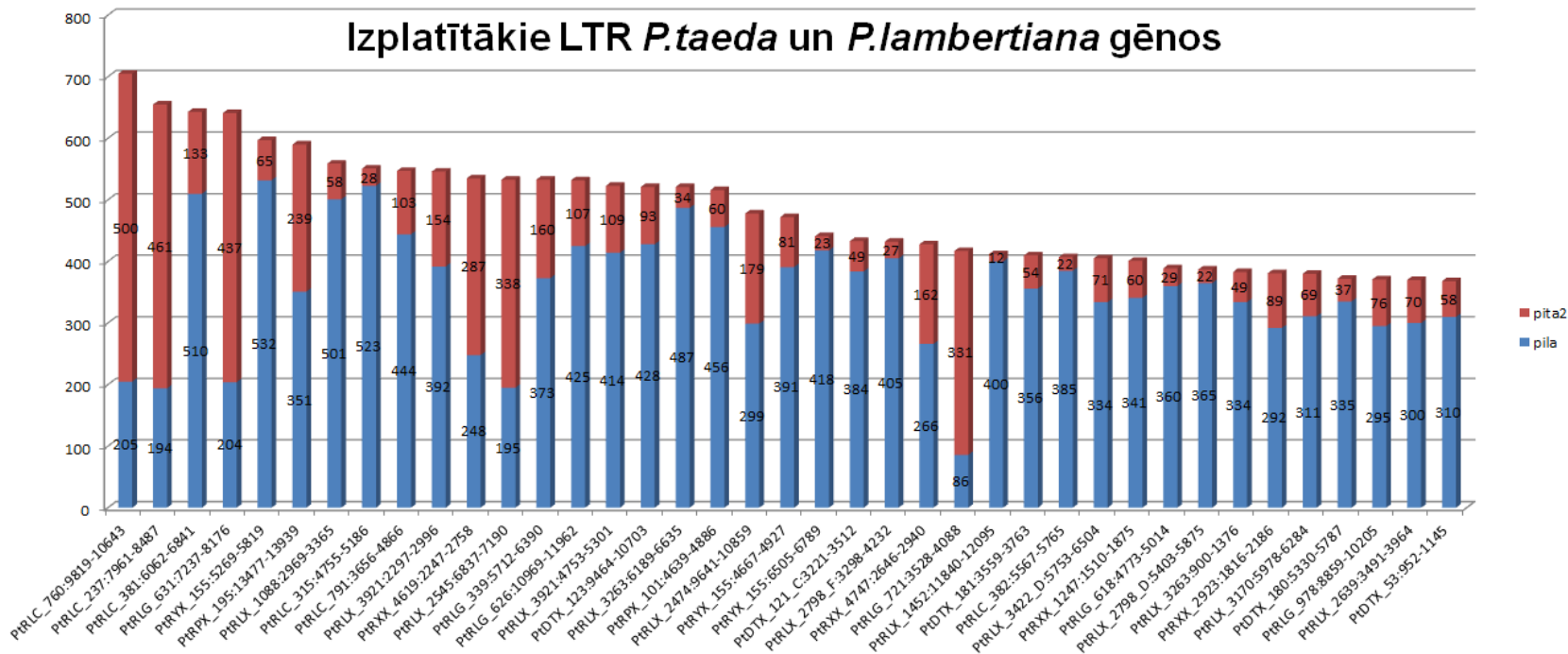
P.lambertiana 40 izplatītākie LTR gēnos

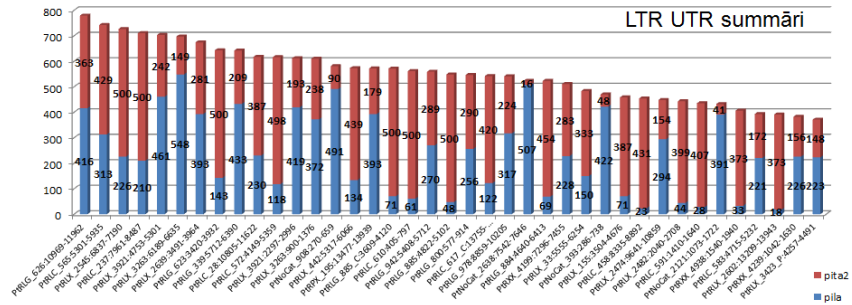
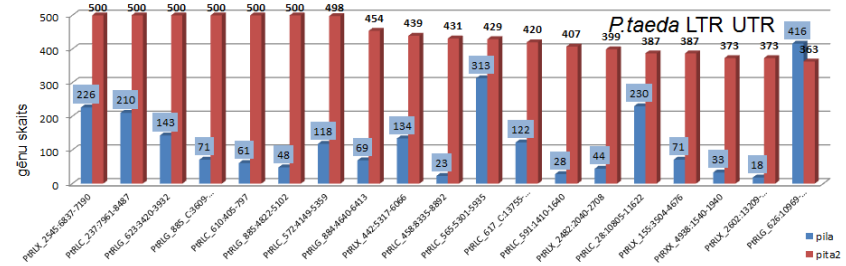
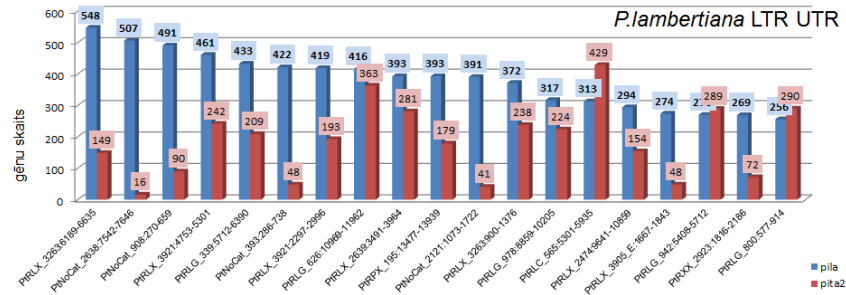


P.taeda 40 izplatītākie LTR gēnos



Izplatītākie LTR *P.taeda* un *P.lambertiana* gēnos





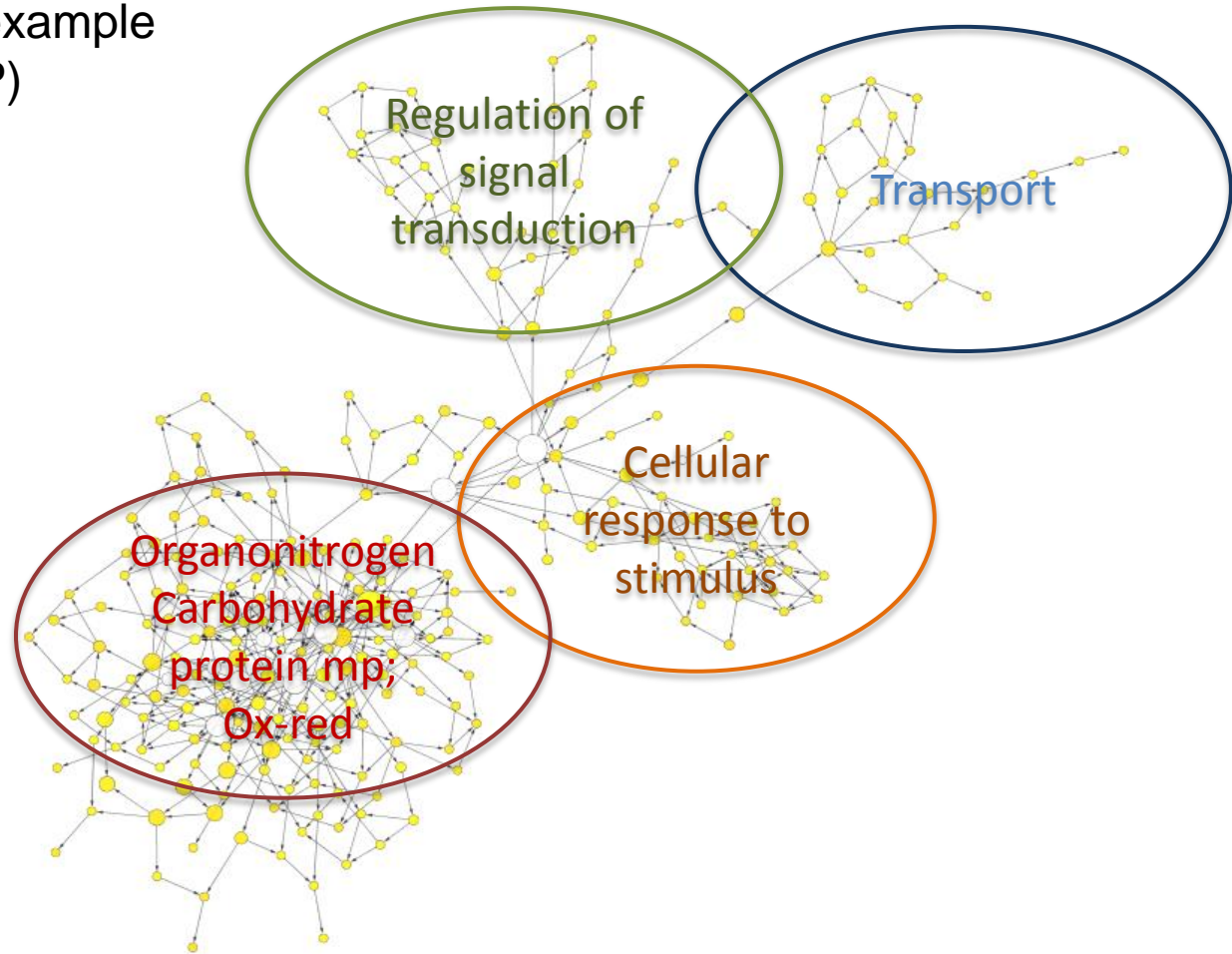


Gēnu ontoloģijas (GO) analīze

- No 36 730 gēniem 15534 varēja pieskirt GO kategoriju- 46663 ierakstu, no kuriem 46589 ir jaunākajā Gene Ontology Consortium anotācijā, 1845 unikālo GO
- *P.lambertiana* 22 282 gēnu, 1368 unikālo GO

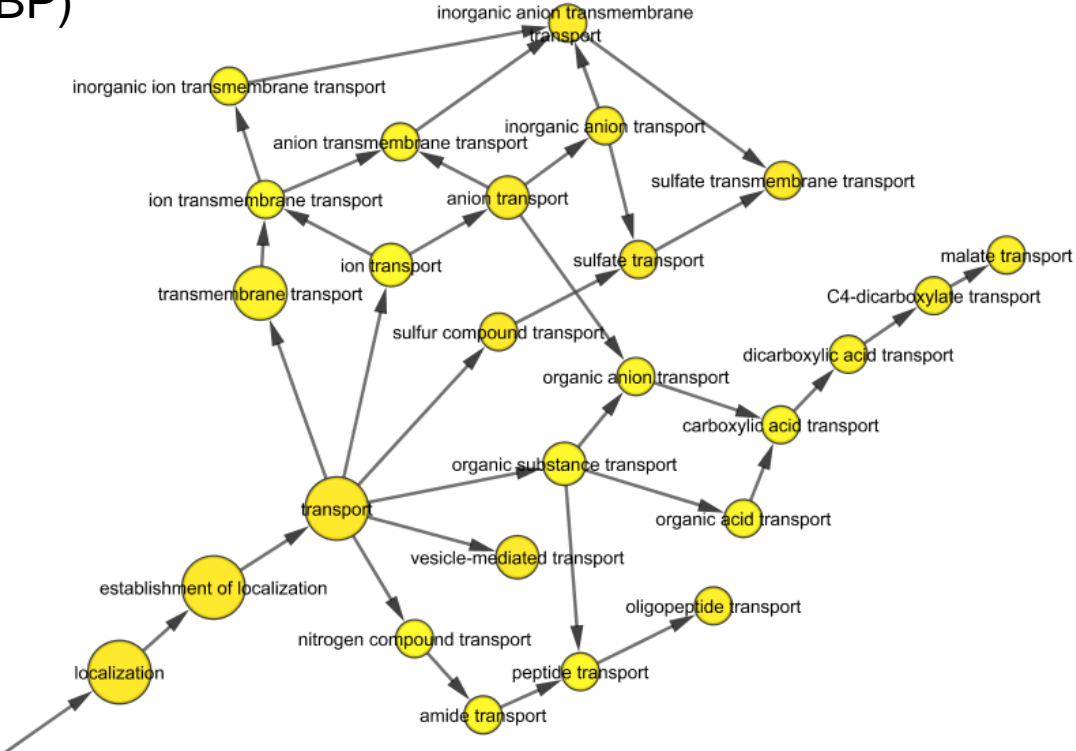
Gene network example

LTR in UTR (BP)



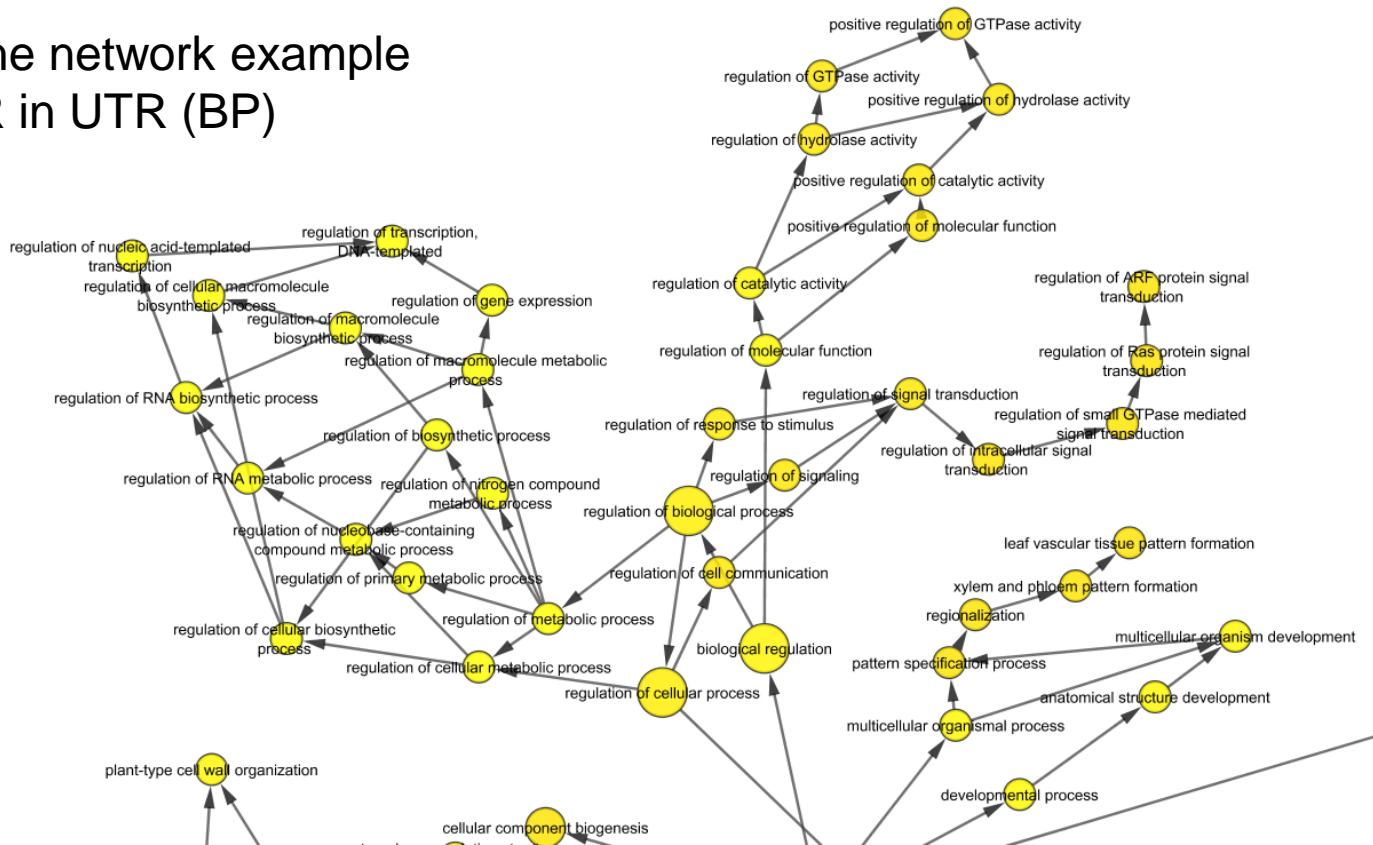
Gene network example

LTR in UTR (BP)



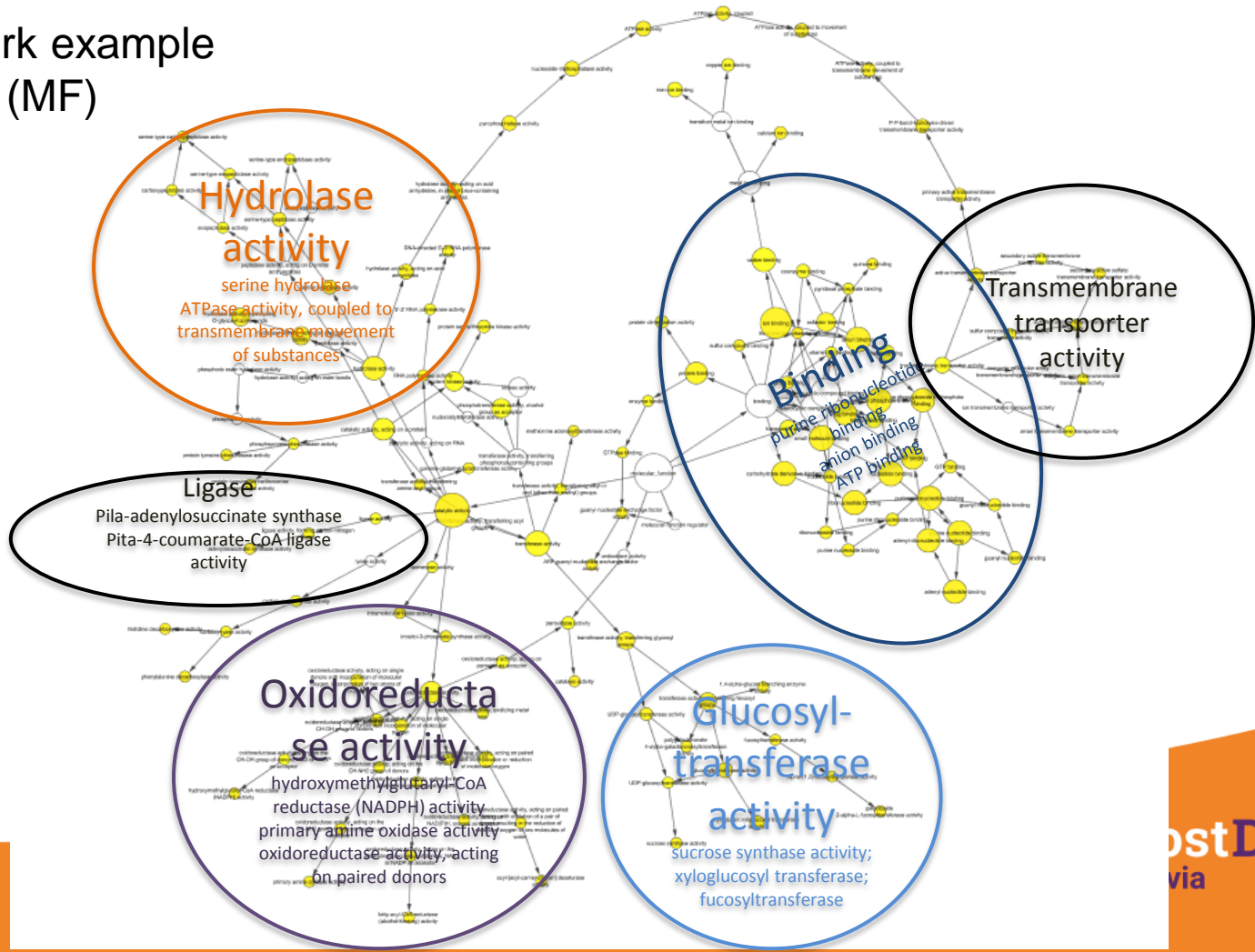
Gene network example

LTR in UTR (BP)



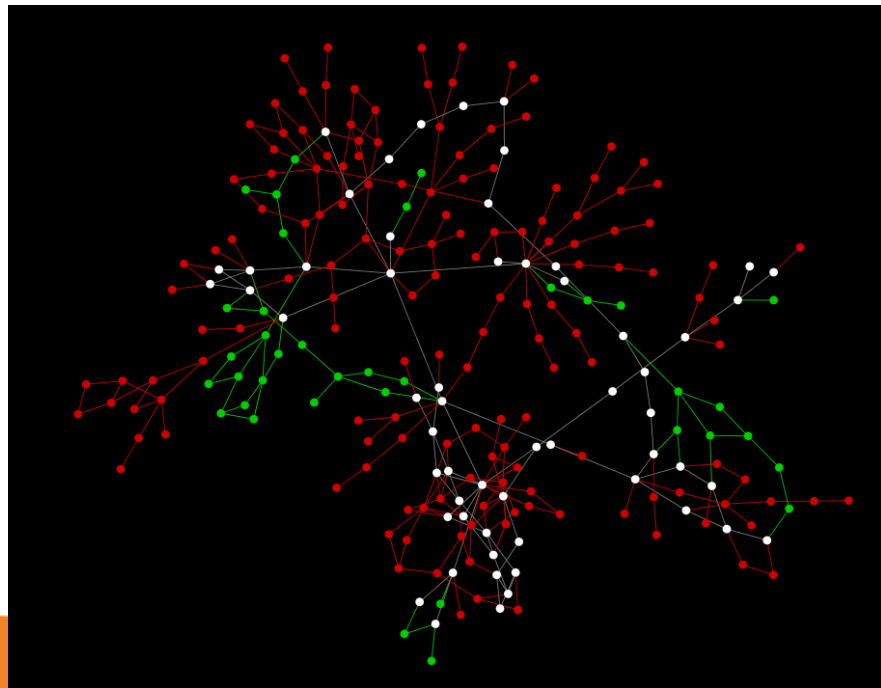
Gene network example

LTR in UTR (MF)

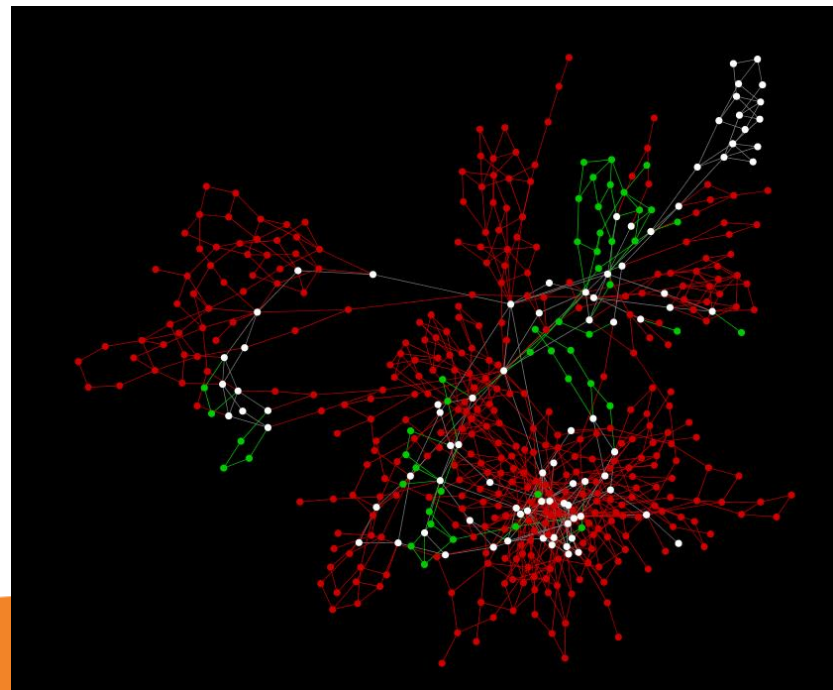


Gēnu tīklu salīdzinājums

MF compare *RLC760*



BP compare *RLC760*



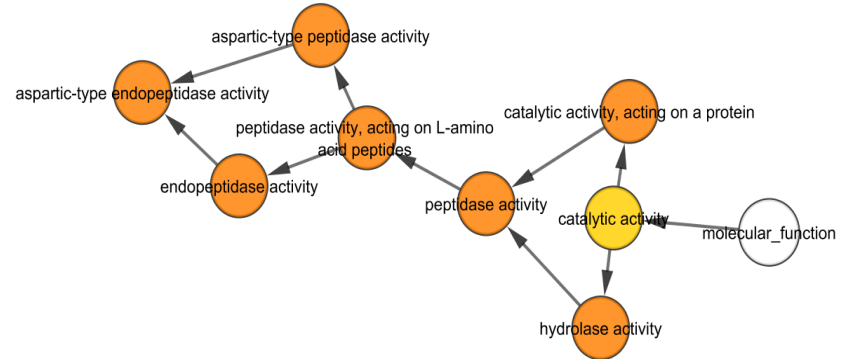
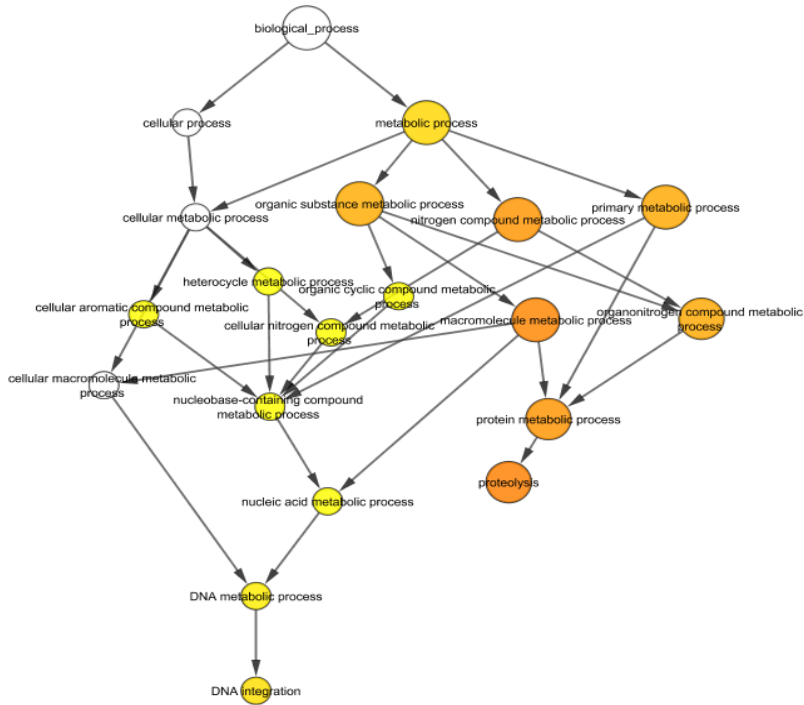
PITA-moIF				PILA moIF			
GO-ID	p-value	Nb	Description	GO-ID	p-value	Nb	Description
5488	1.7487E-1	143	binding	5488	1.4192E-2	26	binding
1901363	7.1010E-2	116	heterocyclic compound binding	1901363	9.4587E-2	20	heterocyclic compound binding
97159	7.2009E-2	92	organic cyclic compound binding	97159	9.5063E-2	20	organic cyclic compound binding
3824	5.5409E-1	112	catalytic activity	3676	1.6884E-3	17	nucleic acid binding
43167	1.0261E-1	92	ion binding	43167	1.3789E-1	16	ion binding
36094	1.2952E-3	58	small molecule binding	46872	1.0940E-3	15	metal ion binding
43168	5.8088E-4	57	anion binding	43169	1.1606E-3	15	cation binding
166	1.9503E-3	56	nucleotide binding	46914	1.3953E-3	13	transition metal ion binding
1901265	1.9503E-3	56	nucleoside phosphate binding	8270	9.9904E-4	11	zinc ion binding
3676	8.1169E-1	54	nucleic acid binding	5515	2.3223E-1	6	protein binding
16740	3.9280E-1	51	transferase activity	46983	4.6775E-4	4	protein dimerization activity
32553	2.9812E-2	50	ribonucleotide binding	16791	1.9519E-2	2	phosphatase activity
97367	4.4430E-3	50	carbohydrate derivative binding	42578	2.3927E-2	2	phosphoric ester hydrolase activity
32555	4.1394E-3	49	purine ribonucleotide binding	16788	5.9186E-1	2	hydrolase activity, acting on ester bonds
17076	4.2202E-3	49	purine nucleotide binding	3677	6.4931E-1	2	DNA binding
32559	3.2736E-3	48	adenyl ribonucleotide binding	9881	7.4290E-3	1	photoreceptor activity
30554	3.2736E-3	48	adenyl nucleotide binding	3868	9.8935E-3	1	4-hydroxyphenylpyruvate dioxygenase activity
8144	3.4752E-2	39	drug binding	43682	9.8935E-3	1	copper-transporting ATPase activity
35639	5.0937E-2	38	purine ribonucleoside triphosphate binding	5375	9.8935E-3	1	copper ion transmembrane transporter activity
5524	4.3112E-2	37	ATP binding	16868	1.7252E-2	1	intramolecular transferase activity, phosphotransferase activity, coupled to transmembrane transport
16772	5.0197E-1	36	transferase activity, transferring phosphorus-containing group	15662	1.9693E-2	1	ATPase activity, coupled to transmembrane transport
5515	1.4815E-1	33	protein binding	42803	1.9693E-2	1	protein homodimerization activity
140096	4.3284E-1	33	catalytic activity, acting on a protein	4673	2.2128E-2	1	protein histidine kinase activity
16301	1.7148E-1	31	kinase activity	16775	2.2128E-2	1	phosphotransferase activity, nitrogenous group transferase activity
16787	8.6520E-1	29	hydrolase activity	155	2.2128E-2	1	phosphorelay sensor kinase activity
16773	1.9432E-1	28	phosphotransferase activity, alcohol group as acceptor	16866	2.2128E-2	1	intramolecular transferase activity
16491	5.2521E-1	27	oxidoreductase activity	42802	2.4557E-2	1	identical protein binding
4672	2.9353E-1	26	protein kinase activity	18024	2.4557E-2	1	histone-lysine N-methyltransferase activity
48037	3.2331E-1	17	cofactor binding	16278	2.4557E-2	1	lysine N-methyltransferase activity
3677	4.5616E-1	15	DNA binding	16279	2.4557E-2	1	protein-lysine N-methyltransferase activity
16788	4.1005E-1	14	hydrolase activity, acting on ester bonds	42054	2.6980E-2	1	histone methyltransferase activity
5215	3.2399E-1	12	transporter activity	8276	2.9398E-2	1	protein methyltransferase activity
43531	1.1650E-2	11	ADP binding	8170	2.9398E-2	1	N-methyltransferase activity
140098	8.3482E-2	10	catalytic activity, acting on RNA	46915	3.9011E-2	1	transition metal ion transmembrane transporter activity
22857	4.6189E-1	10	transmembrane transporter activity	42393	3.9011E-2	1	histone binding
4519	2.9334E-1	9	endonuclease activity	22853	4.1400E-2	1	active ion transmembrane transporter activity
4518	3.4798E-1	9	nuclease activity	19829	4.1400E-2	1	cation-transporting ATPase activity
20037	6.3222E-1	9	heme binding	42625	4.1400E-2	1	ATPase coupled ion transmembrane transporter activity
46906	6.4226E-1	9	tetrapyrrole binding	46982	6.0305E-2	1	protein heterodimerization activity
16758	1.9396E-1	8	transferase activity, transferring hexosyl groups	8757	6.2643E-2	1	S-adenosylmethionine-dependent methyltransferase activity
16757	2.6908E-1	8	transferase activity, transferring glycosyl groups	4721	7.6552E-2	1	phosphoprotein phosphatase activity
16705	4.0089E-1	8	oxidoreductase activity, acting on paired donors, with iron ion as acceptor	3993	8.3432E-2	1	acid phosphatase activity
5506	4.4581E-1	8	iron ion binding	4872	9.7044E-2	1	receptor activity

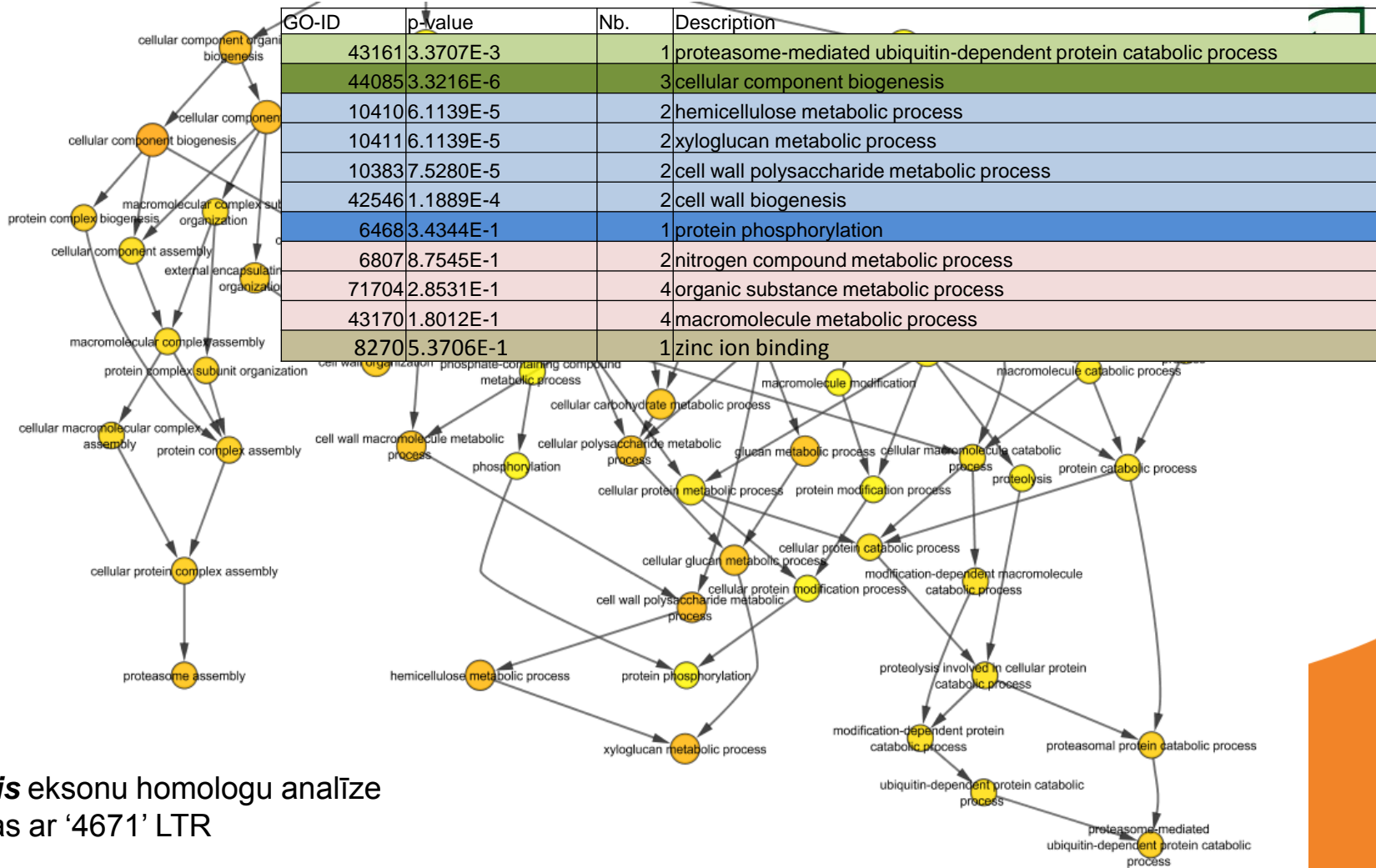
PITA2-bioP				PILA-bioP			
GO-ID	p-value	Nb	Description	GO-ID	p-value	Nb	Description
8152	9.3397E-1	184	metabolic process	8152	2.6606E-1	34	metabolic process
9987	3.2019E-1	170	cellular process	9987	3.4526E-2	33	cellular process
71704	8.3275E-1	155	organic substance metabolic process	44237	1.5174E-2	32	cellular metabolic process
44237	7.5559E-1	149	cellular metabolic process	44238	6.0954E-2	31	primary metabolic process
44238	9.1673E-1	148	primary metabolic process	71704	9.5266E-2	31	organic substance metabolic process
6807	8.0210E-1	138	nitrogen compound metabolic process	6807	2.7511E-2	30	nitrogen compound metabolic process
43170	8.6761E-1	136	macromolecule metabolic process	44260	3.2879E-2	29	cellular macromolecule metabolic process
44260	9.1149E-1	129	cellular macromolecule metabolic process	43170	6.0757E-2	29	macromolecule metabolic process
6725	9.2157E-1	104	cellular aromatic compound metabolic process	6725	1.7315E-2	26	cellular aromatic compound metabolic process
1901360	9.5816E-1	102	organic cyclic compound metabolic process	1901360	1.8057E-2	26	organic cyclic compound metabolic process
46483	9.5649E-1	100	heterocycle metabolic process	90304	2.1275E-2	25	nucleic acid metabolic process
34641	9.6771E-1	100	cellular nitrogen compound metabolic process	6139	2.7306E-2	25	nucleobase-containing compound metabolic process
6139	5.9797E-1	99	nucleobase-containing compound metabolic process	46483	3.0515E-2	25	heterocycle metabolic process
90304	9.6084E-1	97	nucleic acid metabolic process	34641	3.4549E-2	25	cellular nitrogen compound metabolic process
15074	9.8554E-1	77	DNA integration	15074	4.0673E-3	24	DNA integration
1901564	3.9291E-2	48	organonitrogen compound metabolic process	6259	2.4658E-2	24	DNA metabolic process
19538	8.2970E-2	41	protein metabolic process	1901564	7.8109E-1	5	organonitrogen compound metabolic process
44267	7.3284E-2	35	cellular protein metabolic process	6464	6.0410E-1	4	cellular protein modification process
6796	1.5223E-1	33	phosphate-containing compound metabolic process	36211	1.6410E-1	4	protein modification process
6793	1.6082E-1	33	phosphorus metabolic process	43412	6.0773E-1	4	macromolecule modification
6464	1.0198E-1	31	cellular protein modification process	44267	6.9135E-1	4	cellular protein metabolic process
36211	1.0198E-1	31	protein modification process	6796	6.9624E-1	4	phosphate-containing compound metabolic process
43412	1.0636E-1	31	macromolecule modification	6793	7.0109E-1	4	phosphorus metabolic process
16310	1.1952E-1	31	phosphorylation	19538	8.2890E-1	4	protein metabolic process
55114	3.6730E-1	28	oxidation-reduction process	50794	4.1784E-1	3	regulation of cellular process
6468	0.2823E-1	26	protein phosphorylation	50789	4.3696E-1	3	regulation of biological process
9058	1.1206E-1	24	biosynthetic process	65007	5.6729E-1	3	biological regulation
50896	1.3355E-1	24	response to stimulus	16311	2.2844E-2	2	dephosphorylation
1901576	1.4451E-1	22	organic substance biosynthetic process	18193	2.3799E-2	2	peptidyl-amino acid modification
65007	1.4823E-1	22	biological regulation	6810	7.0350E-1	2	transport
44249	1.4937E-1	21	cellular biosynthetic process	51234	7.0408E-1	2	establishment of localization
51716	5.1279E-2	20	cellular response to stimulus	51179	7.0638E-1	2	localization
50794	1.0403E-1	19	regulation of cellular process	50896	8.3907E-1	2	response to stimulus
50789	1.2769E-1	19	regulation of biological process	6468	9.6111E-1	2	protein phosphorylation
7154	4.3740E-4	17	cell communication	16310	3.9368E-1	2	phosphorylation
6810	2.6469E-1	17	transport	55114	9.4344E-1	2	oxidation-reduction process
51234	2.6612E-1	17	establishment of localization	17006	5.1959E-3	1	protein-tetrapyrrole linkage
51179	2.7189E-1	17	localization	9584	5.1959E-3	1	detection of visible light
23052	7.0685E-4	16	signaling	9581	7.7840E-3	1	detection of abiotic stimulus
7165	7.0685E-4	16	signal transduction	9582	7.7840E-3	1	detection of external stimulus
55085	2.0912E-1	13	transmembrane transport	9583	7.7840E-3	1	detection of light stimulus
9059	5.0267E-1	13	macromolecule biosynthetic process	9585	7.7840E-3	1	red, far-red light phototransduction
16070	5.7140E-2	12	RNA metabolic process	51606	7.7840E-3	1	detection of stimulus

Comparison of gene networks

P. Sylvestris pieejamā transkriptoma analīze

118 transkripti sakrīti ar iepriekš atrastā ‘Appalachian’ RE



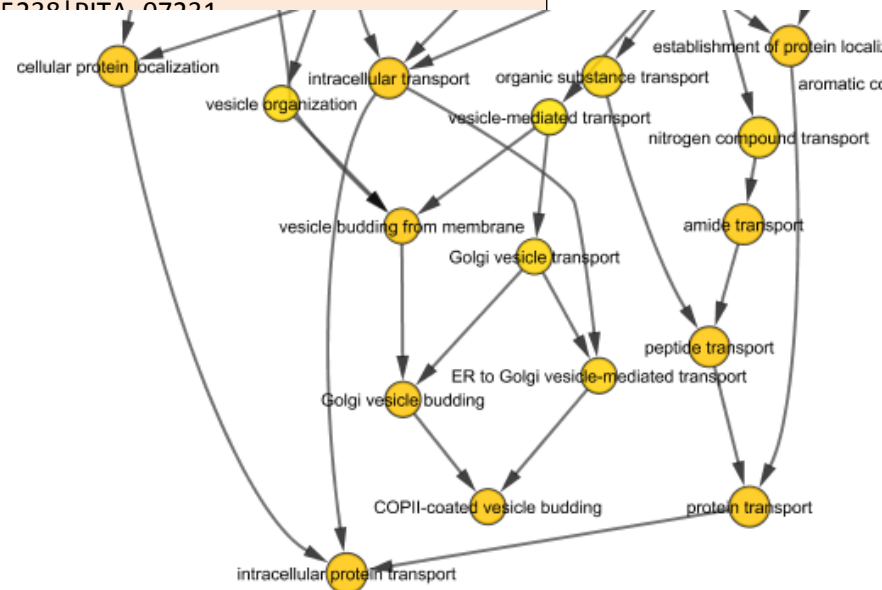


P.Sylvestris eksonu homologo analizē
11 sakritības ar '4671' LTR

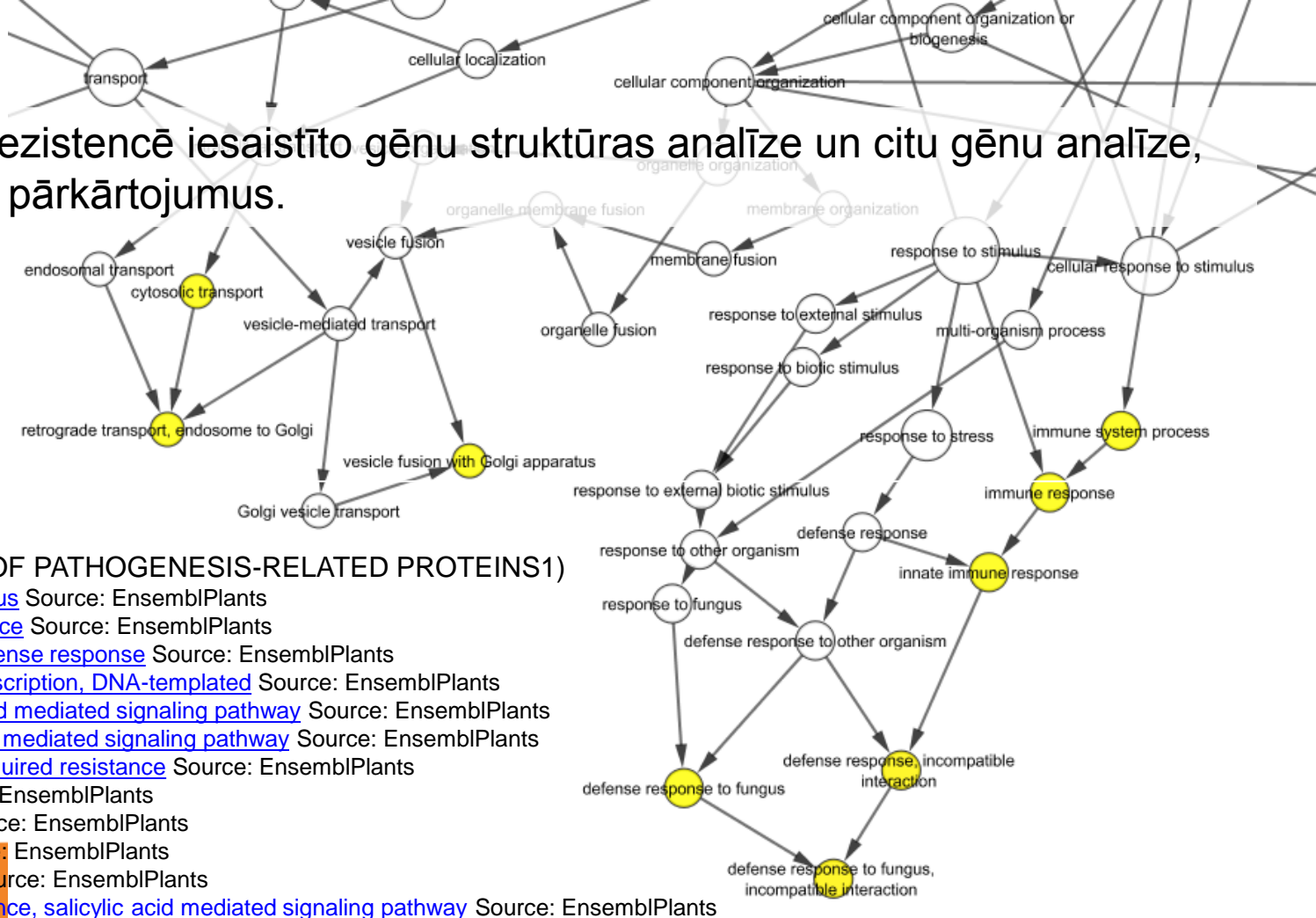
P.Sylvestris eksonu homologu analīze 17 sakritības ar '*PtRLC285*'



GO-ID	p-value	Nb	Description	Genes in test set
15833	1.0243E-3	2	peptide transport	PITA_00121 PITA_00505
90114	9.8404E-4	1	COPII-coated vesicle budding	PITA_00505
48194	9.8404E-4	1	Golgi vesicle budding	PITA_00505
6900	9.8404E-4	1	vesicle budding from membrane	PITA_00505
6278	2.3641E-2	2	RNA-dependent DNA biosynthetic process	PITA_05228 PITA_07224
90304	4.9446E-1	4	nucleic acid metabolic process	PITA_0
90305	2.0163E-2	2	nucleic acid phosphodiester bond hydrolysis	PITA_0
4672	6.1019E-1	1	protein kinase activity	PITA_0
16773	6.1841E-1	1	phosphotransferase activity, alcohol group as acceptor	PITA_0
6468	5.2115E-1	1	protein phosphorylation	PITA_0



*Nozīmīgu augu rezistencē iesaistīto gēnu struktūras analīze un citu gēnu analīze, kas satur līdzīgus pārkārtojumus.



NPR-1 gēns

(NONEXPRESSOR OF PATHOGENESIS-RELATED PROTEINS1)

[defense response to fungus](#) Source: EnsemblPlants

[induced systemic resistance](#) Source: EnsemblPlants

[negative regulation of defense response](#) Source: EnsemblPlants

[positive regulation of transcription, DNA-templated](#) Source: EnsemblPlants

[regulation of jasmonic acid mediated signaling pathway](#) Source: EnsemblPlants

[regulation of salicylic acid mediated signaling pathway](#) Source: EnsemblPlants

[regulation of systemic acquired resistance](#) Source: EnsemblPlants

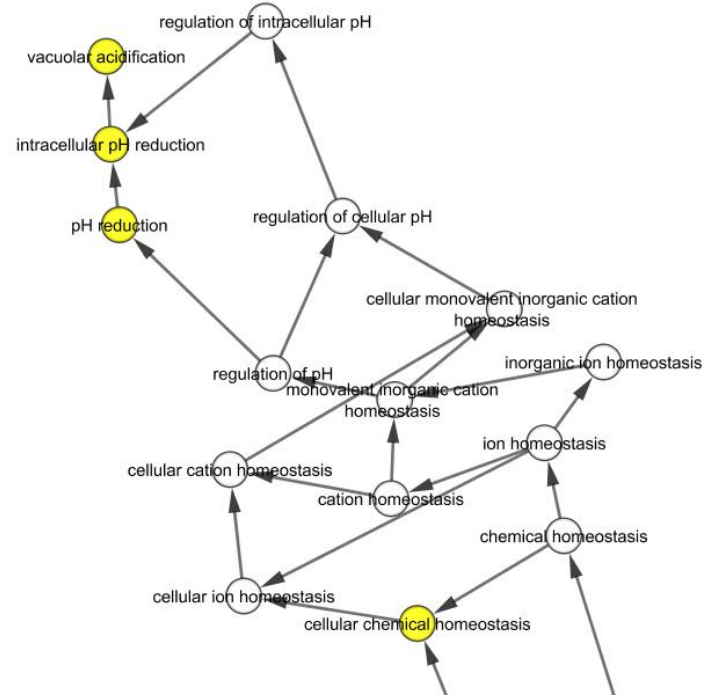
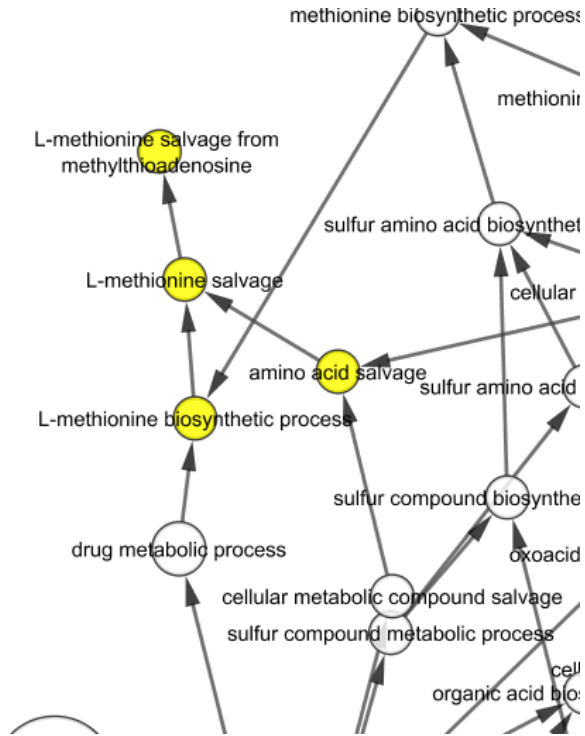
[response to heat](#) Source: EnsemblPlants

[response to hypoxia](#) Source: EnsemblPlants

[response to insect](#) Source: EnsemblPlants

[response to wounding](#) Source: EnsemblPlants

[systemic acquired resistance, salicylic acid mediated signaling pathway](#) Source: EnsemblPlants



Genes significantly overrepresented in the network ($p\text{-value} \leq 0.05$) are indicated



Secinājumi

- *P.taeda* un *P.lambertiana* references genoma sekvenca nesniedz vienoziņīgu informāciju par pilna izmēra transponējamo elementu esamību gēnu rajonos šo sekvenču atkārtotamības un homoloģijas dēļ.
- Gēnos biežāk sastopamo TE ģimeņu profils starp abām sugām ir daļēji līdzīgs pēc sastāva, taču izplatības proporcijas variē.
- Lielāko identificēto gēnu tīklu bioinformātiskā analīze, kura pamatojās uz homologo gēnu salīdzināšanu starp *P.taeda* un *P.lambertiana*, neatrada konservatīvus gēnus ar vienādu TE sastāvu. Taču tīkla dalībnieki tiek kategorizēti un iespējams ņem dalību līdzīgajos stresa inducētos metabolītiskajos ceļos.
- Tā kā *P.sylvestris* genoma labas kvalitātes sekvenca nav pieejama, tiks identificēti gēnu tīkli, kas ir biežāk atrodami abām sugām un izvēti kandidātģēni turpmākiem pētījumiem.



I E G U L D Ī J U M S T A V Ā N Ā K O T N Ē

Paldies!