

EVOLUTION ANALYSIS OF GENES ASSOCIATED WITH APOMIXIS IN BRASSICACEAE FAMILY.

researcher: Rostislav Skitchenko

supervisor: Mike Raiko, CAB SPbU

Research base and project leader

 Theodosius Dobrzhansky Center for Genome Bioinformatics

 Vladimir Brukhin, PhD, Leading Scientist

Goals

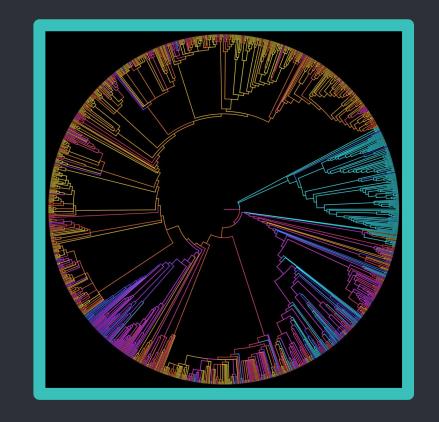
 Perform a comparative phylogenetic assay of the genomes of seven plants

 Find the patterns between specific genes and apomixis plant-forms

Tasks

 Find orthologous genes in other representatives of the Brassicaceae family

 Build the trees of genes of interest



Input data

 7 genomes and proteomes of representatives of the Brassicaceae family

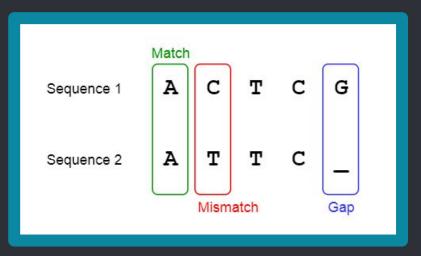
• The set of the 28 proteins of interest

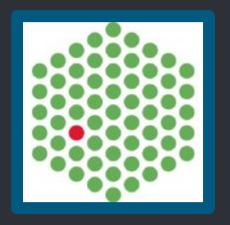
Pairwise local alignment

• Blast

Exonerate

- --model protein2genome
- --showtargetgff
- --percent 90





Exonerate output parsing

- Extract gff annotation
- Parse it to format<chr>< <x><y><prot>
- Extract genes sequences by obtained file
 - \$ bedtools getfasta -name

Preparation to multi-alignment

- Regroup obtained files by genes
- Filter files by
 sequence length
 [x 100 nb . . . x + 1500 nb]
- In case of the absence of any species - add it

```
=====AT4G00020.2 BRCA.fa=====
Bret 1
                        5960
Bret 2
                        146148
Bstricta 278 v1 3
                        5944
Bstricta 278 v1 4
                        159634
Crubella 474 v1 5
                        5812
Crubella 474 v1 6
                        140
Crubella 474 v1 7
                        217
Crubella 474 v1 8
                        116
chi v1 9
                        5816
chi v1 10
                        123227
Athaliana 167 TAIR9 11 7466
Athaliana 167 TAIR9 12
                        5690
Athaliana 167 TAIR9 13 126852
Alvrata 384 v1 14
                        5521
Alvrata 384 v1 15
                        142874
Esalsugineum 173 v1 16 5708
Esalsugineum 173 v1 17 5257
Esalsugineum 173 v1 18
                        191
Esalsugineum 173 v1 19
                        344
Esalsugineum 173 v1 20 129823
```

```
=====AT4G00020.2_BRCA.fa=====

Esalsugineum_173_v1_17_5257

Alyrata_384_v1_14 5521

Athaliana_167_TAIR9_12 5690

Crubella_474_v1_5 5812

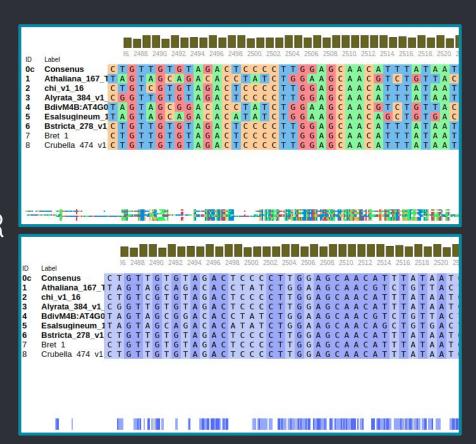
chi_v1_9 5816

Bstricta_278_v1_3 5944

Bret_1 5960
```

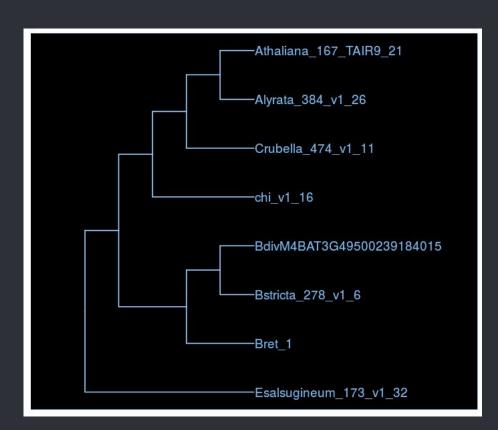
Multiple alignment

- Aligning with MAFFT
- Visualization with R package - msaR
- Cleaning
 alignment by
 conservation
 threshold



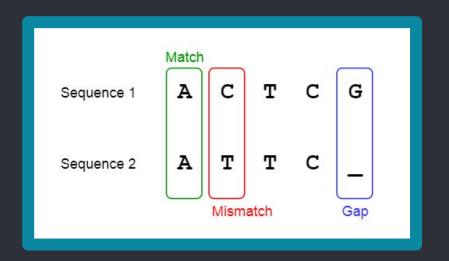
Phylogenetic tree

Building 25
 phylogenetic
 trees for each
 gene



Pairwise local alignment

- Blast
- --makeblastdb
- (7 proteoms)
- --blastp
- --1e*10-4
- Exonerate





EggNog. More orthologues!

get EggNog's id for all genes and all7 plants



from MPEARD... to ENOG410BUQ0

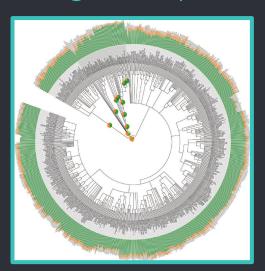
MSSTMF... ENOG410C4KE

... ~(28 x / rows

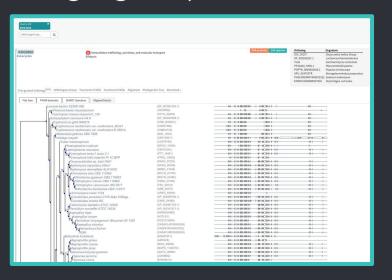
```
PFAM domains
                                 SMART domains
                                                    Aligned blocks
  Flat tree
                                                                                                                                                      PRIMARESETY
                                     Brassica rapa
                                                     (BRA016112)
                                                                             - Bra016112.1-P
                                                                                                   LRRNT 2LRR 1LRRRR 1
                         42 Arabidopsis thaliana
                                                     (SERK1)
                                                                             - AT1G71830.1
                                                                                                                                                      PRIMITIESETVI
                            -Arabidopsis lyrata
                                                                            - scaffold_202099.1
                                                                                                   LRRNT 2LRR 1LRRRR 1
                                                                                                                                                      PRHADESTV
                                                     (SCAFFOLD 202099.1)
                       -Arabidopsis thaliana
                                                                                                                                                   APH PRINCESeTyr
                                                     (SERK2)
                                                                             - AT1G34210.1
0.06
```

EggNog. More orthologues!

- choose groups with duplicates
- search more orthologues for them
- get sequences of orthologs groups







Multiple alignment

- Combine own and orthologs sequences
- Aligning with MAFFT
- --maxiterate 1000 --globalpaii
- Visualization with UGENE
- Cleaning alignment by trimal



--strictplus

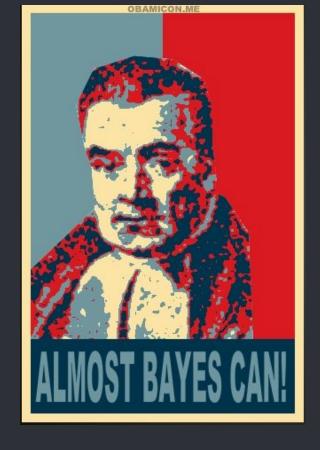
Bayes stat and tree plotting

Mesquite (convert)

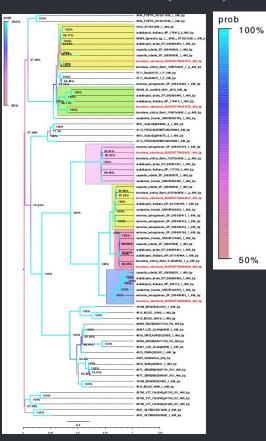
FASTA => NEXUS (mrbayes)

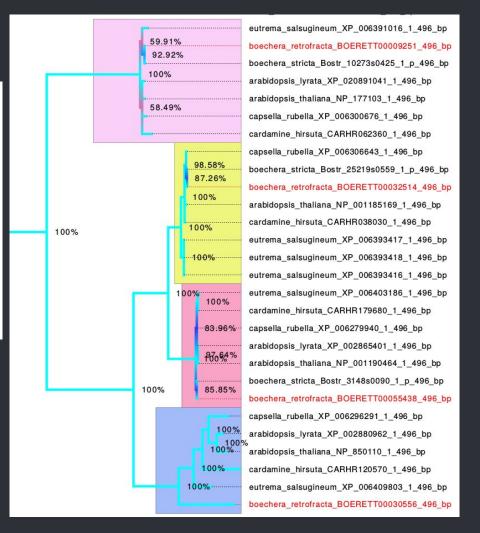
- MrBayes
- --mcmcp ngen = 10 000 000
- --lset nst = 6
- Plot trees with Figtree





ARGONAUTE (AGO)



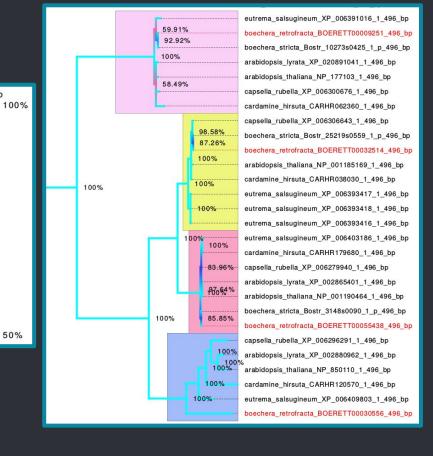


ARGONAUTE (AGO)

Red branch

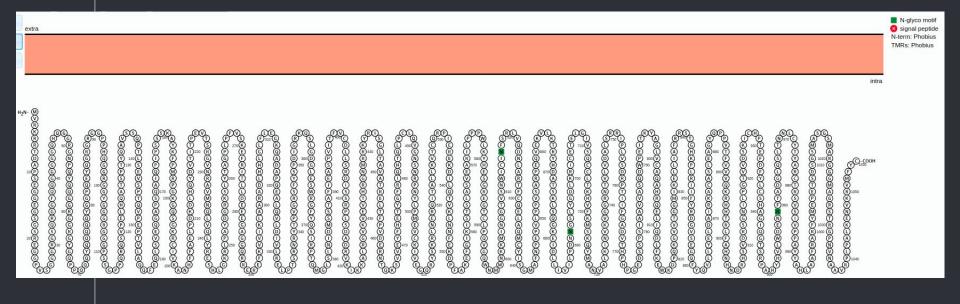


- Yellow branch
- new domains (Gly-rich_Ago1, ArgoMid) (UniProt id O04379)
- new functions, that can be related with apomixis:
 - stem cell development

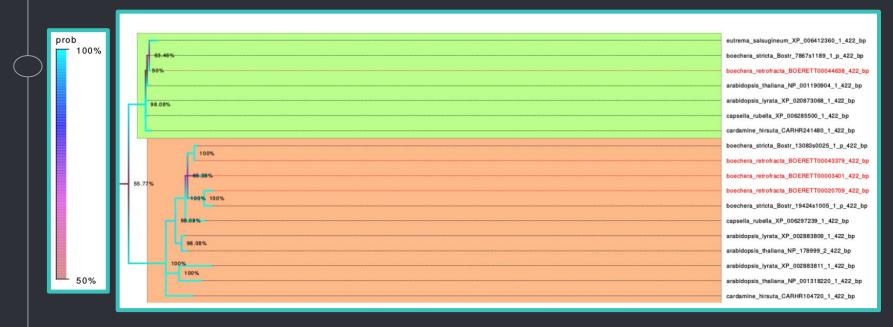


prob

ARGONAUTE (AGO)



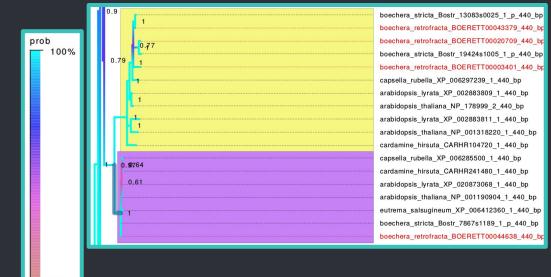
SERK1

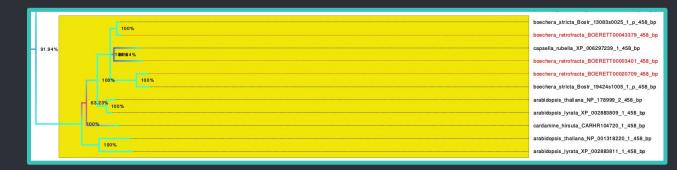


• 3 duplicates in *B.retrofracta*

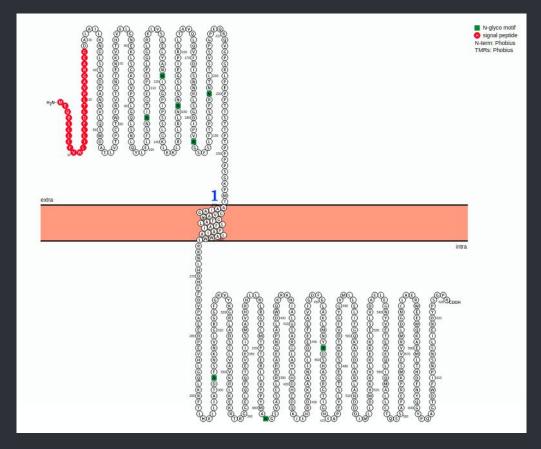
SERK2, SERK3

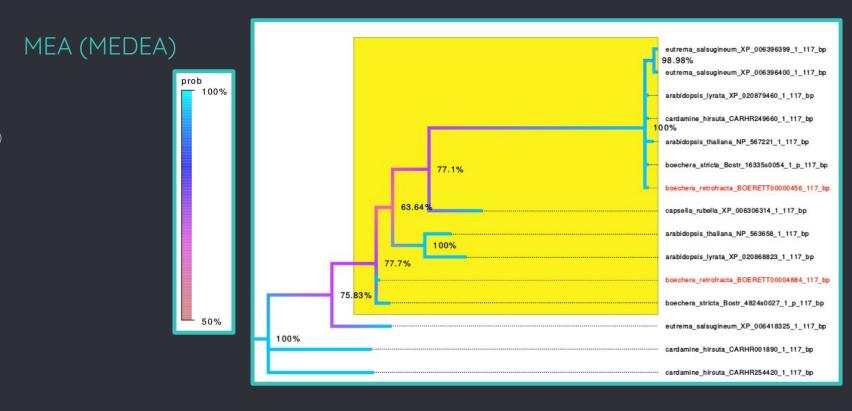
• 3 duplicates in B.retrofracta



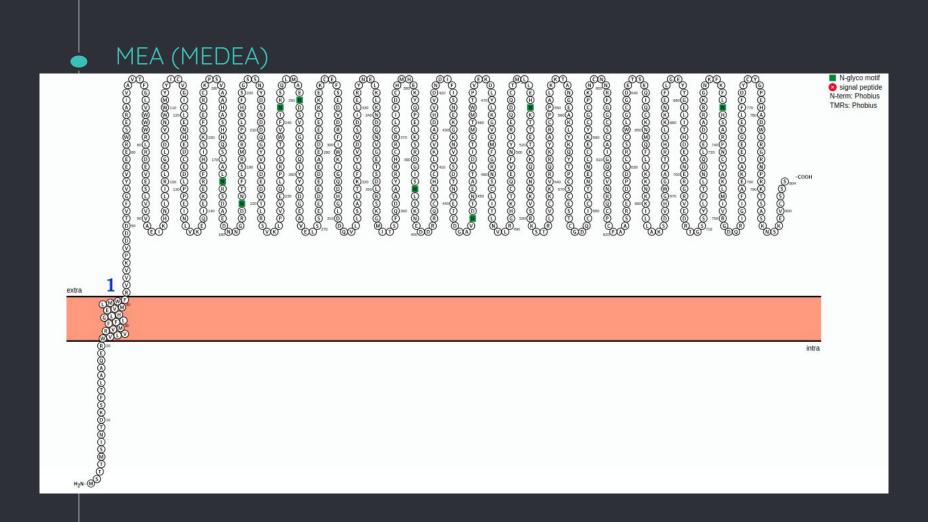


SERK2, SERK3

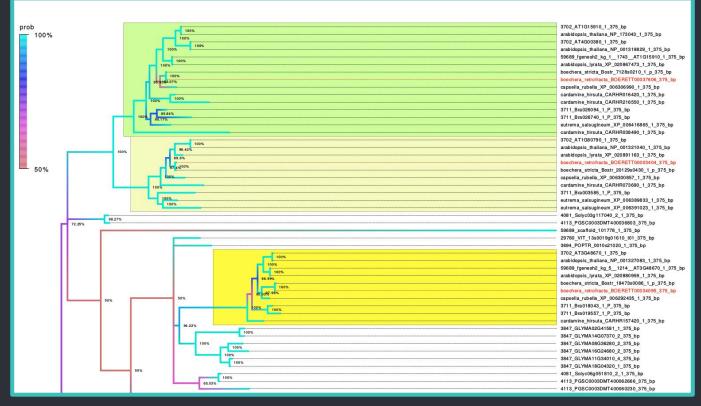




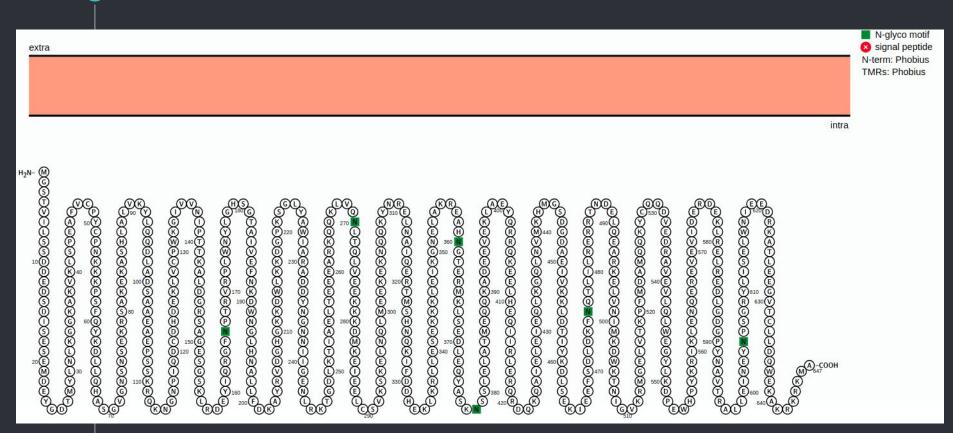
• 2 duplicates - 2 different function



SGS3



• 3 duplicates - 3 different function



DEMETER (DME)

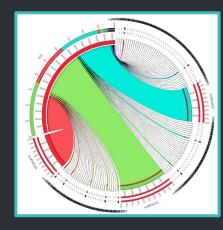
```
3711_Bra005264_1_P_220_bp
                                                                                                                                                                                 cardamine_hirsuta_CARHR130800_1_220_bp
                                                                                                                                                                                 eutrema salsugineum XP 006410826 1 220 bp
                                                                                                                                                                                 capsella_rubella_XP_006295565_1_220_bp
   93.15%
                                                                                                                                                                                 3702_AT2G36490_1_220_bp
                                                                                                                                                                                 arabidopsis thaliana NP 181190 3 220 bp
     96.08%
                                                                                                                                                                                 59689_fgenesh2_kg_4__1689__AT2G36490_1_220_bp
 65.82%
                                                                                                                                                                                 59689_scaffold_402268_1_220_bp
          100%
                                                                                                                                                                                 arabidopsis lyrata XP 020883198 1 220 bp
                                                                                                                                                                                 boechera_retrofracta_BOERETT00053991_220_bp
                                                                                                                                                                                 boechera stricta Bostr 23794s0442 1 p 220 bp
99.27%
                                                                                                                                                                                 capsella rubella XP 006296214 1 220 bp
                                                                                                                                                                                 boechera_stricta_Bostr_23794s0441_1_p_220_bp
                                                                                                                                                                                 boechera retrofracta BOERETT00053989 220 bp
                                                                                                                                                                                 eutrema_salsugineum_XP_006399026_1_220_bp
```

addition duplicate in Boechera

What is planned?

- Hi-C (assemble scaffolds to chromosome)
- Apomictic RNA-seq
 (diff.expr. *B.retrofracta* x
 B.divaricarpa)
- Variant calling





https://github.com/rostkick/Boechera.