

Structural and Evolutionary Features of Active **Sunflower Retrotransposons**

M.Yu. Kazantsev^{1,2*}, P.Yu. Merkulov^{1, 2}, K.N. Tyurin^{1,2}, Y.N. Demurin³, I.V. Kirov^{1,2}

¹FGBNU 'All-Russia Research Institute of Agricultural Biotechnology', Moscow, Russia ²Moscow Institute of Physics and Technology (National Research University), Dolgoprudny, Russia ³All-Russian Research Institute of Oilseed Crops named after Pustovoit, Krasnodar, Russia

Introduction 1

Genome of sunflower is ~75% composed of retrotransposons (RTEs). Most RTEs are inactivated, but active copies are capable of movement, influencing genome evolution and plant adaptation.

- We have activated transposons in the sunflower genome using a combination of epigenetic and thermal stress.
- EccDNA is a byproduct of active RTEs, we have isolated them and sequenced using Oxford Nanopore technology.
- The data obtained, as well as the sequencing results of 12 publicly available sunflower varieties, served to conduct phylogenetic analysis, predict the structure of proteins and study the chromosomal distribution of retrotransposons in the genome of Helianthus annuus L. Special attention was paid to additional open reading frames (aORFs) encoding proteins with unknown functions.

The aim of this work was to study the active LTR retrotransposons of sunflower Helianthus annuus L. and their role in the variability of the genome structure.

2 **Methods**



RTE ACTIVATION:

Plants were treated with a combination of heat stress (37°C) and toxins (zebularin + α -amanitin).



EccDNA DETECTION:

Plasmid-Safe DNAase treatment (PS+) vs. control $(PS-) \rightarrow rolling circle$ amplification (RCA) \rightarrow T7 endonuclease treatment (increased data yield by 4fold).



VALIDATION:

Inverted PCR with specific primers for 9 RTEs.



BIOINFORMATICS:

RTE annotation (DANTE_LTR, HITE), phylogenetic analysis (TESorter), protein structure prediction (AlphaFold 3, Foldseek). Analysis of insertion polymorphism (TIP) in 12 sunflower cultivars (SPLITREADER).

3 **Results**

Active RTEs and their insertional landscape

28 active RTEs have been identified:

- Belonging to the Tekay, Athila, Galadriel, Tork (Ty3/Gypsy) and Ale (Ty1/Copia) clades.
- Three phylogenetic groups of Tekay: Group 2 dominates in activity.



Results

4

Non-canonical aORF proteins

aORF is found in:

- Athila (3'-aORF, transmembrane domain, endoplasmic reticulum).
- Tekay group 2 (5'-aORF, nuclear localisation signals NLS).

Structural homology:

- The Tekay proteins of maize and sorghum are structurally similar to the DRBM protein of Glycine max (double-stranded RNA-binding domain).
- No similar homology was detected in sunflower.



Athila 14 aORF (-)

Tekay 13-2 aORF

- 80% of TIPs are unique to 1-2 varieties, indicating
- Group 1 and 3: Concentrated in pericentromeric regions. Group 2: Insertions in

Tekay 13-2 5'-aORF protein features:

- Highly conserved: 95.6% of 523 Tekay-like elements contain 5'-aORF.
- Structure: Long disordered regions + αhelix at the C-terminus.
- No amino acid similarities, but all 5'-aORF of other analysed species exhibited amino



euchromatin regions (chromosome arms), often close to genes. acid content biases and had predicted nuclear localization



Low (70 > plDDT > 50) Very low (pIDDT < 50)

Confident (90 > pIDDT >70

Very High (pIDDT > 90)

Athila 14 aORF

5 Discussion

Evolutionary role of aORFs:

- 3'-aORFs of Athila encode env-like proteins (possible role in intercellular transport).
- 5'-aORF of Tekay Group 2 are highly conserved and have NLS, that suggest functional significance.

Linkage to silencing:

• Structural similarity of other species Tekay proteins to DRBM proteins suggests a possible role in modulating RNA silencing (similar to viral suppressors of VSRs).

Insertion preferences:

• The different distribution of Tekay groups is not explained by differences in integrase/chromodomain, suggesting a role for aORF proteins in integration site selection.

Perspectives:

Studying the role of aORF-proteins in transposition and interaction with the silencing

Conclusion

6

7

- RTE activity in sunflower is heterogeneous and stress induced.
- Tekay group 2 are the most active RTEs with a unique euchromatin insertion distribution.
- Conserved aORF proteins contain functional domains and are structurally homologous to DRBM proteins in other plants.

Contacts

- miykazantsev@gmail.com R
- @zan_mi \triangleleft
- The study was supported by RSF grant No. 22-64-00076

