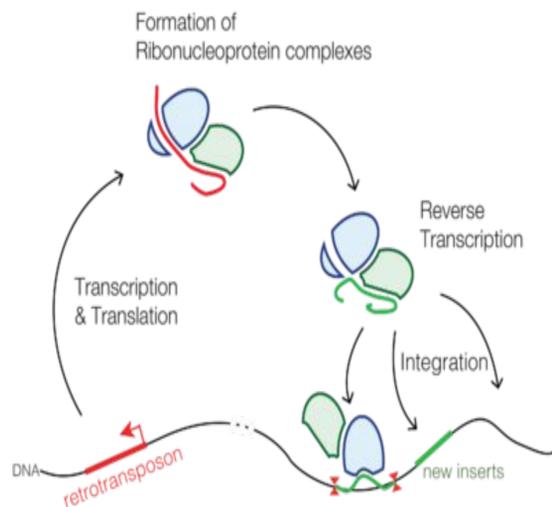


Introduction

Many plant genomes are made up in large part by transposable elements. Of the many transposable element varieties, LTR retrotransposons have become particularly abundant in large part due to their copy and paste replication method.

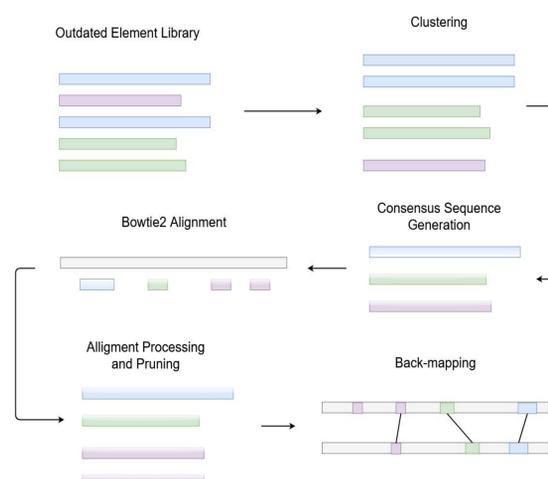


Despite their ubiquity, when a plant reference assembly is updated, the public catalogues of transposons may not receive the same treatment. Such is the case of *Glycine Max* (soybean) which has undergone 4 reference assembly updates since the original 2010 v1.0 release, but its transposable element catalogue is still mapped to the 2010 version.

To address this problem we have created a pipeline utilizing Bowtie 2 global alignments that enable researchers to quickly remap outdated transposable element families from consensus sequences. It is demonstrated here through the remapping of the Gypsy superfamily.

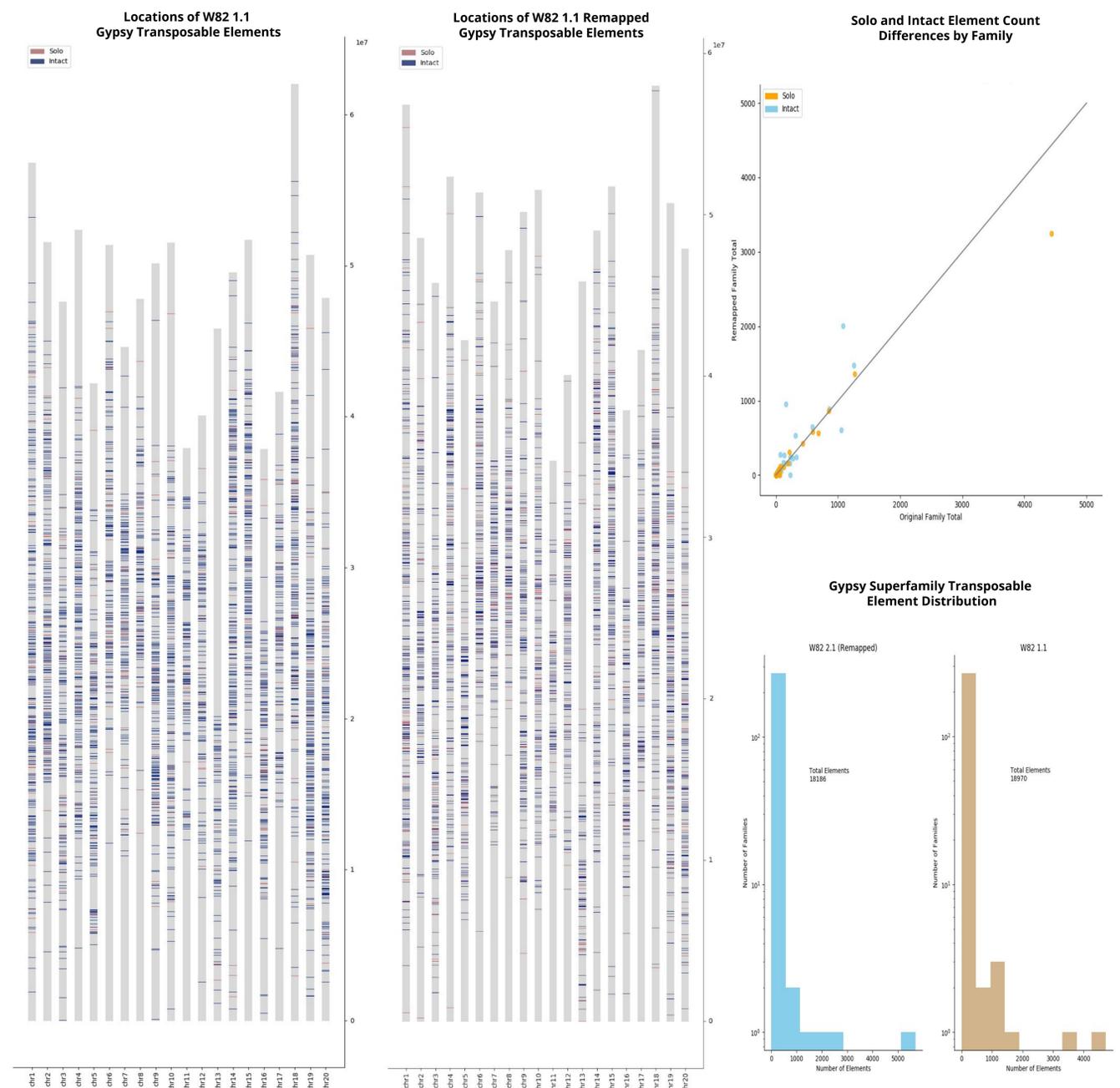
Pipeline Overview

The pipeline is designed to remap transposable elements at the family level through alignment of clustered consensus sequences. This significantly reduces the number of alignments required to remap an entire highly similar family. Additionally, clustered consensus searches have the advantage of being non-specific. Alignments can therefore return novel elements not included in the outdated assembly. A generalization of the pipeline is shown below.



Results

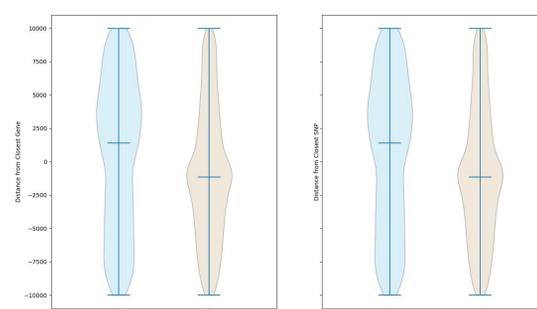
Locations of 277 members of the Gypsy superfamily of transposable elements in *Glycine Max*. On the left, locations per the Soybase transposable element libraries. On the right remapped locations using the TARP pipeline. The histogram (below right) compares the distribution of element count by family for each element library. The scatter plot shows the variation between the number of elements in each library at the family level. Points closer to the line indicate families with more similar inter-assembly element content.



Discussion

Remapping of the Gypsy superfamily using the TARP pipeline revealed significant differences between the current Soybase transposable element library, and the content of the most current reference assembly.

Distribution of Distances to Nearby Genomic Features



Differences between the remapped and Soybase libraries included variation in transposable element content and type, deletion/insertion of elements, and changes in average proximity to genomic features such as genes and SNPs. It is likely similar discrepancies exist across other public transposable element libraries, which presents many opportunities for TARP remapping to provide updated genomic information.

Acknowledgements

1. Camacho, Cistiam, *et al.* "BLAST+: Architecture and Applications." *BLAST+: Architecture and Applications*, 15 Dec. 2009, doi: <https://doi.org/10.1186/1471-2105-10-421>.
2. Langmead, Ben, and Steven L. Salzberg. "Fast Gapped-Read Alignment with Bowtie 2." *Nature Methods*, vol. 9, no. 4, 2012, pp. 357-359, doi: 10.1038/nmeth.1923.
3. "SoyBase.org." *SoyBase.org*, soybase.org/.

This project is supported through the Mulcahy Scholars program of Loyola University Chicago.

Access the quick start guide and project code on GitHub

