Analysis of Transposable Elements in Grasses: Preliminary comparison of pipelines for creating de-novo libraries

Maria Camila Diaz-Rodriguez*; Diego Mauricio Riaño-Pachón*

* Computational, Evolutionary and Systems Biology Laboratory (LabBCES), Agroindustrial Productivity Division, Center for Nuclear Energy in Agriculture (CENA), University of São Paulo (USP), Piracicaba, São Paulo, Brazil

Introduction

As the industry faces the challenge of improving crop resilience and productivity, breeding programs are increasingly turning to molecular approaches. A crucial task is to accurately identify the genetic mechanisms that influence and regulate phenotypes of commercial interest Recently, Transposable Elements (TEs) have been studied as a possible source of variability in evolution and genetic function by their ability to move within the genome and duplicate themselves, transforming them into the most abundant genetic material in eukaryotes.

In sugarcane, the study of TEs has been very limited by the available sequencing technologies, but given the progressive development of high throughput DNA sequencing methodology (next generation sequencing; NGS) in recent years, complete or partially complete genomes are available today, providing a greater amount of information.

Objective

Based on the information available in databases to date, and waiting to find sufficient divergences and robust data to establish relationships between species and genotypes to be studied, will identify the number and types of TEs present in sugarcane (*Saccharum* hybrid cultivar CC_01_1940) and phylogenetically closely related species such as rice (*Oryza sativa*, cultivar Nipponbare), sorghum (*Sorghum bicolor*, cultivar BTx623) and corn (*Zea mays*, cultivar B73).

Methods

Genomes — National Center for Biotechnology Information (NCBI)
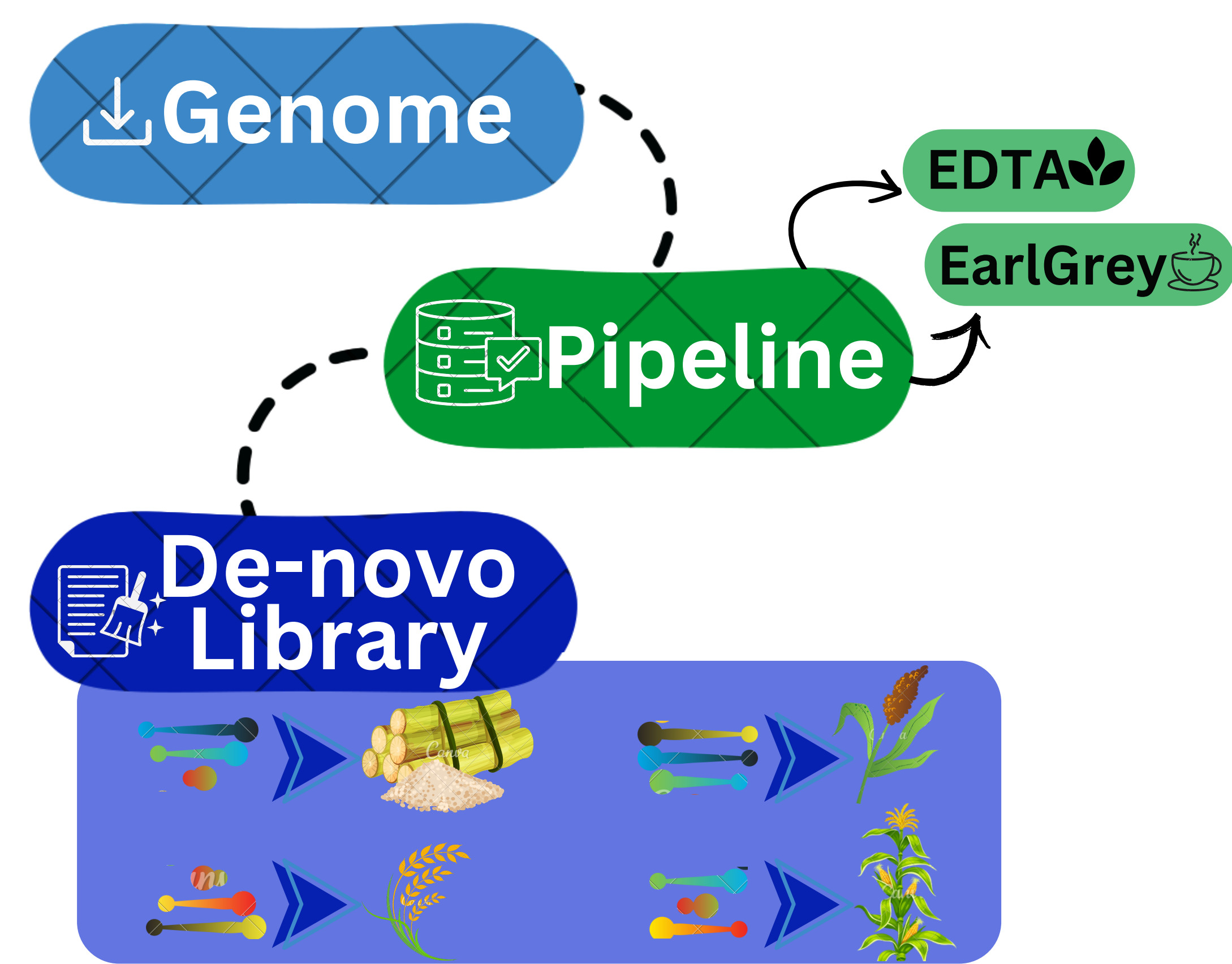- Sugarcane (*Saccharum* hybrid cultivar CC_01_1940)
- Rice (*Oryza sativa*, cultivar Nipponbare)
- Sorghum (*Sorghum bicolor*, cultivar BTx623)
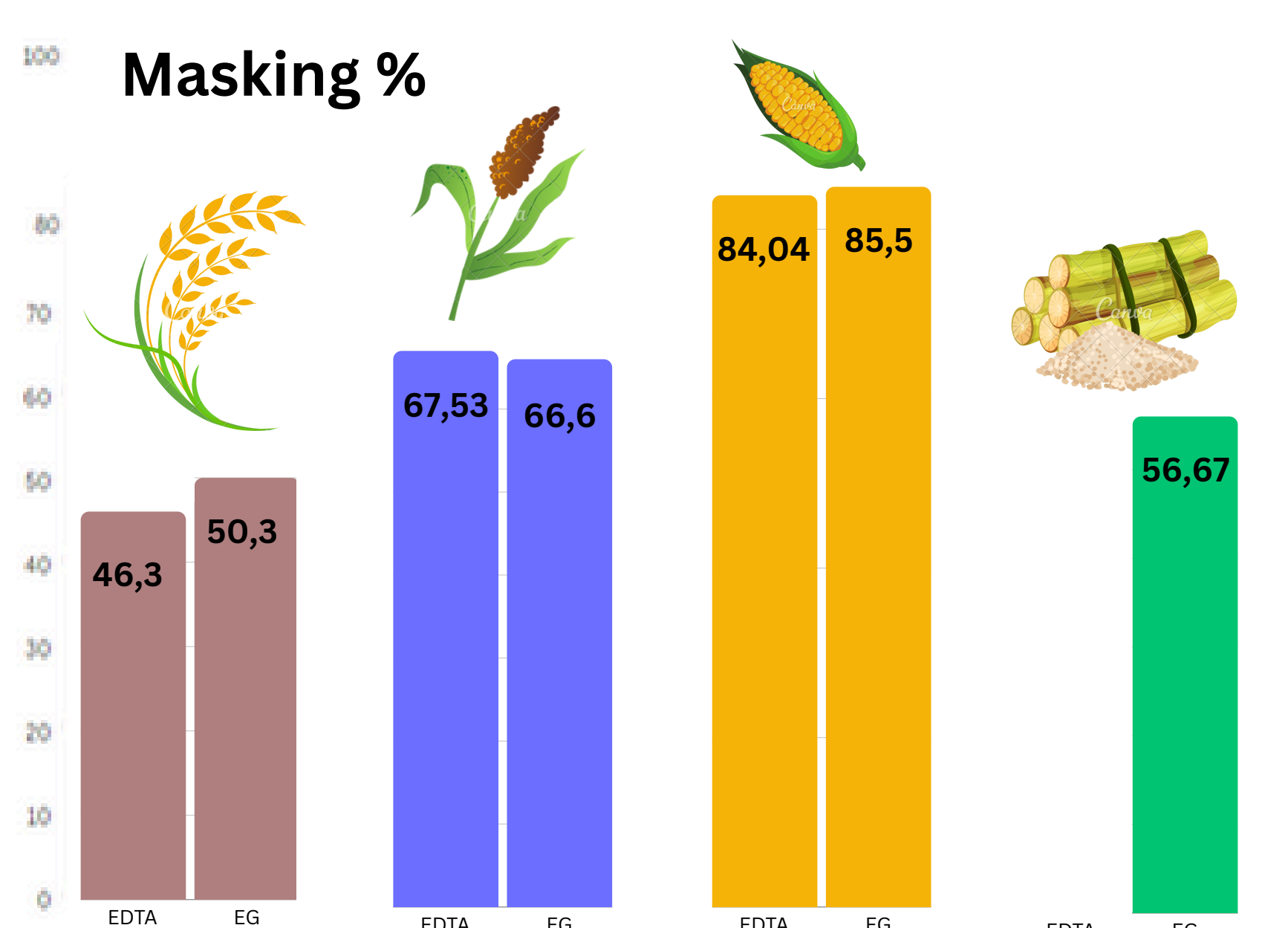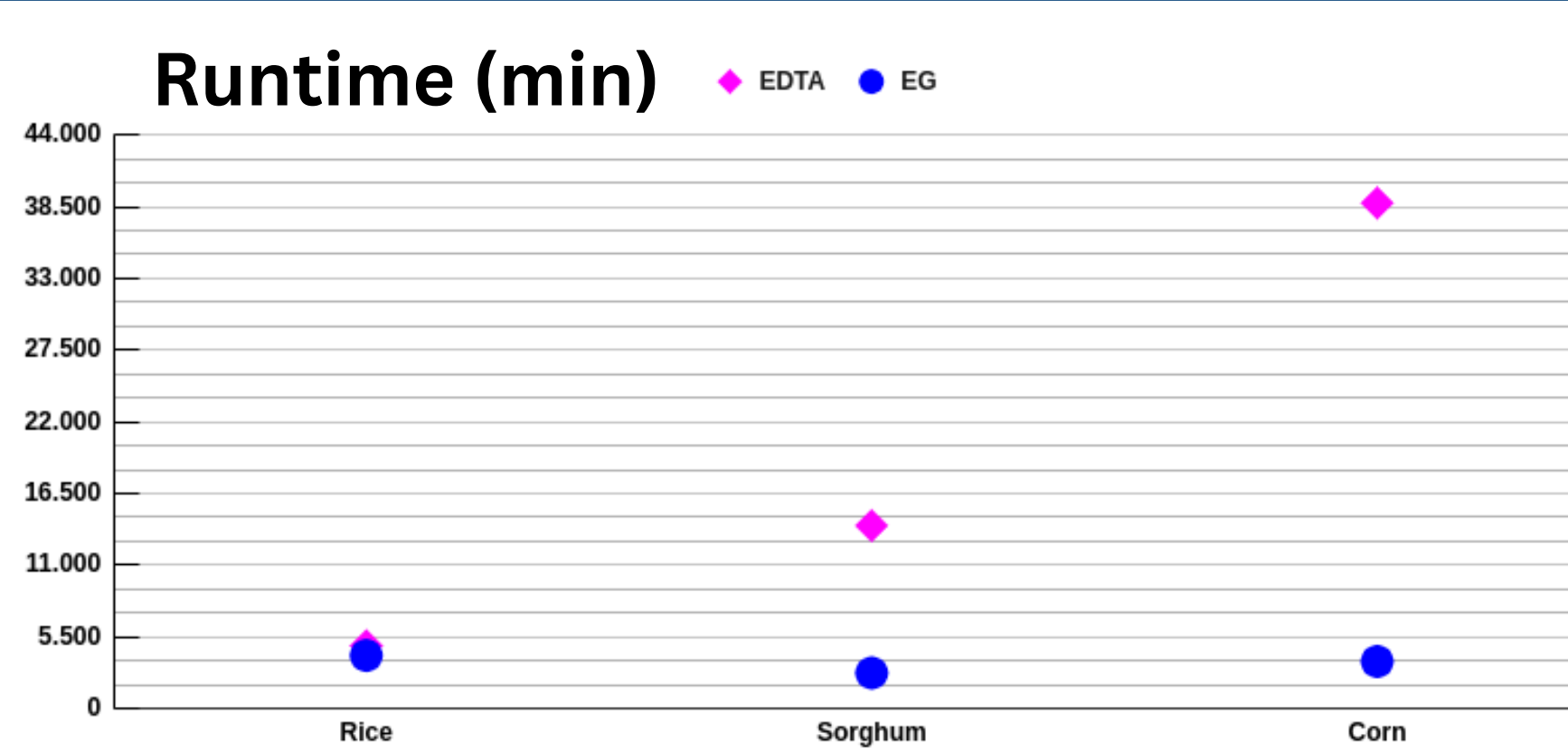- Corn (*Zea mays*, cultivar B73)

Process evaluation
- Runtime
- Masking percentage
- Depth classification level
- TEs in relation to the genome
- Diversity of TEs found

Pipelines (The Extensive de novo TE Annotator (EDTA) / Earl Grey (EG)):
- Search/align algorithms: LTRharvest, LTR_FINDER, rcMergeRepeats, Others
- Repeat element identification/masking: RepeatMasker (EDTA ✓ / EG ✓ >=v 4.1.4), RepeatModeler2 (EDTA ✓ / EG ✓ >=v 2.0.4)
- Curated libraries of repeats: Dfam (EDTA ✓ / EG ✓ >=v 3.7), Repbase
- Instalation: Conda, Singularity, Docker

Genome → Pipeline → De-novo Library → EDTA / EarlGrey

Results

Runtime (min)

Masking %

- Rice: EDTA 46,3 / EG 50,3
- Sorghum: EDTA 67,53 / EG 66,6
- Corn: EDTA 84,04 / EG 85,5
- Sugarcane: EG 56,67

Diversity of TEs (number of different families identified)

	EDTA				Earl Grey			
	Rice	Sorghum	Corn	Cane	Rice	Sorghum	Corn	Cane
DNA	6072	4624	7386	NA	105	133	123	183
LINE	15	20	11	NA	70	76	30	159
LTR	4377	5532	4719	NA	225	365	375	542
Other (SR,Mi,RNA)	0	0	0	NA	6576	7140	8500	9109
Rolling Circle	0	0	0	NA	12	4	6	14
MITE	2982	4079	5180	NA	0	0	0	0
Unclassified	178	286	1011	NA	1499	905	676	1702
TOTAL	13624	14541	18307	-	8487	8623	9710	11709

TEs in relation to the genomes

Depth classification level

Class → Order → Superfamily → Family
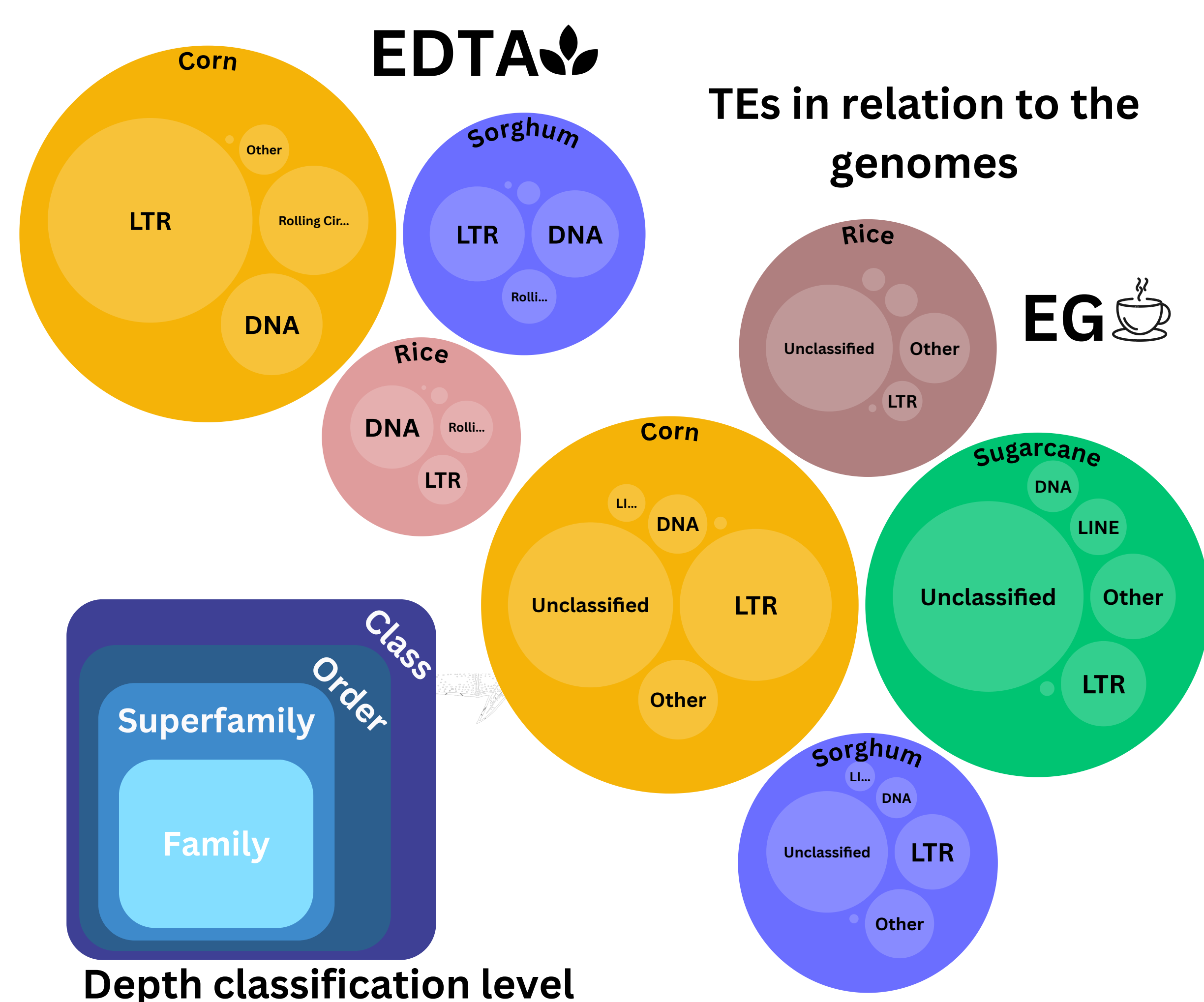
Conclusion

There was a large proportion of unclassified TEs for both Earl Grey and EDTA. Despite this, it was possible to obtain a detailed view of the composition and diversity of TEs in the selected genomes, highlighting the prevalence of LTR and DNA elements, and its difference in diversity. Regarding runtime and variety of information provided, the use of EarlGrey is recommended.

Grants and thanks

cena USP · CAPES · FEALQ · CNPq

References

Baril, T., Galbraith, J.G., and Hayward, A., Earl Grey: A Fully Automated User-Friendly Transposable Element Annotation and Analysis Pipeline, Molecular Biology and Evolution, Volume 41, Issue 4, April 2024, msae068 doi:10.1093/molbev/msae068

Ou S., Su W., Liao Y., Chougule K., Agda J. R. A., Hellinga A. J., Lugo C. S. B., Elliott T. A., Ware D., Peterson T., Jiang N., Hirsch C. N. and Hufford M. B. (2019). Benchmarking Transposable Element Annotation Methods for Creation of a Streamlined, Comprehensive Pipeline. Genome Biol. 20(1): 275.