# Comparative Analysis of Splicing features in Plants and Other Eukaryotes

**Beatriz Rodrigues Estevam [1] ; Dirk Walther [2] ; Diego Mauricio Riaño-Pachón [1]**

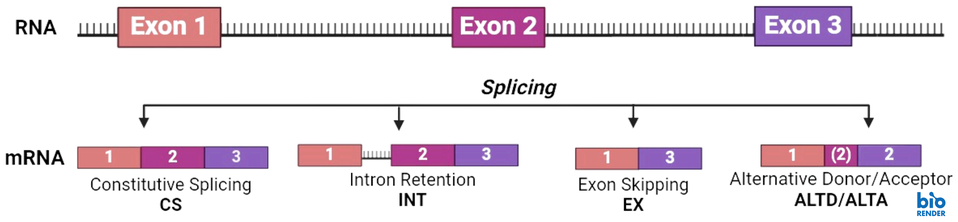
1 Computational, Evolutionary and Systems Biology Laboratory - LabBCES - CENA/USP, Brazil

2 Max Planck Institute of Molecular Plant Physiology - MPIMP, Germany

(bia.estevam.25@usp.br; walther@mpimp-golm.mpg.de; diego.riano@cena.usp.br; )

## INTRODUCTION

Splicing, a crucial process in pre-mRNA maturation, plays a pivotal role in gene expression regulation and proteome diversification across eukaryotic organisms. Despite its significance, the regulatory mechanisms of splicing remain poorly understood, especially in plants, relying heavily on human-centric data.




**In light of this, our ongoing project aims to address this knowledge gap by investigating the differences in splicing features between plant species and other eukaryotes.**

KEY OBJECTIVES

1. Determine the prevalence of splicing types between human and plants;
2. Examine the frequency of dinucleotides present at splicing sites;
3. Analyze the stability of RNA secondary structures (SS) in splice sites flanking sequences.

## METHODS

**SPECIES SELECTION**

- **Human** HG18 and HG38
- ***Arabidopsis thaliana*** TAIR10
- ***Zea mays*** **Maize** "2017_v31"

- Alternative splicing data is publicly available;
- Represent animals and plants (monocots and eudicots).

**DATA ACQUISITION**

**Splicing databases**
- UCSC genome browser
- PastDB
- PASD

**Acquisition of Spicing Events Coordinates, Genomes, and Annotation files.**

**SPLICING TYPE DEFINITION** — KEY OBJECTIVE 1

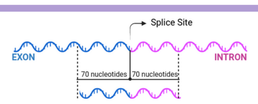
selectConstitutiveExons.py

**Alternative splicing**
Obtained from splicing databases

**Constitutive splicing**
Obtained from annotations.
1. The exon appears in all isoforms.
2. The exon maintains the same positions in all isoforms.
3. The gene has at least three exons.
4. The gene contains at least four isoforms.

**SEQUENCE EXTRACTION**

**Extraction of sequences around the splice sites**
To perform following analysis.




extractor_splicing.py

**SPLICE SITE USAGE** — KEY OBJECTIVE 2

**Sequence logo dinucleotide analysis**
To identify conservation and variations of splice site dinucleotides considering splicing events.

CoCoView

splice sites extraction

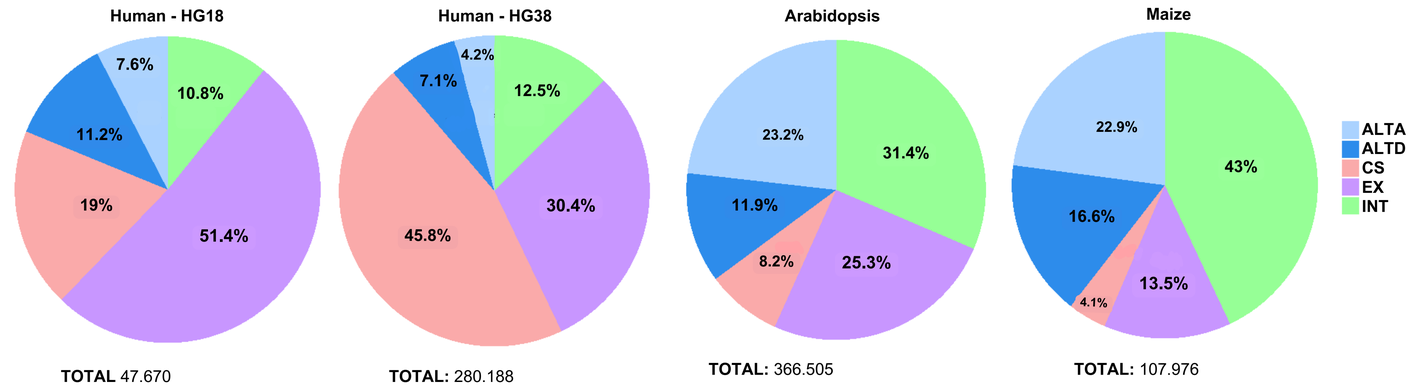
**SECONDARY STRUCTURE** — KEY OBJECTIVE 3

**RNA secondary structure stability for each splicing type.**
To expand the hypothesis made previously for humans in plants.

RNAfold

plots_MFE_analysis.R

## RESULTS




KEY OBJECTIVE 1

**Fig 1. Splicing events type per species.** Concerning alternative splicing, HG18 and HG38 have similar fractions of each type, with exon skipping as the most common alternative splicing type. In contrast, Intron retention is the most common alternative splicing type for both plant species.

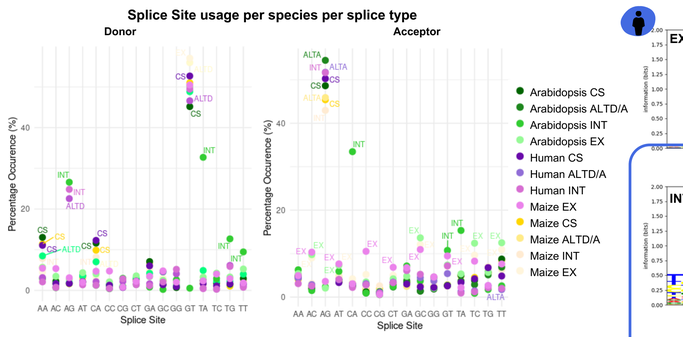
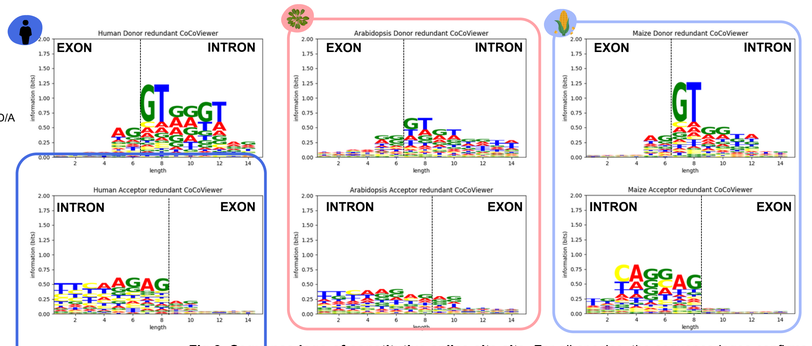
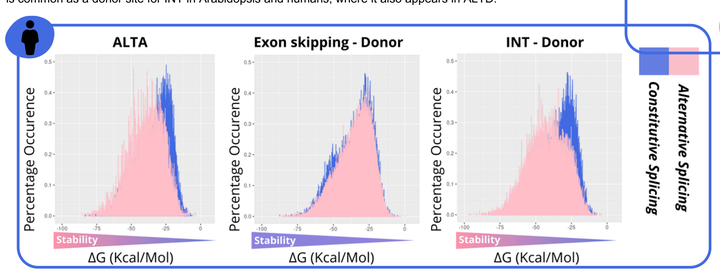



**Fig 2. Percentage occurrence of dinucleotides at splice sites.** Across all species, the canonical GT...AG site has the highest global frequency. However, other dinucleotides are also present, such as "AA" for the Arabidopsis constitutive donor site and "CA" for the human constitutive donor site. Another notable observation is the high occurrence of "TA" and "CA" at the Arabidopsis INT donor and acceptor sites, respectively. Additionally, the "AG" site is common as a donor site for INT in Arabidopsis and humans, where it also appears in ALTD.




KEY OBJECTIVE 2

**Fig 3. Sequence logo of constitutive splice site site.** For all species, the sequence logos confirms that the canonical site GT...AG has a major global frequency. However, attention should be paid to the slight displacement of splice sites, which may indicate potential issues with the annotation and definition of splicing. Dashed lines denote the boundary between exons and introns.



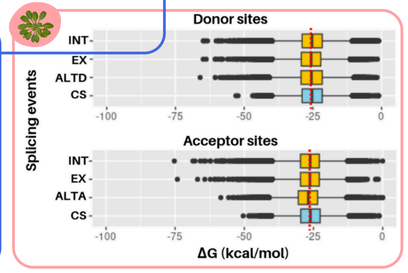




**Fig 5. Comparison of Secondary Structure Stability Between Constitutive and Alternative Splicing Events for *A. thaliana*.** No significant difference in stability dispersion was observed between constitutive and alternative splicing in *A. thaliana.*

KEY OBJECTIVE 3

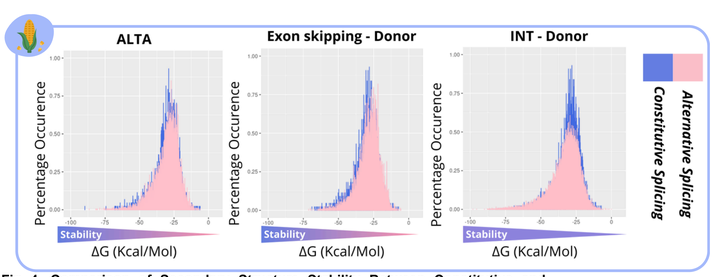



**Fig 4. Comparison of Secondary Structure Stability Between Constitutive and Alternative Splicing Events.** Comparing SS stability of the same events between maize and human shows different patterns, including opposite patterns.

KEY OBJECTIVE 3

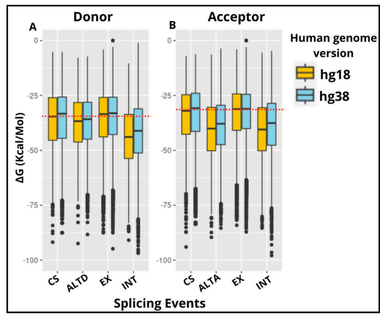
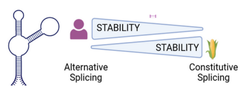



**Fig 6. Box plot of secondary structure stability for human genome versions (hg18 and hg38) to assess variations between the versions.** Signals reported for HG38 are highly similar to previous results in HG18 .




## CONCLUSIONS AND NEXT STEPS

Our comparative analysis reveals significant differences in splicing mechanisms between plants and humans, and even among monocots and eudicots, mainly:

1. Differences in the proportion of splicing types;
2. Variations in dinucleotide usage at splice sites across splicing types (though it raises a question about the definition of these events considering the very slight shift in splice sites in the annotations); e
3. Patterns of secondary structure stability that are different and opposite between humans, monocots, and eudicots.

### NEXT STEPS

- To improve the knowledge on splicing and splicing regulation in plants, given the lack of splicing data in plants and promising results, our next goal is to systematically identify alternative splicing in grasses such as sorghum, maize, rice, barley, and *Arabidopsis thaliana* to perform analysis comprehanding splicing regulation using principally bioinformatics techniques.

@labbces

https://labbces.cena.usp.br/