

Splicing conservation signals in plant long non-coding RNAs

Jose Antonio Corona-Gomez^a, Irving Jair Garcia-Lopez^a, Peter F. Stadler^{c,d,e,f,h,*}, Selene L. Fernandez-Valverde^{b,*}

^aUnidad de Genómica Avanzada, Langebio, Cinvestav, Km 9.6 Libramiento Norte Carretera León, 36821 Irapuato, Guanajuato, México

^bCONACYT, Unidad de Genómica Avanzada, Langebio, Cinvestav, Km 9.6 Libramiento Norte Carretera León, 36821 Irapuato, Guanajuato, México

^cBioinformatics Group, Department of Computer Science, University Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany

^dInterdisciplinary Center for Bioinformatics, University Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany

^eMax Planck Institute for Mathematics in the Sciences, Inselstraße 22, D-04103 Leipzig, Germany

^fDepartment of Theoretical Chemistry, University of Vienna, Währingerstraße 17, A-1090 Wien, Austria

^gFacultad de Ciencias, Universidad Nacional de Colombia, Sede Bogotá, Colombia

^hSanta Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA

SUPPLEMENTAL MATERIAL

Table S1: **Genome versions used in this study.**

Species	Ensembl ID	Source
<i>Aethionema arabicum</i>	GCA_000411095.1	NCBI
<i>Arabidopsis halleri</i>	GCA_900078215.1	Phytosome v12
<i>Arabidopsis lyrata</i>	GCA_000004255.1	Ensembl-Plants
<i>Arabidopsis thaliana</i>	GCA_000001735.1 TAIR10	Ensembl-Plants
<i>Arabis alpina</i>	GCA_000733195.1	NCBI
<i>Boechera stricta</i>	GCA_002079875.1	NCBI
<i>Brassica napus</i>	GCA_000686985.1	Ensembl-Plants
<i>Brassica rapa</i>	GCA_000309985.1	Ensembl-Plants
<i>Camelina sativa</i>	GCA_000633955.1	NCBI
<i>Capsella rubella</i>	GCA_000375325.1	Phytosome v12
<i>Leavenworthia alabamica</i>	GCA_000411055.1	NCBI
<i>Raphanus sativus</i>	GCA_000801105.2	NCBI
<i>Schrenkiella parvula</i>	GCA_000218505.1	NCBI
<i>Sisymbrium irio</i>	GCA_000411075.1	NCBI
<i>Brassica oleracea</i>	GCA_000695525.1	Ensembl-Plants
<i>Eutrema salsugineum</i>	GCA_000478725.1	NCBI

*Corresponding authors

Email addresses: jose.corona@cinvestav.mx (Jose Antonio Corona-Gomez), irving.garcia@cinvestav.mx (Irving Jair Garcia-Lopez), studla@bioinf.uni-leipzig.de (Peter F. Stadler), selene.fernandez@cinvestav.mx (Selene L. Fernandez-Valverde)

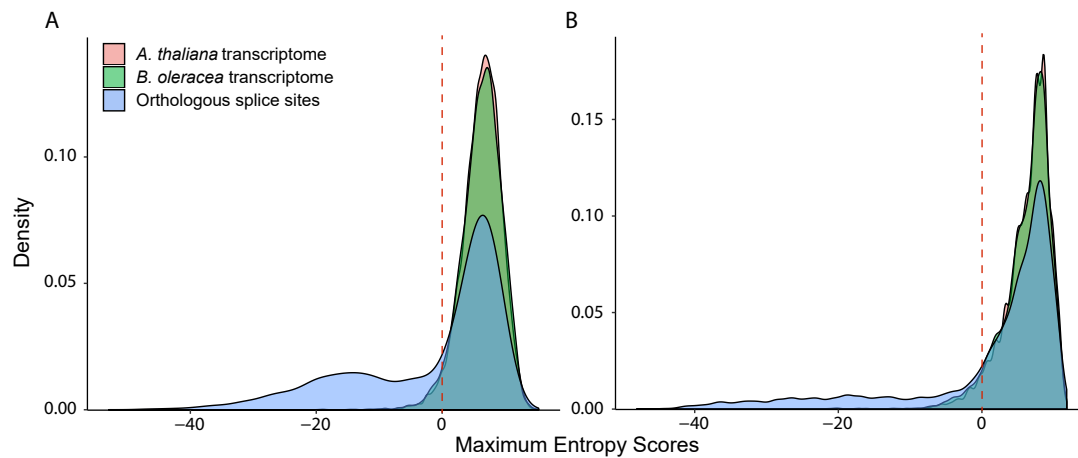


Figure S1: **Density distribution of MES** for splice sites in the donor (A) and acceptor (B) site as identified in *A. thaliana* (pink) and *B. oleracea* (green) transcriptomes compared to orthologous splice sites (blue) identified by position in the Cactus generated WGA. The red line denotes the Maximum Entropy threshold used to consider a splice site as real in the alignment (MES > 0).

Dataset 1:

TrackHubs of WGA, including annotation of splicing sites, Araport11 annotations and own lincRNA annotation.
www.bioinf.uni-leipzig.de/Publications/SUPPLEMENTS/19-001/BrassicaceaeWGA/hub.txt

Dataset 2:

Table of the splice sites, the table contains all the splicing sites that we have predicted for *A. thaliana* and their homologs by position in the 15 species of the WGA. <http://www.bioinf.uni-leipzig.de/Publications/SUPPLEMENTS/19-001/SplicingMap.tsv>

Dataset 3:

Scripts used in creation of splicing map.
bitbucket.org/JoseAntonioCorona/splicing_map_plants.

Dataset 4:

Conservation table by position of TE and lincRNAs overlapping TE.
<http://www.bioinf.uni-leipzig.de/Publications/SUPPLEMENTS/19-001/lincRNA-overlap-TE.tsv>

Dataset 5:

BED files of our lincRNAs for *A. thaliana* and their homologs by genomic position in the 15 species of the WGA
<http://www.bioinf.uni-leipzig.de/Publications/SUPPLEMENTS/19-001/lincRNAs-position/>