

## Supplementary Tables

**Table S1** List of all genomes and annotation versions used for the experiments. The species with \* are the model species used for the comparison of TSEBRA and EVM.

Species	Annotation version	Genome size (Mb)	Number of genes
<i>Arabidopsis thaliana</i> *	Tair Araport 11 (Jun 2016)	119	27,445
<i>Bombus terrestris</i>	NCBI Annotation Release 102(Apr 2017)	249	10,581
<i>Caenorhabditis elegans</i> *	WormBase WS271 (May 2019)	100	20,172
<i>Danio rerio</i>	Ensembl GRCz11.96 (May 2019)	1345	25,254
<i>Drosophila melanogaster</i> *	FlyBase R6.28 (Jun 2019)	138	13,929
<i>Medicago truncatula</i>	MtrunA17r5.0-ANR-EGN-r1.6 (Feb 2019)	430	44,464
<i>Parasteatoda tepidariorum</i>	NCBI Annotation Release 101 (May 2017)	1445	18,602
<i>Populus trichocarpa</i>	JGI Ptrichocarpa 533 v4.1 (Nov2019)	389	34,488
<i>Rhodnius prolixus</i>	VectorBase RproC3.3 (Oct 2017)	707	15,061
<i>Tetraodon nigroviridis</i>	TETRAODON8.99 (Nov 2019)	359	19,589
<i>Xenopus tropicalis</i>	NCBI Annotation Release 104 (Apr 2019)	1449	21,821

**Table S2** F1-score on CDS, transcript, and gene level for BRAKER1 (RNA-seq hints), BRAKER2 (protein hints), TSEBRA\_EVM, EVM using comparable evidence, and TSEBRA with hints generated by the BRAKER runs. The methods were run three times for each species, allowing different minimum evolutionary distances ((i) species excluded, (ii) family excluded or (iii) order excluded) for the proteins used as support. A different set of genome partitions, each totaling 90% of the genome size from a species, was sampled for the evaluation of all methods for each model species and test setting.

	CDS level F1-score				
	BRAKER1	BRAKER2	EVM	TSEBRA_EVM	TSEBRA
<i>A. tha.</i> (i)	81.67	84.62	85.10	<b>87.70</b>	87.34
<i>A. tha.</i> (ii)	81.67	84.00	84.07	86.46	<b>86.93</b>
<i>A. tha.</i> (iii)	81.87	84.01	84.41	86.21	<b>86.90</b>
<i>C. ele.</i> (i)	85.76	87.26	88.15	88.03	<b>88.27</b>
<i>C. ele.</i> (ii)	85.83	80.91	<b>86.46</b>	85.05	84.46
<i>C. ele.</i> (iii)	85.87	81.13	<b>86.14</b>	85.13	84.48
<i>D. mel.</i> (i)	79.88	84.03	83.74	85.37	<b>85.71</b>
<i>D. mel.</i> (ii)	79.20	79.86	80.78	81.81	<b>82.95</b>
<i>D. mel.</i> (iii)	79.82	76.79	79.67	79.89	<b>81.66</b>
	Transcript level F1-score				
	BRAKER1	BRAKER2	EVM	TSEBRA_EVM	TSEBRA
<i>A. tha.</i> (i)	53.52	60.25	61.01	<b>64.92</b>	64.05
<i>A. tha.</i> (ii)	53.48	57.02	56.63	61.97	<b>62.14</b>
<i>A. tha.</i> (iii)	53.78	56.63	57.32	61.35	<b>62.00</b>
<i>C. ele.</i> (i)	53.44	55.13	57.41	59.66	<b>60.54</b>
<i>C. ele.</i> (ii)	53.36	42.52	53.79	54.58	<b>56.09</b>
<i>C. ele.</i> (iii)	53.30	42.71	52.76	54.46	<b>55.94</b>
<i>D. mel.</i> (i)	51.44	58.50	57.86	<b>62.26</b>	61.61
<i>D. mel.</i> (ii)	51.22	50.98	52.23	56.37	<b>57.01</b>
<i>D. mel.</i> (iii)	51.33	46.94	49.90	53.76	<b>55.18</b>
	Gene level F1-score				
	BRAKER1	BRAKER2	EVM	TSEBRA_EVM	TSEBRA
<i>A. tha.</i> (i)	65.20	75.03	75.09	<b>81.45</b>	81.15
<i>A. tha.</i> (ii)	65.22	71.08	70.07	78.92	<b>79.77</b>
<i>A. tha.</i> (iii)	65.51	70.58	70.88	78.35	<b>79.69</b>
<i>C. ele.</i> (i)	63.25	67.18	69.37	73.80	<b>76.31</b>
<i>C. ele.</i> (ii)	63.12	52.15	65.18	69.30	<b>70.92</b>
<i>C. ele.</i> (iii)	63.13	52.29	63.98	68.90	<b>70.78</b>
<i>D. mel.</i> (i)	64.56	75.31	74.75	80.66	<b>81.46</b>
<i>D. mel.</i> (ii)	64.12	66.08	67.81	74.46	<b>76.07</b>
<i>D. mel.</i> (iii)	64.44	61.25	64.94	71.34	<b>73.93</b>

**Table S3** Weights for each input source used by the EvidenceModeler for the comparison with TSEBRA\_EVM for each model species and all test settings.

	BRAKER	BRAKER2	RNA-seq	Protein
<i>A. tha.</i> (i)	30	6750	22	22
<i>A. tha.</i> (ii)	30	6750	7	6750
<i>A. tha.</i> (iii)	30	6750	22	22
<i>C. ele.</i> (i)	30	480	900	480
<i>C. ele.</i> (ii)	30	22	60	7
<i>C. ele.</i> (iii)	30	22	60	22
<i>D. mel.</i> (i)	30	6750	45	6750
<i>D. mel.</i> (ii)	30	480	6750	0
<i>D. mel.</i> (iii)	30	22	60	30

**Table S4** Hyperparameters used by TSEBRA\_EVM for the comparison with TSEBRA\_EVM for each model species and all test settings. The hyperparameter sets are described in Implementation.

	Evidence source weight		Low Evidence support threshold		Transcript score threshold			
	Protein ( $w_P$ )	RNA-seq ( $w_R$ )	Intron	Start/stop-codon	$\epsilon_1$	$\epsilon_2$	$\epsilon_3$	$\epsilon_4$
<i>A. tha.</i> (i)	15.0	10.0	0.75	1.0	0.0	0.0	20.0	0.0
<i>A. tha.</i> (ii)	10.0	0.5	0.75	1.0	0.0	0.0	10.0	0.0
<i>A. tha.</i> (iii)	5.5	10.0	0.5	1.0	0.0	0.5	30.0	2.5
<i>C. ele.</i> (i)	5.5	15.0	0.0	0.0	0.0	1.0	10.0	2.5
<i>C. ele.</i> (ii)	0.5	10.0	0.0	0.0	0.0	0.5	20.0	10.0
<i>C. ele.</i> (iii)	0.5	10.0	0.0	0.0	0.0	0.5	10.0	15.0
<i>D. mel.</i> (i)	1.0	0.5	0.875	0.5	0.0	0.0	5.0	0.0
<i>D. mel.</i> (ii)	5.5	0.75	0.25	1.0	0.0	0.5	20.0	10.0
<i>D. mel.</i> (iii)	5.5	1.0	0.25	0.5	0.0	0.5	15.0	15.0