

Arabidopsis Intron Splice-Site Tables

Given below are tables of *Arabidopsis* consensus intron sequences.

The first set were kindly supplied by John W.S. Brown, Scottish Crop Research Institute. See, Brown, J.W.S., Smith, P., and C.G. Simpson (1996) *Arabidopsis* consensus intron sequences. *Plant Mol. Biol.* 32: 531-535.

The second set is older and so were compiled from an uncorrected dataset. They were prepared by J. Mike Cherry, as part of his previous work with the original AAtDB database.

Arabidopsis thaliana DNA.

Set One (SCRI)

*Percentage occurrence of the consensus nucleotides are given in subscript

Table 1: Nucleotide composition of *Arabidopsis* introns at the 5' splice site.

	-3	-2	-1	1	2	3	4	5	6
A	352	621	85	0	0	682	565	197	222
C	361	131	45	0	9	50	140	76	157
G	154	87	777	998	0	100	43	519	115
U	132	159	91	0	989	166	250	206	504
Consensus*	C _{36/A₃₅}	A ₆₂	G ₇₈	G ₁₀₀	U ₉₉	A ₆₈	A ₅₇	G ₅₂	U ₅₁

Table 2: Nucleotide composition of 3' splice sites of *Arabidopsis* introns.

	-18	-17	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2
A	213	203	202	208	176	172	175	168	205	198	197	182	185	156	304	58	997	0	242	251
C	149	140	130	117	137	119	125	108	137	118	107	133	116	103	60	658	0	0	87	156
G	167	182	182	172	166	175	147	195	206	182	177	194	169	109	385	11	0	998	549	163
U	469	473	484	501	519	532	551	527	450	500	517	489	525	630	249	271	1	0	120	428
Consensus*	U ₄₇	U ₄₇	U ₄₈	U ₅₀	U ₅₂	U ₅₃	U ₅₅	U ₅₃	U ₄₅	U ₅₀	U ₅₂	U ₄₉	U ₅₃	U ₆₃	G ₃₉	C ₆₆	A ₁₀₀	G ₁₀₀	G ₅₅	U ₄₃

Table 3: Nucleotide composition of putative branchpoint sequences in *Arabidopsis* introns.

(a) Nearest to 3' splice site

	-5	-4	-3	-2	-1	1	2
A	268	290	0	0	311	979	203
C	124	125	439	0	150	0	191
G	181	184	0	0	314	0	92
U	406	380	540	979	204	0	493
Consensus*	U ₄₁ /A ₂₇	U ₃₉ /A ₃₀	U ₅₅ /C ₄₅	U ₁₀₀	A ₃₂ /G ₃₂	A ₁₀₀	U ₅₀
			Y ₁₀₀	U ₁₀₀	R ₆₄	A ₁₀₀	Y ₇₀

(b) CUNAN preferred to UUNAN (18-60nt)

	-5	-4	-3	-2	-1	1	2
A	265	292	0	0	289	979	257
C	130	100	717	0	151	0	154
G	167	168	0	0	275	0	107
U	417	419	262	979	264	0	461
Consensus*	U ₄₃ /A ₂₇	U ₄₃ /A ₃₀	U ₂₇ /C ₇₃	U ₁₀₀	A ₃₀ /G ₂₈ /U ₂₇	A ₁₀₀	U ₄₇
			Y ₁₀₀	U ₁₀₀	R ₅₈	A ₁₀₀	Y ₆₃

Set Two

Produced by ACEDB 3.0 software (Durbin and Thierry-Mieg) using the AAtDB release 3-5 database.

Of the 1016 Arabidopsis DNA sequences (excluding EST sequences) within AAtDB, 970 introns were defined.

The "|" denotes the exon-intron boundary. The columns represent the nucleotide position relative to the exon-intron boundary. The rows present the total number of the different nucleotides observed in the dataset.

5' consensus

	--- intron --->											
A	281	338	588	97		11	21	635	545	201	226	307
C	230	347	128	45		4	8	47	136	87	147	173
G	174	158	98	740		933	29	107	52	474	116	112
T	285	127	156	88		22	912	181	237	208	481	378
X	0	0	0	0		0	0	0	0	0	0	0

3' consensus

	<--- intron ---											
A	174	183	172	291	77	912	27		235	249	274	272
C	123	117	107	61	600	14	4		90	142	167	182
G	194	161	101	360	29	27	931		529	174	215	270
T	479	509	590	258	264	17	8		116	405	314	246
X	0	0	0	0	0	0	0		0	0	0	0