Supplementary materials

Intron presence and its association with edited sites

ccmFc

Two parallel losses of the *ccm*Fci1 were found in Alismatales. These are the only cases where these introns have been reported missing in the Angiosperms. In *Hydrocharis* the loss of the intron was coupled with the removal of 15 (of 18) edited sites, strongly supporting a retroprocessing event. In contrast, the lack of the intron in *Helanthium* is not associated with modification in the editing pattern of this gene, indicating that another mechanism may be involved.

cox2

*cox*2i1 has been lost once in the ancestor of all Alismatales, except for Araceae and Tofieldiaceae (only represented here by *Spirodela* and *Triantha*). The loss of this intron is not associated with changes in editing frequency, supporting an alternative mechanism of intron removal. In contrast, *cox*2i691 has been lost at three times: once in the ancestor of Alismataceae, once in *Lagarosiphon*, and once in the ancestor of a clade comprising seven members of Hydrocharitaceae. In all cases intron loss is associated with a reduction in RNA editing. The loss of *cox*2i691 in Alismataceae is associated with the removal of the only two edited sites in intron 3 together with six edited sites in the 3' end of exon 2 (**Fig S2**). In the *Hydrilla-Enhalus* clade nine edited sites have been simultaneously removed from exons 2 and 3. A second reduction of edited sites occurred in some members of Hydrocharitaceae, this time eliminating all five edited sites remaining in the 5' end of the intron-lacking gene.

nad1

*Nad*1i2 was lost once in *Hydrocleys* and *Limnocharis* clade, and once in the *Hydrilla-Enhalus* clade, once in *Caldesia* and once in *Helanthium*. In the first two cases, intron removal is coupled with complete removal of edited sites in exons 2

and 3 (**Fig S3**), supporting a RT-mediated model of intron removal. In *Caldesia* and *Helanthium* only four and three out of five edited sites where lost simultaneously with the intron.

nad2

*nad*2i156 has been lost independently three times within Alismatales. None of the events were associated with simultaneous loss of edited sites (**Fig S4**). In members of Alismataceae we observed that the intron was removed prior to the removal of edited sites (in *Sagittaria* and *Helanthium* eight of nine sites were lost), whereas in some Hydrocharitaceae removal of the intron occurred later than edited sites removal. In the ancestor of the *Hydrilla* – *Enhalus* clade 13 out of 18 edited sites were removed, with a posterior removal of the intron in a more inclusive clade formed by *Halophila, Thalassia,* and *Enhalus*. It is noteworthy that most edited sites are lost in the 3' end of exon 2, which is adjacent to a *trans*-spliced intron (*nad*2i542). However, the *trans*-spliced intron was never lost.
In contrast, *nad*2i709 was lost in the ancestor of *Enhalus, Thalassia,* and *Halophila*, and in the ancestor of *Nechamandra* and *Vallisneria*, in both cases coupled with lack of editing in exons 4 and 5 (with the exception of one edited site in exon 5 in *Nechamandra-Vallisneria,* **Fig S3**). However, the removal of edited sites in exon 4 occurred earlier in the phylogeny of Hydrocharitaceae.

nad4

*nad*4i2 was lost in *Hydrocleys* together with 11 flanking edited sites in the 3' end of exon 2 and 12 edited sites in the 5' end of exon 3 (**Fig S5**). This intron was also lost in *Caldesia*, with the removal of two and 16 edited sites, and in the ancestor of *Echinodorus, Sagittaria* and *Helanthium* together with all 12+20 adjacent edited sites. All these instances strongly support a RT-mediated model of intron loss.
The situation in the Hydrocharitaceae is, however, quite different. Most or all edited sites in exon 2 are lost in all members of Hydrocharitaceae with the exception of *Stratiotes*, but *nad*4i1 is present in all taxa. Similarly, editing frequency is reduced in exon 3 in *Najas* and in *Ottelia*, without being associated

to the lack of the *nad*2i542 intron. This intron is, however, lost at a later stage in the common ancestor of *Enhalus*, *Thalassia* and *Halophyla*, the common ancestor of *Vallisneria* and *Nechamandra* and in *Hydrilla*. In all cases the intron is lost in parallel with the flanking edited sites in exon 4.

nad5

*nad*5i1872 has been lost in *Vallisneria*, together with all six edited sites in exon 4 and a single edited site in exon 3 (**Fig S6**). *nad5*i1872 has also been lost in the common ancestor of *Enhalus, Thalassia* and *Halophila*, but in this case the whole *nad*5 gene has lost its edited sites. The reduction in edited sites has not occurred as a single event. Whereas the lack of edited sites in exons 4 and 5 is associated with the loss of *nad*5i1872 in this clade, the removal of edited sites in intron 2 seems to have occurred earlier in evolution. Although no editing is needed for *nad*5, its two trans-spliced introns (*nad*5i1455 and *nad*5i1477) are conserved as well as the *cis*-spliced *nad*5i1.

nad7

Within Alismataceae (**Fig 3**) the intron *nad*7i676 was lost in the common ancestor of the family, simultaneously with the loss of all but one edited site in the adjacent exons (17 edited sites lost). If the removal of all edited sites and the intron was caused by retroprocessing, the only edited site present in exon 4 most likely has been gained prior to this event. This edited site is lost later in *Echinodorus*, associated with the loss of intron *nad*7i917 and 5 edited sites present in the exon 5. The third intron loss in this family occurred in the common ancestor of *Hydrocleys* and *Limnocharis*, where *nad*7i140 was lost without any change in editing frequency.

In the Hydrocharitaceae, the earliest intron loss involves the lack of *nad*7i676 in all Hydrocharitaceae with the exception *Stratiotes*. Although edited sites are later on lost in the flanking exons 3 and 4, this reduction in editing frequency is not associated with the loss of *nad*7i676. The second intron loss involves the *nad*7i917, lacking in the seven species of the *Enhalus – Hydrilla* clade. In this case intron loss is associated with the loss of five edited sites in intron 4 and the four

adjacent edited sites in exon 5. *Nad*7i209 has been lost twice in Hydrocharitaceae, once in the *Nechamandra-Vallisneria* clade and once in *Hydrilla*. In both cases the two edited sites present in one of the flanking exons (exon 2) were removed simultaneously. Similarly, the *nad*7i140 intron was removed in the *Enhalus* – *Halophila* clade simultaneously with the six flanking edited sites, in *Hydrilla* together with three flanking edited sites, and in *Lagarosiphon*, with the five edited sites present in exon 1. This intron was also removed in *Nechamandra* and *Vallisneria*, but in this case no edited sites was lost simultaneously.

Table S1. Number of edited sites (ES) present in 21 genes in members of the Alismatales.

	atp1	*atp4*	*atp6*	*atp8*	*atp9*	*ccmB*	*ccmC*	*cob*	*cox1*	*cox2*	*cox3*	*mttB*	*nad1*	*nad2*	*nad3*	*nad4*	*nad4L*	*nad5*	*nad6*	*nad7*	*nad9*	Total ES	*cis*-spliced introns
Triantha	6	10	22	5	7	43	37	21	27	18	16	32	27	31	21	55	13	39	20	34	12	496	16
Spirodela	5	10	18	5	7	44	36	17	22	17	15	35	26	29	19	55	12	31	19	34	11	467	16
Aponogeton	6	9	16	4	2	40	25	17	26	14	14	30	24	33	19	52	14	36	14	34	11	440	15
Caldesia	0	10	0	4	1	19	7	2	0	8	14	28	15	25	15	29	13	34	12	16	9	261	11
Echinodorus	1	10	0	4	1	19	7	1	0	8	14	13	17	22	15	22	13	34	13	8	9	231	10
Helanthium	1	10	0	4	1	19	7	1	0	8	14	28	14	17	15	22	13	34	13	15	9	245	9
Luronium	1	10	0	4	1	19	7	3	0	8	14	28	18	27	15	22	13	15	12	15	9	241	13
Ranalisma	1	10	0	4	1	19	0	1	0	8	14	28	18	25	15	48	13	34	13	14	9	275	12
Sagittaria	1	10	0	4	1	19	7	1	0	8	14	28	16	15	15	22	13	34	13	15	9	245	11
Hydrocleys	1	10	0	4	0	9	0	4	0	4	12	16	11	20	16	14	13	28	13	14	9	198	10
Limnocharis	0	10	0	4	1	19	7	12	0	8	15	28	12	27	15	47	13	34	13	16	9	290	11
Maundia	7	9	16	4	2	42	29	16	27	16	12	30	24	33	17	54	13	37	13	34	12	447	15
Scheuchzeria	7	10	16	4	2	41	29	17	28	16	13	30	28	34	18	53	13	38	12	36	12	457	15
Triglochin	7	10	16	4	2	29	29	18	25	15	14	30	24	31	17	52	12	35	14	34	12	430	15
Potamogeton	7	8	17	5	2	40	28	17	26	15	13	29	24	34	17	51	10	35	13	33	12	436	15
Zannichellia	7	7	17	5	2	40	28	17	26	15	13	28	25	32	17	50	9	35	13	33	12	431	15
Zostera	7	8	16	5	0	38	28	17	26	15	13	29	24	30	17	50	11	35	14	34	11	428	15
Amphibolis	7	7	16	4	2	39	27	12	18	15	11	30	25	31	16	51	10	36	12	32	11	412	15
Halodule	7	6	16	5	2	40	27	13	21	15	12	30	22	32	15	49	7	35	13	30	11	408	15
Butomus	7	10	19	4	2	41	27	15	24	17	14	32	27	32	19	53	13	36	14	33	11	450	15

Stratiotes	5	10	19	4	2	38	28	16	25	17	13	30	24	32	16	53	13	35	14	30	10	434	15
Hydrocharis	1	6	18	0	0	26	28	0	1	6	0	8	21	30	16	18	0	28	12	25	10	254	13
Ottelia	5	8	19	4	2	34	26	0	22	13	14	28	23	8	1	8	13	33	13	13	10	297	14
Lagarosiphon	3	10	18	1	2	34	28	1	0	6	0	21	23	25	1	20	14	34	0	10	3	254	11
Hydrilla	0	8	3	2	2	34	4	0	1	7	1	25	7	4	1	6	8	18	?	6	10	147	7
Najas	5	6	18	3	2	36	26	0	2	5	0	27	18	5	14	11	12	15	13	6	10	234	11
Nechamandra	0	6	5	4	0	23	26	0	1	0	0	20	13	0	7	6	12	13	1	5	1	143	7
Vallisneria	0	6	3	4	0	23	26	0	1	0	0	20	11	0	7	6	12	7	1	4	1	132	6
Halophila	0	3	4	2	2	4	0	0	0	0	0	2	9	4	0	4	12	0	0	0	5	51	6
Thalassia	0	10	5	4	2	7	0	0	0	0	0	4	12	4	1	6	12	1	0	1	10	79	6
Enhalus	0	10	5	4	0	7	0	0	0	0	0	4	12	4	1	6	12	1	0	1	9	76	6

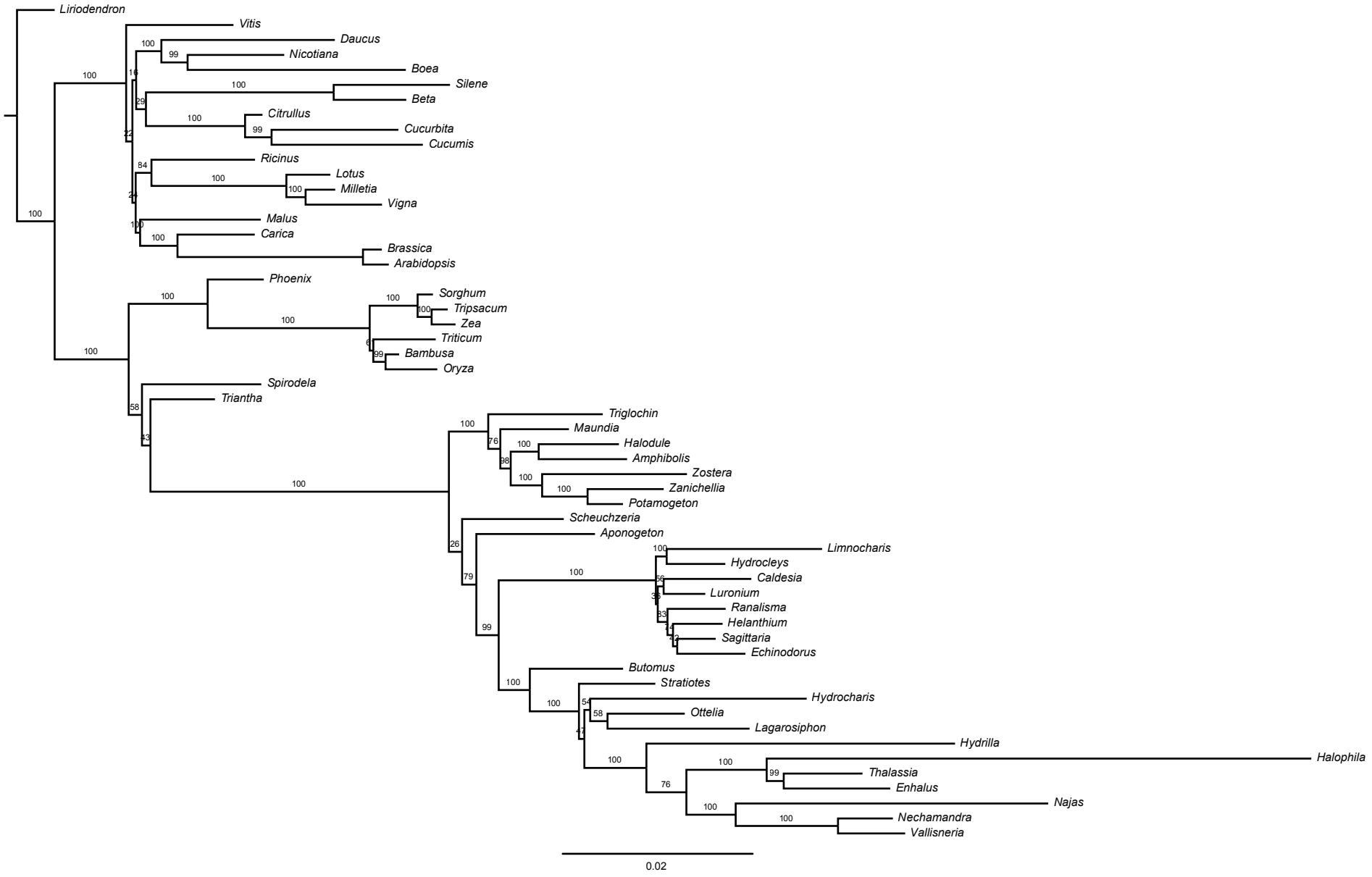

Fig. S1. ML tree based on 20 mitochondrial genes for 46 angiosperm taxa

The tree was obtained using RAxML ver. 7.2.8 with a GTR+gamma model. Bootstrap values are indicated above the branches, and are based on 10000 replicates of rapid bootstraping.

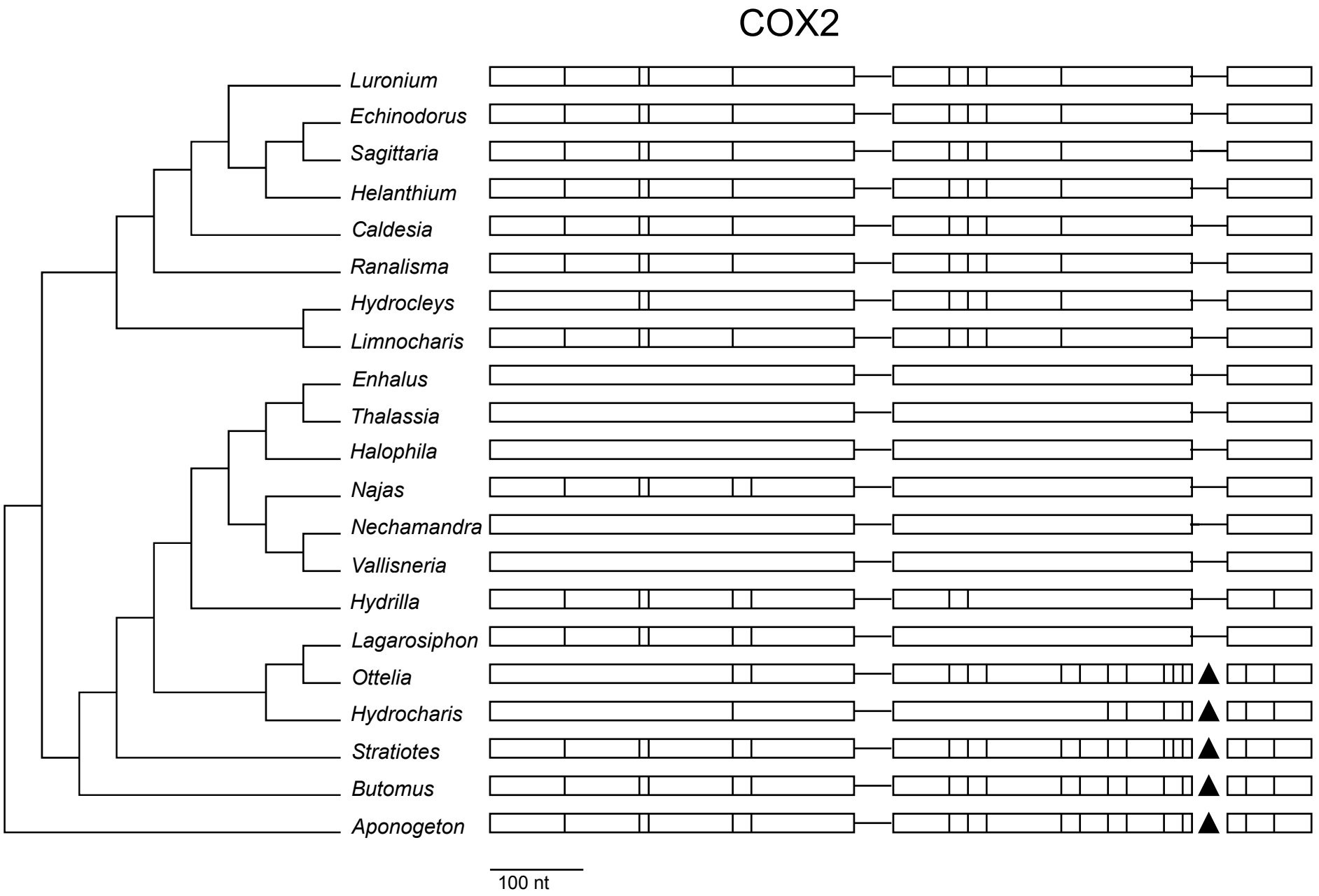

Fig. S2 - Intron and edited sites distribution in the *cox*2 gene for selected members of the Alismatales.
Rectangles represent exons, and black triangles represent a *cis*-spliced intron. Horizontal lines joining rectangles indicate that no intron is present at that position. Vertical lines inside rectangles represent positions that need editing (edited sites).

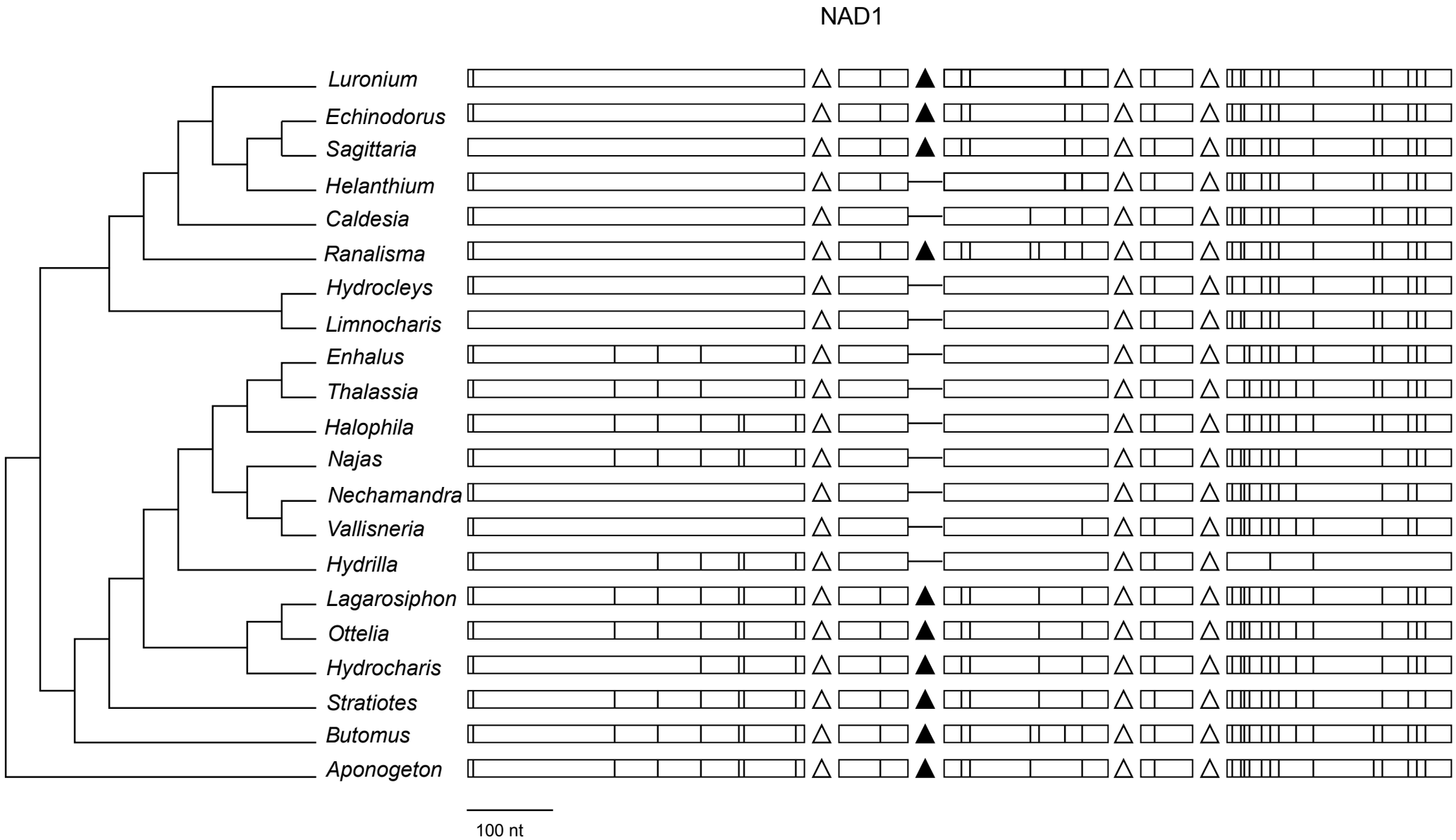

Fig S3. Intron and edited sites distribution in the *nad*1 gene for selected members of the Alismatales

Rectangles represent exons, a black triangle represent a cis-spliced intron and an open triangle represents a trans-spliced intron. Horizontal lines joining rectangles indicate that no intron is present at that position. Vertical lines inside rectangles represent positions that need editing (edited sites).

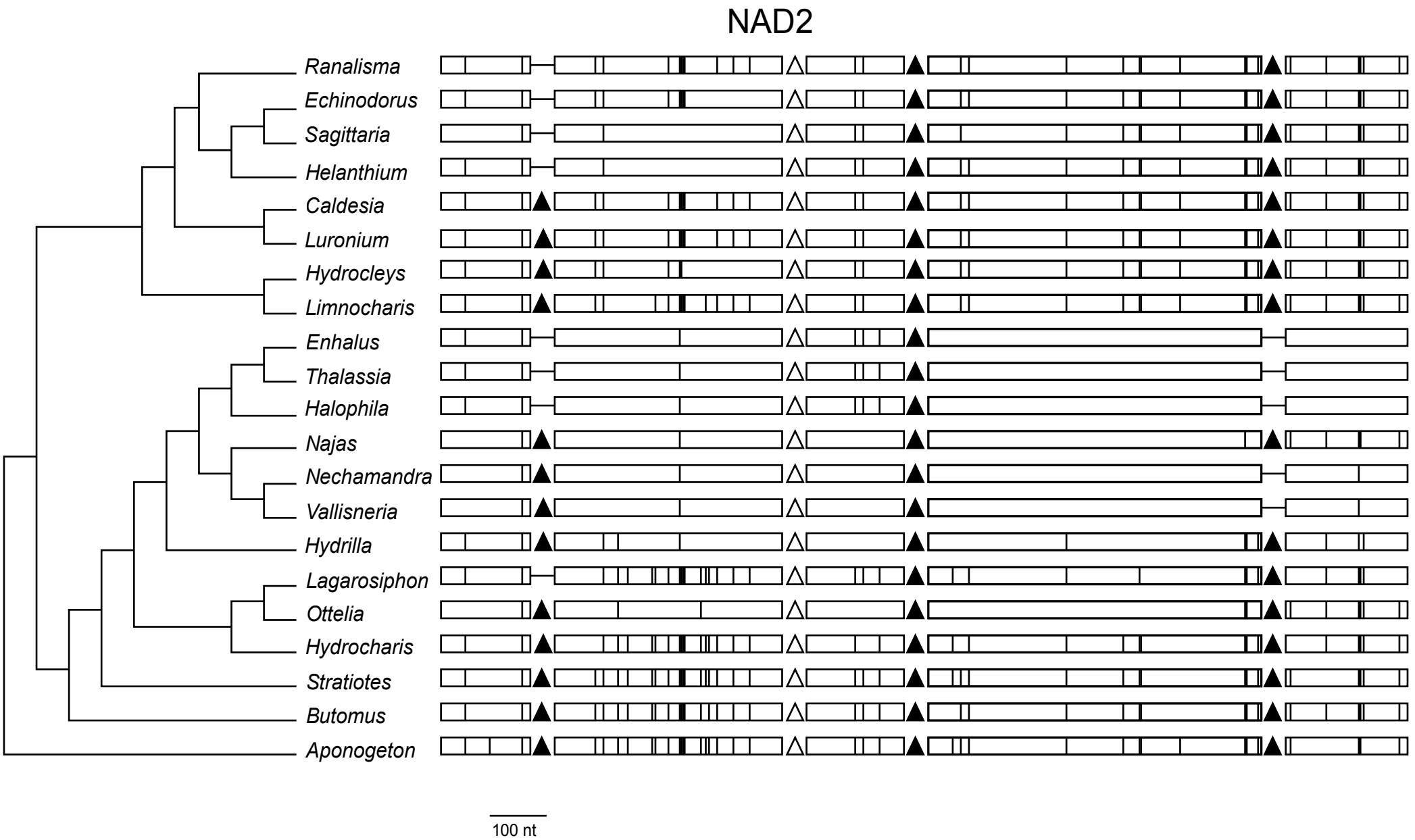

Fig S4. Intron and edited sites distribution in the *nad*2 gene for selected members of the Alismatales

Rectangles represent exons, a black triangle represent a cis-spliced intron and an open triangle represents a trans-spliced intron. Horizontal lines joining rectangles indicate that no intron is present at that position. Vertical lines inside rectangles represent positions that need editing (edited sites).

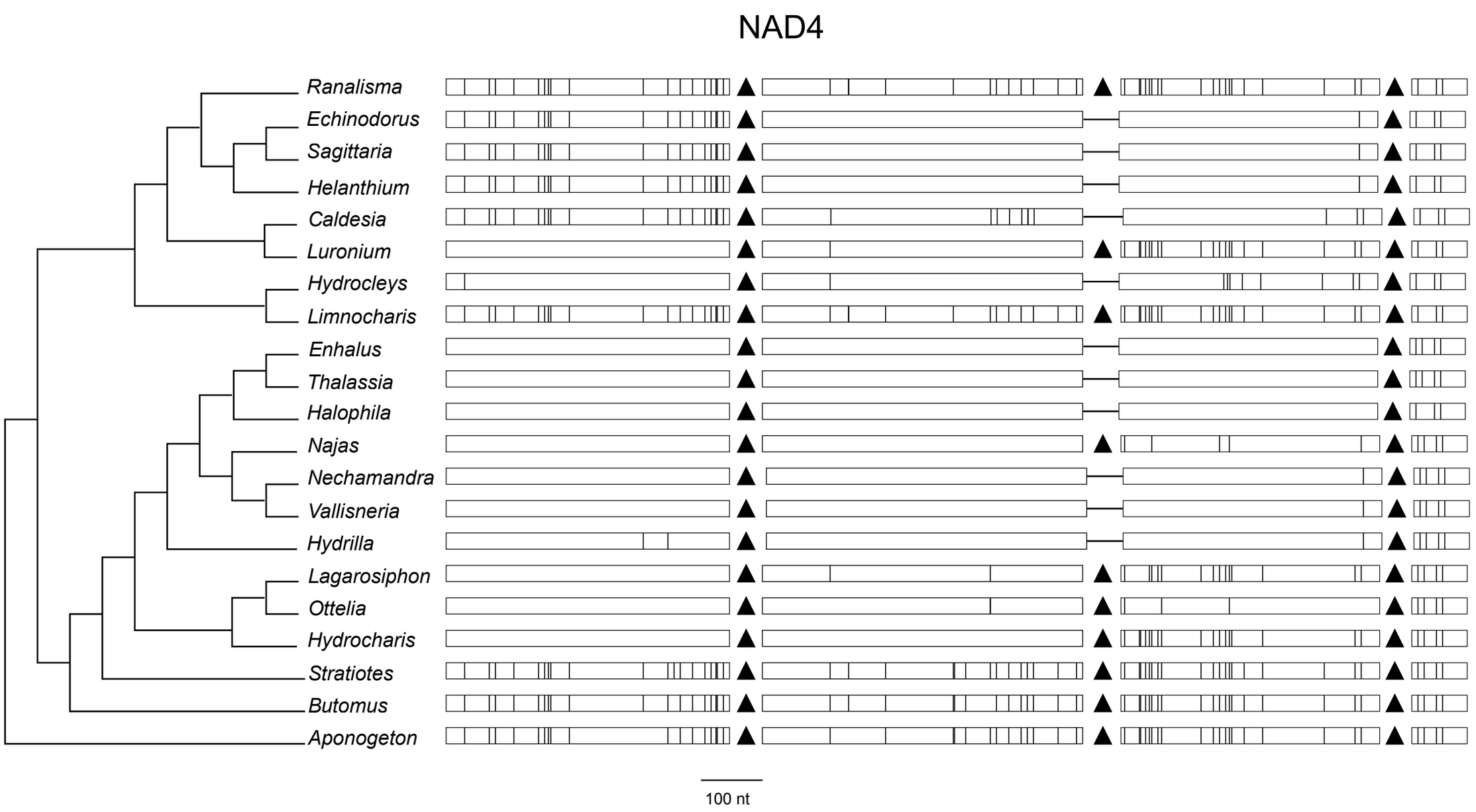

Fig S5. Intron and edited sites distribution in the *nad*4 gene for selected members of the Alismatales
Rectangles represent exons, and a black traingle represent a *cis*-spliced intron. Horizontal lines joining rectangles indicate that no intron is present at that position. Vertical lines inside rectangles represent positions that need editing (edited sites).

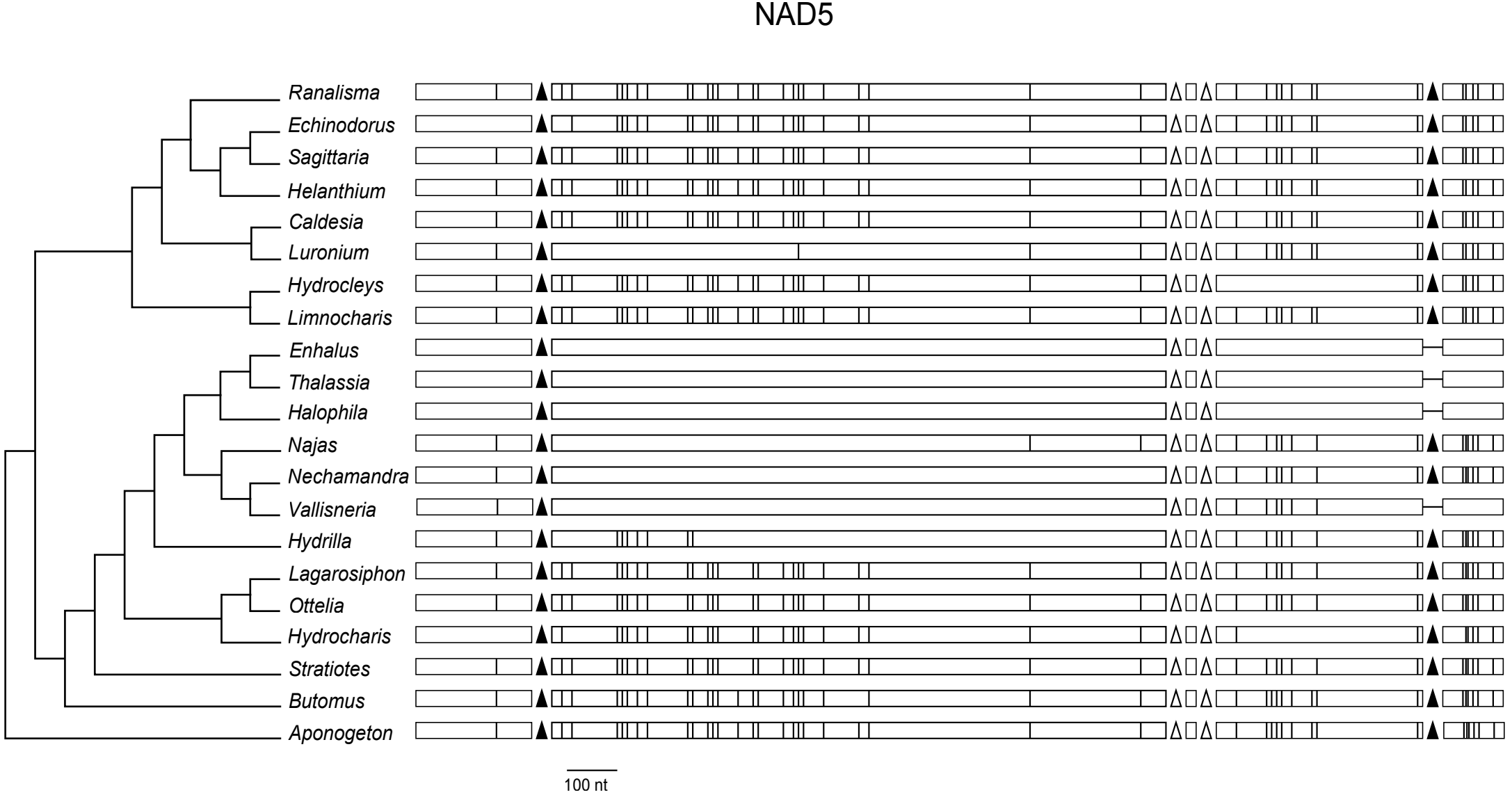

Fig S6. Intron and edited sites distribution in the *nad*5 gene for selected members of the Alismatales

Rectangles represent exons, black triangles represent cis-spliced introns, open triangles represent trans-spliced introns. Horizontal lines joining rectangles indicate that no intron is present in that position. Vertical lines inside rectangles represent positions that need editing (edites sites).