

Supporting Information

Genome-wide identification and characterization of amino acid permease gene family in *Nicotiana tabacum*

Yingying Zhao^{a,b, §}, Yalong Xu^{b, §}, Zhong Wang^b, Jianfeng Zhang^b, Xia Chen^b, Zhengfeng Li^c, Zefeng Li^b, Lifeng Jin^b, Pan Wei^b, Lin Zhang^d,
Xiaoquan Zhang^d, Ran Wang^{b, *}, Fang Wei^{a, *}.

^a. School of Life Sciences, Zhengzhou University, No.100 Science Road, Zhengzhou, Henan, 450001, China. Email: zyy244792855@163.com; fangwei@zzu.edu.cn

^b. Zhengzhou Tobacco Research Institute of CNTC, No. 2 Fengyang Street, Zhengzhou, Henan, 450001, China. Email: xuyalong2007@163.com; zhongwang1123@126.com; zhangjianfeng_83@outlook.com; chenxia0372@163.com; 379359757@qq.com; jin_lf@126.com; weipan83 @126.com; wangr@ztri.com.cn

^c. Technology Center, China Tobacco Yunnan Industrial Co., Ltd., Kunming, Yunnan, 650231, China. Email: 1208913744@qq.com

^d. College of Tobacco Science, Henan Agricultural University, No.63 Agriculture Road, Zhengzhou, Henan, 450002, China. Email: hnndzhang@126.com, zxq013415@163.com

*To whom correspondence should be addressed. §These authors contributed equally to this work.

Table S 1 Primers used in this research

Primer set	Target gene	Upstream primer	Downstream primer
1	Ntab0475230	TACAGCAGTTATTGGTTCTGGAG	AAGGAGCGCAGAAGTGTAGTAA
2	Ntab0126970	ACTATTGGACTAGGCTTAGGAGTT	AACTGTTGTTACTGCTACACTAAT
3	Ntab0329990	TTGTTACTTATTACACTTCCGCT	GATAGCAACTCCAAAAAGATTCG
4	Ntab0126360	TAACAGGGACGACAGTAGGAGTT	CTTGCTTTCTGGAGGTTGTGAC
5	Ntab0759460	TTCCACCACAACCTTTATTCTATATT	AGAAAACCTGGTAAGCTCCAAT
6	Ntab0648520	ACTACAGTTACATGGAGGTTGTC	ATAAGGGTAATTTGATACTGAGC
7	Ntab0439400	CGCTAACATTTGCATCGCTATT	ACTATAACATATGCTGTCCTCCAC
8	Ntab0331750	TGATGGAGTAAATGGAGACACG	CCCAATAAGTCCACAGAGCATG
9	Ntab0821150	TATGATAGCATTGGTGTCACTG	AAGTTACAGTTCCAATATGTGTTAC

10	Ntab0800540	ATCAAAAGGAAACTGAGCCAAT	AAAGCAAAAAGGATCATTACAGC
11	NtAAP2-2-CDS	ATGTTGCCAAGGAGTCGAACTCTTC	TTAGTAAATAGTCTTGAAAGGCTTATACGT TTTG
12	NtAAP2-2-RNAi	GGGGACCACTTTGTACAAGAAAGCTGGGTCTC TACTGTTGGTCTTGTCTTGG	GGGGACAAGTTTGTACAAAAAAGCAGGCT CGAGATTCCTGGAGCATCATCT
13	NtAAP2-2-OE	CGAGGTACCATGTTGCCAAGGAGTCGAACTCT TC	CCCACTAGTTTAGTAAATAGTCTTGAAAGG CTTATACGTTTTG
14	nppt	TCGCCGCCAAGCTCTTCAGCAAT	GTGGAGAGGCTATTCGGCT ATGACT
15	Hyg	GTGCTTTCAGCTTCGATG	AACCAAGCTCTGATAGAG
16	NtAAP2-2 flanking- Flag	GTTTACCAGCTTAACTTCTTTCGTGT	CTTATCGTCAT CGTCCTTGTAATCG

Table S2 Three potential amino acid permease genes identified in tobacco which were removed

Gene ID	MW(PI	Amino acid	CDS	gDN	Instability	Conserved
Ntab0599	25.97	8.5	230	693	1333	32.12	1 Aa_trans
Ntab0860	87.12	6.6	798	2397	1952	32.13	2 Aa_trans
Ntab0931	42.78	9.5	384	1155	2337	27.63	1 Aa_trans

Table S3 Identities on amino acid sequence and their coding sequence between the fifteen members of *N. Tabacum*.

		Amino acid identity 69.96%														
		Ntab047	Ntab0186	Ntab0484	Ntab0049	Ntab0126	Ntab0889	Ntab0329	Ntab0800	Ntab0821	Ntab0126	Ntab0759	Ntab0648	Ntab0439	Ntab0331	Ntab0927
		5230	340	720	050	970	300	990	540	150	360	460	520	400	750	350
Nucleotide identity 71.04%	Ntab0475230		100	98.33	98.33	86.16	86.16	85.95	63.48	63.55	57.49	58.84	54.43	56.21	56.76	57.14
	Ntab0186340	100		98.33	98.33	86.16	86.16	85.95	63.48	63.55	57.49	58.84	54.43	56.21	56.76	57.14
	Ntab0484720	97.37	97.37		100	86.1	86.1	85.89	64.45	63.94	58.48	59.46	54.56	56.85	56.99	57.38
	Ntab0049050	97.37	97.37	100		86.1	86.1	85.89	64.45	63.94	58.48	59.46	54.56	56.85	56.99	57.38
	Ntab0126970	82.92	82.92	82.61	82.61		100	99.17	64.65	64.72	56.97	58.51	53.81	55.17	57.47	57.47
	Ntab0889300	82.92	82.92	82.61	82.61	100		99.17	64.65	64.72	56.97	58.51	53.81	55.17	57.47	57.47
	Ntab0329990	82.51	82.51	82.54	82.54	96.67	96.67		65.82	64.33	61.1	63.55	58.57	54.88	57.26	57.26
	Ntab0800540	64.16	64.16	63.64	63.64	64.57	64.57	63.92		94.42	55.95	57.96	53.42	56.13	54.78	54.97
	Ntab0821150	63.9	63.9	63.32	63.32	64.33	64.33	63.88	93.76		55.26	57.4	53.32	55.51	55.51	55.8
	Ntab0126360	60.78	60.78	60.49	60.49	60.54	60.54	56.77	50.47	50.93		93.01	72.42	77.38	63.87	64.27
	Ntab0759460	63.49	63.49	62.93	62.93	62.77	62.77	58.51	52.13	52.13	91.43		76.57	79.02	65.28	65.49
	Ntab0648520	57.87	57.87	59.16	59.16	58.3	58.3	53.81	47.87	47.97	72.87	77.52		88.91	63.13	62.92
	Ntab0439400	60.78	60.78	62.04	62.04	60.27	60.27	60.32	51.05	50.95	78.94	79.95	88.25		63.66	64.27
	Ntab0331750	60.1	60.1	60.83	60.83	61.08	61.08	61.84	49.41	50.1	63.68	65.56	62.44	65.57		97.25
	Ntab0927350	61.2	61.2	61.74	61.74	62.18	62.18	62.11	50.29	50.97	64.74	66.6	62.79	66.12	96.19	

Table S4 Putative conserved motif in the amino acid sequence of the 15 NtAAP sequence

Motif	Width	Best position match
1	119	<p>GPRNYTYMEAVKGILGGKKVKVCGLIQYLNLFGV</p> <p>AIGYTIAASVSMLAIKRSNCFHRNHRRDPCHMSSNG</p> <p>YMIAFGVTEILFSQIPGFDQVWWLSIVAAIMSFTYST</p> <p>VGLVLGIAKVAG</p> <p>EIQDTIKSPPAEHKTMKKASMLSIGVTTIFYLLCGCI</p>
2	115	<p>GYAAFGNDAPGNLLTGFGFFDPYWLLDIANAAIVIH</p> <p>LVGAYQVYCQPLFAFVEKWSAKKWSKSNFVTAEH</p> <p>DIPIPFVG</p> <p>NLQTNYSKCFDDDGHLRRTGTFWTTSHITAVIGS</p>
3	79	<p>GVLSLAWAIGQLGWVAGPTVMILFAFVVLYTSNLL</p> <p>SQCYRTGD</p> <p>NFFRVIWRTIFVILTTIAMLLPFFNDVVGLLGALGF</p>
4	79	<p>WPLTVYFPIEMYIKQKKIGRWTNQWIGLEMLSAAC</p>

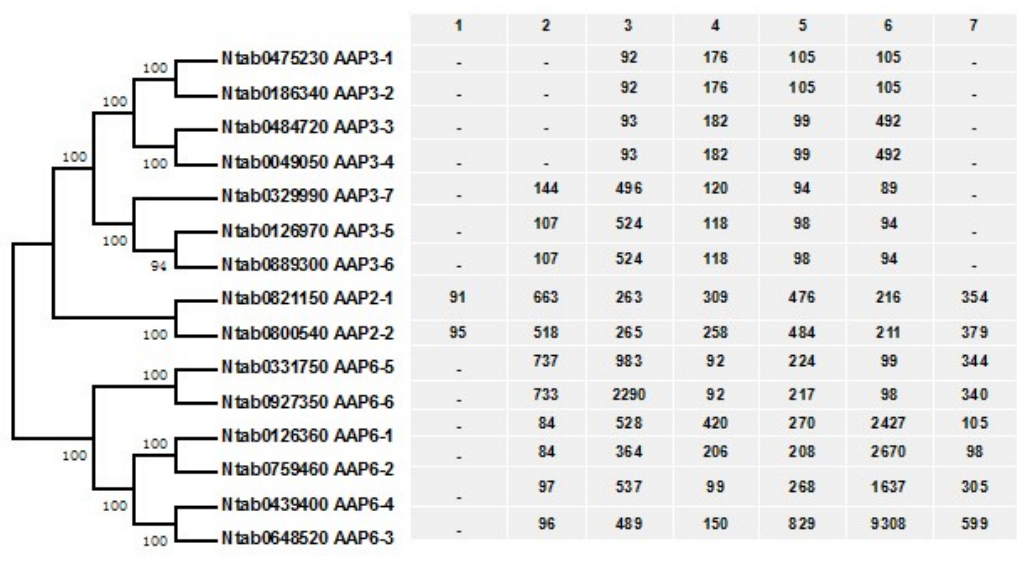
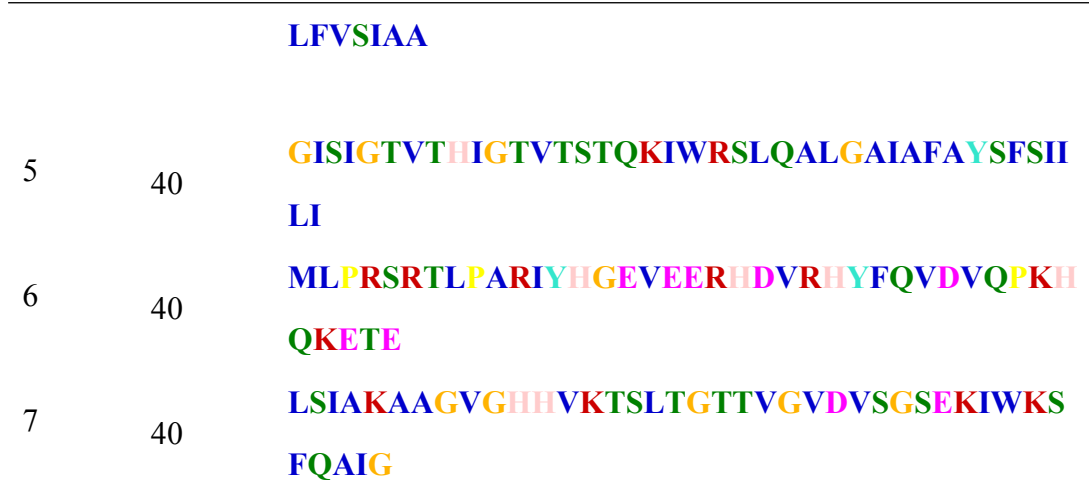


Fig.S 1 Schematic presentation of size and positions of the conservation of intron

numbers of *AAP* genes from *N. tabacum*.

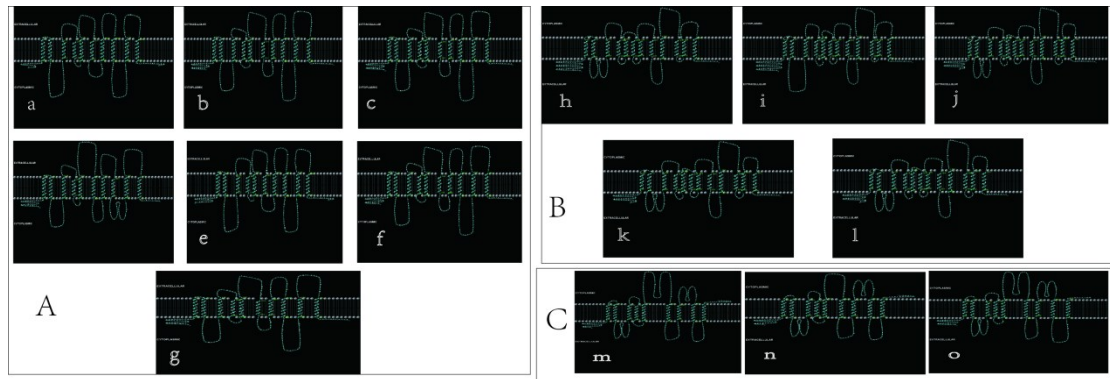


Fig.S 2 The hydrophobic transmembrane domain of *NtAAP* genes. A present N/C terminals are all in cytoplasmic and contains 10 TMs, a-g is NtAAP6-5、NtAAP3-3、NtAAP3-2、NtAAP6-1、NtAAP6-6、NtAAP3-4、NtAAP3-1 respectively; B present N/C terminals are all in extracellular and contains 12 TMs, h-l is NtAAP2-1、NtAAP2-2、NtAAP3-5、NtAAP3-6、NtAAP3-7; N/C-terminal are in extracellular and cytoplasmic, respectively, containing 9 TMs, m-o is NtAAP6-3、NtAAP6-4、NtAAP6-2 respectively.

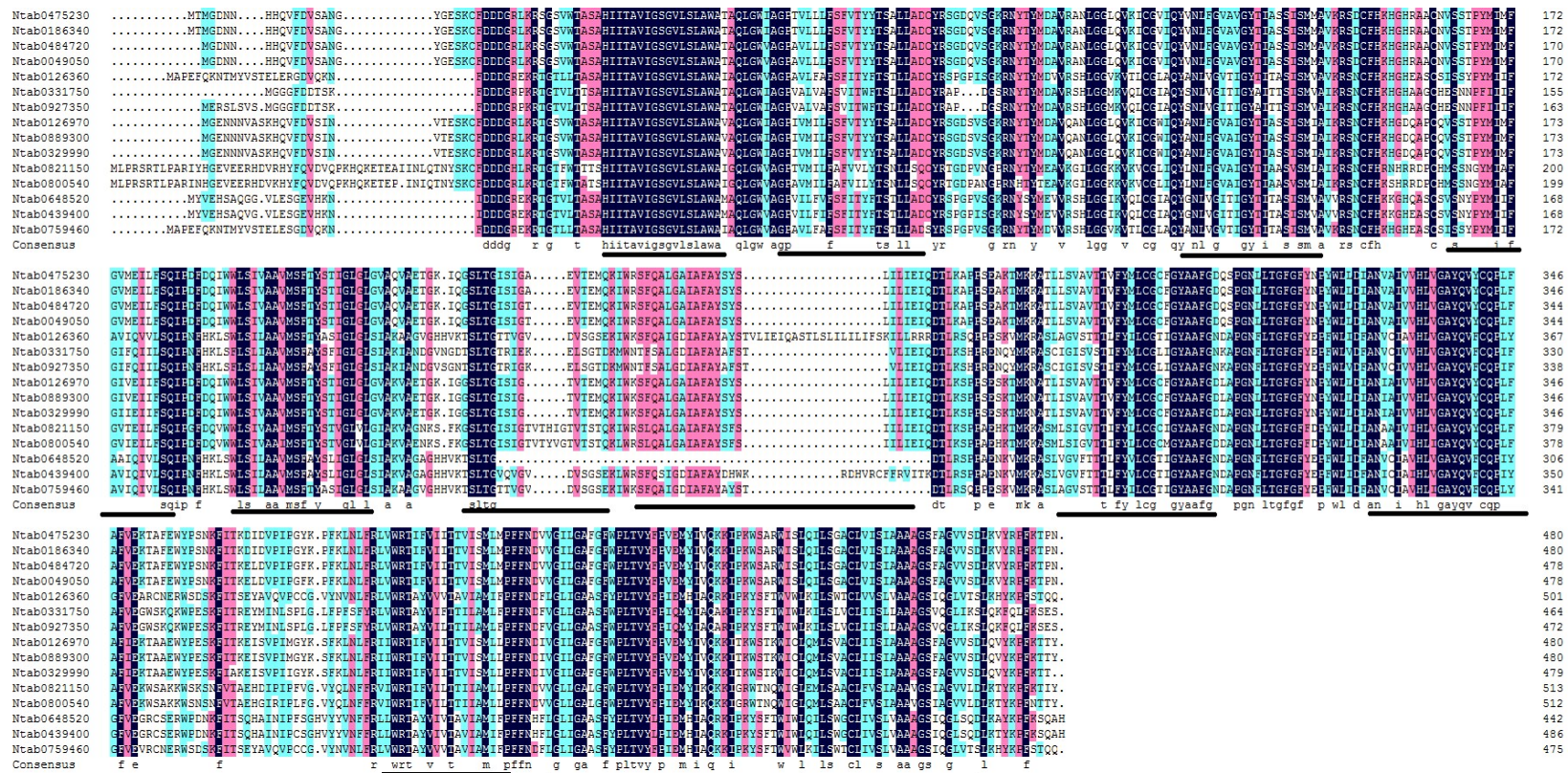


Fig. S3 Multiple alignment of the NtAAP protein sequences in *N. tabaccum* by DNAMAN. The black underline areas represented the transmembrane regions.

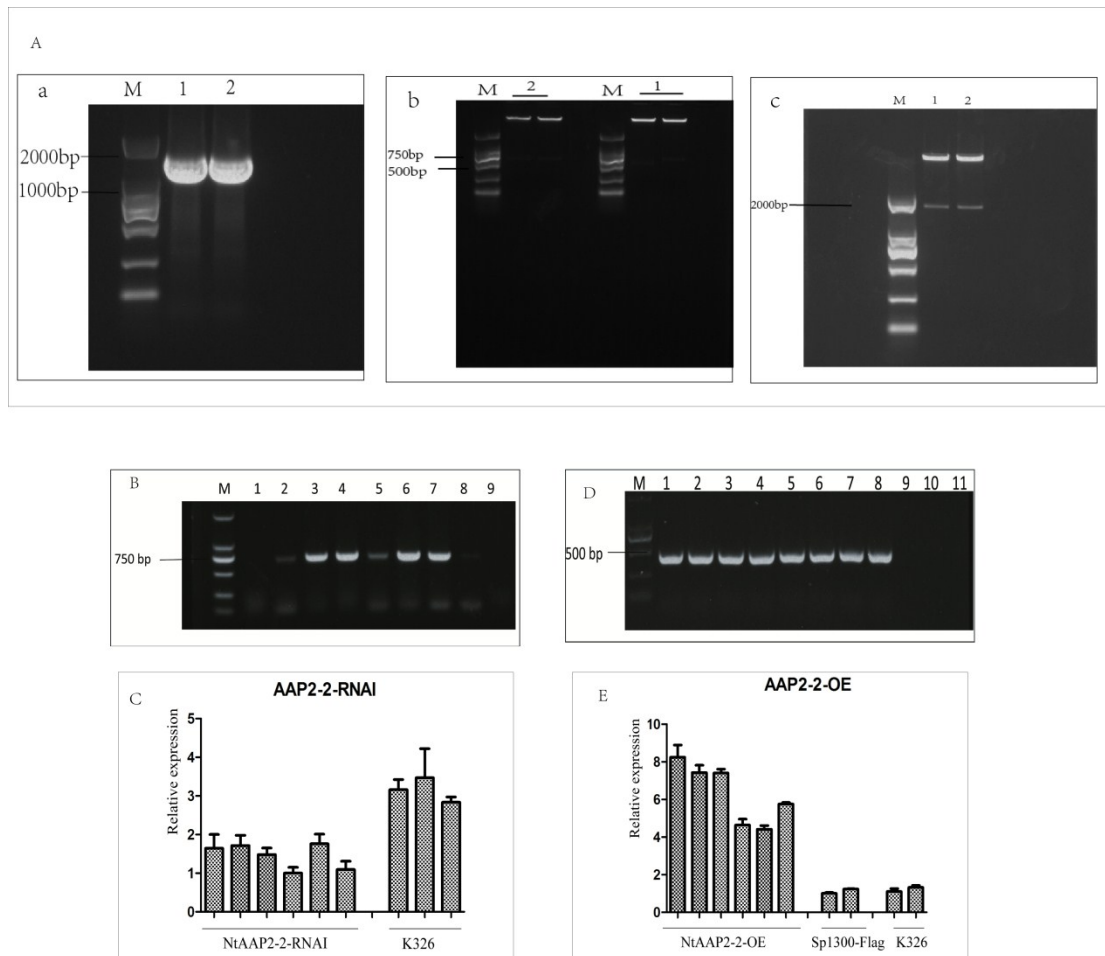


Fig. S4 Clone and molecular analyses of *NtAAP2* transgenic lines and wild-type (K326). A: a, clone of *NtAAP2* gene in *Nicotiana tabacum*; b, the digestion of *NtAAP2-2-RNAi* vector plasmid with *Xho*I and *Xba*I respectively; c, the digestion of *NtAAP2-2-OE* vector plasmid with *Spe*I and *Kpn*I. Lane M, DL Marker 2000; B: Confirmation of the positive *NtAAP2-RNAi* transgenic lines via amplification of the kanamycin-resistance gene by PCR. Lane M, DL Marker 2000; Lane 3-7, positive lines; Lane 8-9, wild type. C: Expression level of *NtAAP2* gene in *NtAAP2-RNAi* transgenic lines by qPCR. D: Confirmation of the positive *NtAAP2-RNAi* transgenic lines via amplification of the *NtAAP2-2* gene flanking primers by PCR. Lane M, DL Marker 2000; Lane 1-8, positive lines; Lane 9-11, wild type. E: Expression level of *NtAAP2* gene in *NtAAP2-OE* transgenic lines by qPCR.