## **Supporting Information**

Title:

Analysis of the conserved N-terminal domains in major ampullate spider silk proteins

## Authors and affiliations:

Dagmara Motriuk-Smith<sup>1\*</sup>, Alyson Smith<sup>1</sup>, Cheryl Y. Hayashi<sup>2</sup>, Randolph V. Lewis<sup>1</sup> <sup>1</sup>University of Wyoming, Department of Molecular Biology, Laramie WY 82071 <sup>2</sup>University of California, Department of Biology, Riverside CA 92521

Corresponding author: Dagmara Motriuk-Smith University of Wyoming College of Agriculture Department of Molecular Biology Dept. 3944 1000 E. University Ave Laramie, WY 82071 e-mail: motriuk@uwyo.edu phone: 307-766-6380 fax: 307-766-5098





Figure 1. Molecular architecture of At.MaSp2 major ampullate silk fibroin. (A) Arrangement of exons (dark green) interrupted by introns (light green) in the At.MaSp2fibroin gene. (B) Alignment of intron sequences showing a high level of conservation between introns *I*2 and *I*3. Abbreviations used: E, exon; I, intron. Exons 1 through 4 (*E*1-*E*4) represent the 5'end of silk fibroin coding region. The 3' end of the fibroin coding region is represented by exon c (*Ec*). Dots represent missing DNA sequence.

# Supporting Information; Figure 1B

12 13	1 1	G T G A G T G A	T T G T T G	СТ	A	тт тт	т с т с	C /	A T A T	T T	G A G A	T T	A C A C	ст	G d	A T A T	T T	ТС		G G	T T	тт тт	G / G /	A A A A	A A	G 1 G 1	A A	c ·	T G T G	G G	G A G A	T T	G T G T	T T	C A C A	T T	тт тт	A G A G	G G	60 60
12 13	61 61	ттст ттст	G G T G G T	C A C A	T	тт тт	T A T A	A	G A G A	A	A G A G	T	C A C A		A A	с с с с	C C	A A A A	A	A A	C I	G T G T	T T	G T G T	T T	ТА	A	ТО		T T	C A C A	A A	C A C A	T	т т т т	A	тт тт	A A A A	C C	120 120
12 13	121 121	T T T T T T T T	ТАТ ТАТ	G T G T	т А т А	A C A C	A T A T	G ·	тс тс	A A	т т т т	C C	G T G T	гт гт	C C	т с т с	G G	A A A A	A	T T	Т	G A G A	C /	A C A C	T T	C 1 C 1	T T	с ( с (	G A G A	A A	C A C A	Ти	A T A T	A	A C A C	Т	G A G A	C T C T	T T	180 180
12 13	181 181	G T A G G T A G	G A A G A A	тт	с с с с	G T G T	G T G T	Т	T G T G	A . A .	A A A A	T T	с с с с	C A C A	A A	C A C A	A A	то	G C G C	C C	G .	T G T G	G G	ст ст	T T	A A A A	AA	A 0 A 0		A A	A T A T	G G		) A ) A	т т т т	A	C A C A	СТ	A A	240 240
12 13	241 241	G A C C G A C C	A T T A T T	T A T A	. A 4 . A 4	A A A A	A A A A	A	C A C A	T T	т с т с	A	T T T T	r c r c	G . G .	A G A G	A A	A 1 A 1	гт гт	T T	A A A A	A A A A	A A	G T G T	A A	A T A T	G G	A 0 A 0	G A G A	C C	A A G G	A A	G A G A	G	T C T C	Т	T A T A	G T G T	T T	300 300
12 13	301 301	A T A A A T A A	A A C A A C	A T A T		A G A G	тт тт	A	тт тт	T T	G T G T	A A	C A C A		T T	тт тт	T T	A A A A	G	A A	C A C A	а т а т	C /	A A A A	T T	T T T T	т т	с ( с (	C A C A	T T	G G G G	G	A A A A	, Т Т	T A T A	A	C A C A	A T A T	G G	360 360
12 13	361 361	T C A C T C A C	T G C T G C	СТ	AA	A A A A	т т т т	ТО	ст ст	T T	т т т т	A A	T A T A	r c r c	C C	T G T G	T T	тт	r c r c	T T	A I	C A C A	A A	А Т А Т	A A	G A G A	A	ТИ	A A A A	G G	A A A A	A	A G A G	6 C 6 A	G A G A	A	T G T G	G T . G T .	A A	420 420
12 13	421 421	T A T G T A T G	A T T A T T	A T A T	т	G A G A	тт тт	A A A A	A A A A	A . A .	A G A G	T	G C G C		C I	а т а т	A A	тт	T T T	T T	A A A A	A T A T	G G	тт	A A	C A C A	C C	A 0 A 0	G A G A	G G	G T G T	T T	СТ	A	G A G A	Т	T G T G	C A C A	A A	480 480
12 13	481 481	C T C A C T C A	A T T A T T	T T C T	т и ти	A A A A	A T A T	A A A A	A C A C	A . A .	A G A G	G G	G A G A	A A	T T	ТА ТА	T T	A T A T	ГА ГА	A A	A ·	T G T G	C /	а т а т	G G	A A A A	T	A 0 A 0	G G G	C C	G T G T	G	A G A G	G G	G T G T	A	тт	A A A A	Т	540 540
12 13	541 541	A T C A A T C A	T T G T T G	C G	6 A <i>A</i> 6 A <i>A</i>	A T A T	A A A A	ТА	A A A A	G . G .	A A A A	A A	T A T A	, т , т	T T	т т т т	A A	C A C C	A	T T	G A G A	A A A A	C C	T A T A	T T	C A C A	A	ТИ	A A A A	A A	A T A T	T T	т с т с	; c ; c	T G T G	C C	т т т т	T G T G	T T	600 600
12 13	601 601	T T A A T T A A	A A G A A G	A A A A	. A .	тт	T A T A	G (	G T G T	T T	G G G G	G	A A A A	A A A A	Т	A A A A	T T	A C A C	ЭТ ЭТ	G G	G A G A	A A A A	A A	ст ст	T T	Т А Т А	T	G G	T G T G	A A	A T A T	A	A G A G	ЭТ ЭТ	A A A A	AA	A G A G	T C T C	T T	660 660
12 13	661 661	A T T A A T T A	A T A A T A	A A A A	. A . . A .	тт тт	A A A A	A	тт тт	Т	A A A A	A A	A A A A	G	A A	C A C A	G G		G A G A	T T	A I	G C G C	C C	тт	C C	G A G A	AA	A ·	ГА ГА	G G	C G C G	A	A G A G	G G	A A A A	G	T A T A	G A G A	A A	720 720
12 13	721 721	A T A C A T A C	C A T C A T	T G T G	G A A	A T A T	A G A G	т	тт	Т	A G A G	G	T T T T	гт гт	C C	G C G C	T T	C A C A	ι Τ ι Τ	T T	ТА	АТ АТ	T T	T G T G	C C	Т А Т А	G	Т	C A C A	T T	G G G G	A	A G A G	G G	T A T A	Т	тт	A A A A	T T	780 780
12 13	781 781	A T T A A T T A	A T A A T A	C A C A	G A	A T A T	т т т т	G A G A	A A A A	T T	G T G T	A A	T A T A	A A A A	A A	ТА ТА	T T	A A A A	A	T T	ТА	A C A C	Т	а т а т	T T	Т А Т А	A	A ·	T G T G	A A	A A A A	T T	тт	с с	G A G A	Т	C A C A	T T T T	C C	840 840
12 13	841 841	A T A T A T A T	A T T A T T	тт	т и ти	A C A C	A C A C	C /	A G A G	C .	A T A T	A A	T A T A	ч т ч т	A A	G T G T	T T	C A C A	ι Τ ι Τ	T T	A ( A (	C A C A	T T	тт	T T	C A C A	A	A ·	ГА ГА	C C	т т т т	G G	СТ СТ	с с	T A T A	A	тт	T C . T C .	A A	900 900
12 13	901 901	C G A A C G A A	G C G G C G	A A A A	ТА	A G A G	A T A T	A A	T G T G	Т	A T A T	A A	тт тт	ГА ГА	A A	G G G G	A A	A A A A		A A	Т	G T G T	T T	ТА ТА	G G	G A G A	T T	A ·	T G T G	A A	A C A C	T T	C A C A	A	A G A G	Т	тт	T A . T A .	A A	960 960
12 13	961 961	ттсс ттсс	T A G T A G	G A G A	. A (	G C G C	T A T A	ТА	A T A T	T T	С Т С Т	G G	T A T A	ч т ч т	Т	A A A A	G G	то	G C G C	T T	ти ти	А Т А Т	T T	ст ст	G G	ТА	T	ти	A A A A	T T	T A T A	T T	тт	с С	С Т С Т	C C	A A C A	A T A T	T T	1020 1020
12 13	1021 1021	T G A A T G A A	A A A A A A	A A A T	T T	T G T G	G A G A	G (	G A G A	A A	A T A T	G G	тт тт	гт гт	T T	T G T G	A A	G A G A	. Т . Т	A A	A ( A (	с с с с	G / G /	а т а т	A A	A A A A	T T	A A A A	АТ АТ	A A	T A T A	C /	A G A G	G G	C T C T	C /	A C A C	T G T G	T T	1080 1080
12 13	1081 1081	T C A G T C A G	G T T G T T	A G A G	6 A <i>4</i> 6 A <i>4</i>	A T A T	С Т С Т	G A G A	A A A A	T T	T A T A	T T		G C	T T	т т т т	C C	тс	C A C A	A A	A A A A	A T A T	A	а т а т	T T	A C A C	G G G	A ·	ГА ГА	A A	T A T A	AA	ТА ТА	C C	T A T A	Т	тт тт	C C C	A A	1140 1140
12 13	1141 1141	G A A T G A A T	Т А А Т А А	T A T A	T	T A T A	A T A T	G (	C G C G	A A	T A T A	T T	TT TT	гт гт	T T	T G T G	A A	A C A C	C G C G	A A	A ·	T G T G	C /	A C A C	T T	ТА	T	A 0 A 0	G A G A	A A	A G A G	G G	Т А Т А	T	т т т т	G /	A A A A	T G T	A A	1200 1200
12 13	1201 1201	A A C T A A C T	Т Т А Т Т А	тт	т. т.	C A C A	A A A A	A A A A	A G A G	G . G .	A T A T	A A	тс	ст	Т	A A A A	T T	Т А Т А	A	G G	Т	C G C G	G G	тт	A A	T T T T	T T	T T	ГА ГА	T T	т т т т	G G	Т А Т А	G	A T A T	Т	G G G G	C T C T	G G	1260 1260
12 13	1261 1261	A C A A A C A A	A T A A C A	СТ	т	C A C A	T A T A	ТА	A T A T	G G	T A T A	T T	T A T A	тт	G d	A C A C	A A	A A A A	ι Τ ι Τ	T T	A A	тт	T T	ст ст	G G	C A C A	G	A A	A A A A	A A	A A A A	A A	G T G T	T T	A T A T	T C	С Т С Т	A G A G	A A	1320 1320
12 13	1321 1321	G A A T G A A T	G C A G C A	тс	A C	C A C A	A T A T	A	т с т с	C .	A A A A	A A	тс	C A C A	T T	G C G C	A A	Т А Т А	A	C C	ТА	A A A A	G / G /	A G A G	A A	тс	ст	C /	A A A A	T T	с с с с	A A	G A G A	T	T A T A	T T	т с т с	C A C A	G G	1380 1380
12 13	1381 1381	A G G A A G G A	A T G A T G	A A A A	. A 4 . A 4	A G A G	т с т с	т ( т (	с с с с	T T	G T G T	A A	C A C A	ч т ч т	G d	A A A A	T T	A C A C	3 A 3 A	T T	G # G #	A A A A	A	т с т с	G G	A T A T	c c	G G	ГА ГА	T T	C A C A	A	тт	A	т т т т	C	G A G A	T T T T	G G	1440 1440
12 13	1441 1441	А Т А Т А Т А Т	G G G G G G	тт	T T	T T T T	T A T A	A	ТА ТА	T T	T A T A	A A	A C A C	G A G A	A A	A G A G	A A	то	G A G A	A A	A A	тт	T T	т с т с	T T	A A A A	T	ти	ат ат	T T	т с т с	A A	C A C A	Т	A T A T	A	G T G T	T A T A	T T	1500 1500
12 13	1501 1501	T A A C T A A C	G T T G T T	G A G A	. A (	C G C G	С Т С Т	ТА	A A A A	T T	G A G A	A A	G C G C	G A G A	T T	т т т т	G G	A A A A		A A	Т	G A G A	A	а т а т	A A	T T T T	T T	G / G /	A G A G	A A	G A G A	G G	G T G T	G G	т т т т	Т	C A C A	T T T T	T T	1560 1560
12 13	1561 1561	A A A A A A A A	G A A G A A	A A A A	. A 4 . A 4	A T A T	с с с с	AA	A C A C	G G	T A T A	C C	ст ст	гт гт	C .	A A A A	G G		ЭТ ЭТ	G G	T ·	т с т с	Т	а т а т	T T	G 1 G 1	G	G / G /	A A A A	G G	T A T A	C C	G A G A	A	G T G T	A	ст ст	T C T C	C C	1620 1620
12 13	1621 1621	А Т А Т А Т А Т	ТС G ТС G	A T A T	A	с с с с	A T A T	A	тт	C .	A A A A	A A	A A A A	ч т ч т	T T	C G C G	A A	T I T I	гт гт	C C	c ·	тт	G / G /	A A A A	A A	A T A T	A A	ти	A C A C	A A	T G T G	A	A A A A	Т	T A T A	Т	C A C A	A T A T	C C	1680 1680
12 13	1681 1681	T G A A T G A A	Т Т А Т Т А	т т т т	AA	A T A T	т с т с	ти	A A A A	T T	т с т с	C C	T A T A	A A	A A	A A A A	T T	A C A C	ЭТ ЭТ	G G	A A A A	АТ АТ	A	А Т А Т	A A	A T A T	A A	A 0 A 0	ЭТ ЭТ	G G	G T G T	T T	СТ	T T	т т т т	T T	- A T A	A A A A	T T	1739 1740
12 13	1740 1741	G A G T G A G T	ттт ттт	A A A A	ТА	A T A T	A G A G	A	G A G A	A A	тт	T T	A A A A	A A	A A	т с т с	T T	A 1 A 1	r c r c	T T	C I	C A C A	A	а т а т	T T	T T T T	G G	ТИ	A A A A	T T	G G G G	T T	С Т С 1	Г А	G 17 G 17	92 93				

## Supporting Information; Figure 2A

	<b>N-terminus</b> MNMSIRLALLGFVVLSTQTVFSAGQGATPWENSQLAESFISRFLRFIGQSGA SPNQLDDMSSIGDTLKTAIEKMAQSRKSSKSKLQALNMAFASSMAEIAVAEQ GGLSLEAKTNAIASALSAAFLETTGYVNQQFVNEIKTLIFMIAQASSNEIS
	<b>Repeat I</b> GPGYGPGAGQQGPGSQGPGSGGQQGPGGQGPYGPS (A) <sub>8</sub>
$\bigcirc$	Repeat II GPGYGPGAGQQGPGSGQQGGQGSGQQGPGGAGQGGPRGQGPYGPG (A) $_9$
1	<b>Repeat III</b> GGYGPGAGQQGPGSQGPGSGGQQGPGSQGPYGPS (A) <sub>8</sub>
0	<b>Other silk-like sequences</b> GSAAAAGGSSGGGGGGGGQGGYGQGAYASASAAAAYG SAPQGTGGPASQGPSQQGPVSQPSYGPSATVAVTAV
	GGQGSGQQRPGGAGQGGLGPYGPG (A) 9
	${\tt GPGYGPGAGQQGPGSQAPV} ({\tt A})_5$
	<b>C-terminus</b> SRLSSPQASSRVSSAVSTLVSSGPTNPASLSNAISSVVSQ VSASNPGLSGCDVLVQALLEIVSALVHILGSSSIGQINYA ASSQYAQMV

Figure 2. Arrangement of N-terminal, repetitive, and C-terminal domains within the major ampullate silk protein. (A) At.MaSp2, (B) Lg.MaSp1, and (C) Nim.MaSp2. Silk proteins are divided into non-repetitive N-terminal regions (blue), repetitive units, and C-terminal regions (red). Types of repetitive units (short blocks ending with  $(A)_n$ ) are designated as repeat I, repeat II, and repeat III. The consensus sequence of each of these units was generated using the MacVector software (Accelerys, San Diego, CA). Silk-like sequences represent regions that did not have any significant similarity to any other sequences, thus considered as non-repetitive units. Dotted lines represent missing protein sequences.

### Supporting Information; Figure 2B



#### N-terminus

MTAPSQSGAFSSDQIDDMSVISNTLMAAMDNMGGRITPTK LQALDMAFASSVAEIAAVEGQNIGVTTNAISDALTSAFYQ TTGVVNNKFISEIRSLINMFAQASANDVY



Repeat I PGQGGYGQGGFGQGASG (A) 7



X=A,E,L,P

### Other silk-like sequences



SSAGASSGAKGYGSVSSSVSAAQRSGMI YQAPAKAQVTFSMTRQQQQGGSG (A) 6







GPGQVYYGPQSVAAPAA

### C-terminus



SALAAPATSARISSHALTLLSNGPTNPASISNVISNAVSQI SSSNPGYSSCDILVQALLELVTALLTIIGSSNVNDINYGSS GQYAQMVSQSVQNVFG

## Supporting Information; Figure 2C

	<i>N-terminus</i>	$\bigcirc$
	MSWSTLALAIIAVLSTQCIFIAGQANTPWSDTATADAFIQNFI AVSGSGAFTPDQLDDMSTVGDTIMSAMDKMARSNKSSKSKLQA	L C
	NMAFASSMAEIAAVEQGGQSMDVKTNAIANALDSAFYMTTGST QQFVNEMRSLINMLSAAAVNEVS	
		$\sim$
$\Leftrightarrow$	Repeat I GRGPGGYGPGQQGPGGPG (A) 6	$\sim$
		$\sim$
$\bigcirc$	Repeat II GPGGYGPGQQGPG (A) 7	$\sim \sim \sim$
		$\sim$
	Repeat III GRGPGGYGPGOOGPGOOGPGGSG(A)	$\bigcirc$
		$\bigcirc$
	Other silk-like sequences	$\sim \sim \sim$
$\bigcirc$	YGGGASAAAATAGSYGQGPSGYAQGSSAASAAAPSGYVPSQ TGQSGLGAAAAAAAVAPSGYGPSQQGPSGPGAAT(A)4	$\sim \sim \sim$
$\bigcirc$	GRGPEGYGPRQQGPG (A) 6	$\sim$
	GPGGYGPRQQGPGGYGPGQQGPG (A) $_7$	$\sim$
	GSEGYGPGQQGPRGPG (A) $_4$	$\sim$
$\overline{\mathbf{O}}$	GPGGYGPGGYGPGQQGPGGPG (A) 7	$\sim$
$\overline{\bigcirc}$	GPGGYGPGQQRPGQQGPGGPG (A) 6	$\mathbb{O}$
	GRGPGGSGPGGYGPGQQGPG (A) 8	$\sim$
		$\sim$

 $\diamondsuit$ 

 $\langle \rangle$ 

# Supporting Information; Figure 3

Lg.MaSp1 At.MaSp2	GAATTTCCAGGTGTATTTGTCAACCTGTATTTTAT <b>CACG</b> GAAAAC CAGGTTACGTATGTAGAAAT-AGCTGACAGTGATG <b>CACG</b> CAAATG	
Nim.MaSp2	GAGGTCAGTTTTGT-TGAATCAGCTCAT-TTTTAACACGCACCAA	
La MaSpl		9
At MaSn2	CATCA TATAAAACACACACAAACTIICAAAAAC	9
Nim Magn2		0
MIM.Ma5pz	$\Lambda(2) \qquad \qquad \Lambda(3)$	9
Lg.MaSp1	GGATTTCCCAATGATTAAAATGAATTTGTCAACTCGACTTGCTCT	54
At MaSp2	AGTTCCATTCTCAAAGaAA <b>ATG</b> AATTGGTCAATTCGTCTTGCTCT	54
Nim MaSp2	AGATCAGTTCTCAAAACaAA <b>ATC</b> AGTTGGTCCACTCTAGCTTT	51
interno pr		51
Lg.MaSp1	ATCATTCCTCGTG-TGCTTTGCACTCAAGGTCTGTATGTTCTGGG	98
At.MaSp2	TTTAGGTTTCGTGGTGCTCAGCACCCAAACTGTATTTTCTGCTGG	99
Nim.MaSp2	AGCGATTATCGCGGTGCTTAGCACCCAGTGCATTTTTATTGCAGG	96
1		
Lg.MaSp1	ACAAGCAAACACTCCATGGTCTAGTAAACAAAATGCTGACGCTTT	143
At.MaSp2	CCAGG GTGCAACTCCATGGGAGAACTCGCAACTGGCGGAGAGCTT	144
Nim.MaSp2	ACAAGCAAACACACCATGGAGCGACACTGCCACAGCAGATGCTTT	141
Lg.MaSpl	TATAAGTGCATTC <b>ATG</b> ACTGCTCCTTCACAAAGTGGAGCATTTTC	188
AL.Maspz		10C
MIIII.Ma5pz	CATICAGAATTICTIAGGAGCIGTITCAGGAAGTGGAGCCTITAC	100
Lq.MaSp1	ATCGGAT CAGATCGATGACATGCTGTCATCAGCAATACATTAAT	233
At.MaSp2	CCCAAAC	234
Nim.MaSp2	TCCAGAT CAACTTGATGAT ATG	231
Lg.MaSp1	GCAGCAATGGATaATATGGGAGGAAGAATTACACCCAC	272
At.MaSp2	GACTGCAATTGAA <b>a</b> AA <mark>ATG</mark> GCTCAAAGCCGAAAAAGTTCTAAATC	279
Nim.MaSp2	<b>G</b> TCAGCA <b>ATG</b> GAC <b>a</b> AA <mark>ATG</mark> GCTCGCAGTAACAAGAGCTCCAAATC	276
Ia Magni		217
At Magn2	CAMPTICCACCONTRACTOCOCCONTRACTOCOCCACCACCONTRACTOCOCCACAC	321
Nim Magna	AAAATTACAACCTCTAAACATOCCATTTACTTCCTCCATOCCAA	221
MIM.Ma5pz	AAAAIIACAAGCICIAAACAIGGCIIICGCIICAICGAIGGCAGA	771
Lq.MaSp1	AATTGCTGCTGTGGAAGGTCAAAATATAGGGGTAACTAC	356
At.MaSp2	AATTGCTGTAGCAGAGCAGGGAGGTTTAAGCTTAGAAGCAAAAAC	369
Nim.MaSp2	AATTGCAGCGGTGGAACAAGGTGGTCAG <b>a</b> GC <b>ATG</b> GATGTCAAAAAC	366
Lg.MaSp1	AAATGCAATTTCAGACGCCTTGACATCAGCTTTCTATCAAACAAC	401
At.MaSp2	CAATGCCATCGCAAGTGCCCTCAGTGCAGCCTTTTTGGAAACCAC	414
Nim.MaSp2	AAATGCAATTGCCAATGCTCTAGATTCAGCTTTTTAT <b>ATG</b> ACAAC	411
-		
Lg.MaSp1	AGGGGTAGTTAATAACAAATTTATCAGCGAAATTAGAAGTTTGAT	446
At.MaSp2	TGGCTATGTAAACCAACAGTTTGTCAATGAAATAAAAACATTAAT	459
Nim.MaSp2	TGGTTCAACAAATCAACAGTTTGTCAACGAA <b>ATG</b> AGAAGCTTAAT	456
La Magni	Ͳ <b>϶</b> ϪͲ <b>ϪͲϾ</b> ͳͲͲϾϹϪϹϪϪϾϹϾͲϹͳϾϹϪϪϪͲϾϪͲϾͲͲͲϪͲ_483	
At.MaSp1	ATTT <b>ATG</b> ATCGCTCAGGCATCATCAAATGAAATTTCT 496	
Nim.MaSp2	TaACATGTTGAGTGCAGCTGCCGTTAATGAAGTATCA 493	

Figure 3. Alignment of genomic DNA sequences. Aligned DNA ranges from upstream of the predicted transcription start site (approximately 80 nucleotides) to the downstream of the predicted transcription start site (almost 500 nucleotides). The entire N-termini coding regions are included in this alignment. Start codons of proposed translation start sites of long isoforms are shaded in yellow. Start codons of the predicted translation start sites of short isoforms are shaded in red. The remaining start codons in frame are bold and italicized. Adenines at -3 position in respect to the ATG are in a lower case and bold. Carat symbols point at the first nucleotide of the following: (1) CACG motif; (2) TATA motif; (3) predicted transcription start site that was designated as +1. Boxed sequences correspond to regions used to design oligonucleotides that were utilized in the Northern blotting experiment. Colors of boxes represent numbers following species names (see Materials section): 1, red; 2, blue; 3, green; 4, pink; 5, brown; 6, orange.