

Supplementary Materials

Intramolecular homologous recombination event occurred in the spider egg case silk

gene CySp2 of Wasp Spider *Argiope bruennichi*.

Leng Han and Masao Nakagaki *

Contents

Supplementary Material 1(S1): The genomic DNA sequence of 5' end of <i>CySp2</i>	2
Supplementary Material 2(S2): Alignment result of CySp2O and CySp2N	3
Supplementary Material 3(S3): CySp2N amino acid sequence and signal peptide analysis	5
Supplementary Material 4(S4): A homology comparison between CySp1 and CySp2N for the non-repetitive N-terminal region.	7

Supplementary Material 1(S1): The genomic DNA sequence of 5' end of *CySp2*

S1a

>Sequencing result of *CySp2* by Han with the primers from 217 to 872

```
217 GGATGCCATT GCCCAGGTGA TACTCAATGC CGTTTCAAGC AACACTGGCG CCACAGCATC
277 GGCCAGAGCT CAAGCTTTAA GTACAGCGCT TGCATCTTCT CTGACAGATC TGCTCATTGC
337 AGAGTCGGCA AAAAGCAATT ACAGCAATCA GCTGTCTGAA CTAACAGGAA TCCTCTCCGA
397 CTGTTTTATC CAAACTACTG GATCGGACAA TCCAGCATTT GTGTCCAGAA TTCAATCTCT
457 CATTTCAGTG CTCTCCAGA ATGCAGATGT AAATATAATT TCCTCGGCAG GGATACCTTC
517 AGTGTCAAGG AGAAGTGGAG CTAGAGGAGG ATTCTCTAAC GCCGCAAGCC AGTCCGCAAG
577 CAGYGCTGCC TCGCAAGCAA GTGCCTCTTC ATTCGCACAA GCATCCTCTG CTTCCTTGC
637 AGCCTCAAGT TCTTTCTCCA GTGCCTTCTC TTCGGCCAAT TCCCTCTCAG CTCTCGGCAA
697 CGTTGCTTAT CAATTAGGCT TCAACGTAGC TAACACTCTC GGTATCGGCA ACGCTCCAGG
757 CCTCGGTGCT GCCTTATCTC AAGCTGTCTC TTCAGTCGGC GTTGGAGCCA GTTCCAGCAC
817 GTACGCTAAT GTCGTTTCCA ATGCAGTTGG ACAATTCTTA GCCGGTCAGG GCGTTT
```

S1b: New genomic DNA sequence for the 5' end of *CySp2*, *CySp2N*.

>genomic DNA sequence of 5'end of *CySp2* by Han

```
1 GACAGTGAGG AACCCTTTCA ATCAAGCAAC TATGATGTGG TTCACTACAG TAGCGTTCCT
61 CTTATGCCTT TTAGGAGCAG TGTCGTCACA ATCAGTTGCA GTTACC GCCG TTCCCTCAGT
121 CTTCAGCAGT CCAAATTTGG CCAGCGGTTT CCTTCAGTGT CTCACATTTG GAATCGGAAA
181 TTCTCCCGCA TTTCTACTC AAGAACAACA AGACTTGGAT GCCATTGCC AGGTGATACT
241 CAATGCCGTT TCAAGCAACA CTGGCGCCAC AGCATCGGCC AGAGCTCAAG CTTTAAGTAC
301 AGCGCTTGCA TCTTCTCTGA CAGATCTGCT CATTGCAGAG TCGGCAAAAA GCAATTACAG
361 CAATCAGCTG TCTGAACTAA CAGGAATCCT CTCCGACTGT TTTATCCAAA CTACTGGATC
421 GGACAATCCA GCATTTGTGT CCAGAAATCA ATCTCTCATT TCAGTGCTCT CCCAGAATGC
481 AGATGTAAAT ATAATTTCCCT CGGCAGGGAT ACCTTCAGTG TCAGGAAGAA GTGGAGCTAG
541 AGGAGGATTC TCTAACGCCG CAAGCCAGTC CGCAAGCAGY GCTGCCTCGC AAGCAAGTGC
601 CTCTTCATTC GCACAAGCAT CCTCTGCTTC CCTTGCAGCC TCAAGTTCTT TCTCCAGTGC
661 CTTCTCTTCG GCCAATTCCC TCTCAGCTCT CGGCAACGTT GCTTATCAAT TAGGCTTCAA
721 CGTAGCTAAC ACTCTCGGTA TCGGCAACGC TCCAGGCCTC GGTGCTGCCT TATCTCAAGC
781 TGTCTCTTCA GTCGGCGTTG GAGCCAGTTC CAGCACGTAC GCTAATGTGC TTTCCAATGC
841 AGTTGGACAA TTCTTAGCCG GTCAGGGCGT TT
```

Supplementary Material 2(S2): Alignment result of CySp20 and CySp20

S2a: Alignment result of CySp20 and CySp2N by SIM4

seq1 =CySp20 , 523 bp seq2 =CySp2n, 872 bp

1-311 (1-311) 99%;312-523 (661-872) 100%

```
seq1 1 GACAGTGAGGAACCCCTTTCAATCAAGCAACTATGATGTGGTTCACTACAG
|||
Seq2 1 GACAGTGAGGAACCCCTTTCAATCAAGCAACTATGATGTGGTTCACTACAG

51 TAGCGTTCCTCTTATGCCTTTTAGGAGCAGTGTTCGTCAACAATCAGTTGCA
|||
51 TAGCGTTCCTCTTATGCCTTTTAGGAGCAGTGTTCGTCAACAATCAGTTGCA

101 GTTACCGCCGTTCCCTCAGTCTTCAGCAGTCCAAATTTGGCCAGCGGTTT
|||
101 GTTACCGCCGTTCCCTCAGTCTTCAGCAGTCCAAATTTGGCCAGCGGTTT

151 CCTTCAGTGTCTCACATTTGGAATCGGAAATTCTCCCGCATTTCTACTC
|||
151 CCTTCAGTGTCTCACATTTGGAATCGGAAATTCTCCCGCATTTCTACTC

201 AAGAACAACAAGACTTGGATGCCATTGCCAGGTGATACTCAATGCCGTT
|||
201 AAGAACAACAAGACTTGGATGCCATTGCCAGGTGATACTCAATGCCGTT

251 TCAAGCAACACTGGCGCCACAGCATCGGCCAGAGCTCAGGCTTTAAGTAC
|||
251 TCAAGCAACACTGGCGCCACAGCATCGGCCAGAGCTCAAGCTTTAAGTAC

301 AGCGCTTGCAT CTTCTCTTCGGCCAATTCCCTCTCAGCTCT
|||<<<...<<<|||
301 AGCGCTTGCATCTT...TGCCTTCTCTTCGGCCAATTCCCTCTCAGCTCT

342 CGGCAACGTTGCTTATCAATTAGGCTTCAACGTAGCTAACACTCTCGGTA
|||
691 CGGCAACGTTGCTTATCAATTAGGCTTCAACGTAGCTAACACTCTCGGTA

392 TCGGCAACGCTCCAGGCCTCGGTGCTGCCTTATCTCAAGCTGTCTCTTCA
|||
741 TCGGCAACGCTCCAGGCCTCGGTGCTGCCTTATCTCAAGCTGTCTCTTCA

442 GTCGGCGTTGGAGCCAGTTCAGCACGTACGCTAATGTCGTTTCCAATGC
|||
791 GTCGGCGTTGGAGCCAGTTCAGCACGTACGCTAATGTCGTTTCCAATGC

492 AGTTGGACAATTCTTAGCCGGTCAGGGCGTTT
|||
841 AGTTGGACAATTCTTAGCCGGTCAGGGCGTTT
```

S2b: Alignment result of CySp2O and CySp2N by Blast N

Query =CySp2O , 523 bp

Sbjct =CySp2n, 872 bp

```
Query 1 GACAGTGAGGAACCCCTTCAATCAAGCAACTATGATGTGGTTCACTACAGTAGCGTTCCT 60
      |||
Sbjct 1 GACAGTGAGGAACCCCTTCAATCAAGCAACTATGATGTGGTTCACTACAGTAGCGTTCCT 60

Query 61 CTTATGCCTTTTAGGAGCAGTGTGTCACAATCAGTTGCAGTTACCGCCGTTCCCTCAGT 120
      |||
Sbjct 61 CTTATGCCTTTTAGGAGCAGTGTGTCACAATCAGTTGCAGTTACCGCCGTTCCCTCAGT 120

Query 121 CTTCAGCAGTCCAAATTTGGCCAGCGTTTCCTTCAGTGTCTCACATTTGGAATCGGAAA 180
      |||
Sbjct 121 CTTCAGCAGTCCAAATTTGGCCAGCGTTTCCTTCAGTGTCTCACATTTGGAATCGGAAA 180

Query 181 TTCTCCCGCATTTCCTACTCAAGAACAACAAGACTTGGATGCCATTGCCAGGTGATACT 240
      |||
Sbjct 181 TTCTCCCGCATTTCCTACTCAAGAACAACAAGACTTGGATGCCATTGCCAGGTGATACT 240

Query 241 CAATGCCGTTTCAAGCAACACTGGCGCCACAGCATCGGCCAGAGCTCAGGCTTAAAGTAC 300
      |||
Sbjct 241 CAATGCCGTTTCAAGCAACACTGGCGCCACAGCATCGGCCAGAGCTCAAGCTTAAAGTAC 300

Query 301 AGCGCTTGCATCTTCTCT 318
      |||
Sbjct 301 AGCGCTTGCATCTTCTCT 318

Query 312 CTTCTCTTCGGCCAATTCCTCTCAGCTCTCGGCAACGTTGCTTATCAATTAGGCTTCAA 371
      |||
Sbjct 661 CTTCTCTTCGGCCAATTCCTCTCAGCTCTCGGCAACGTTGCTTATCAATTAGGCTTCAA 720

Query 372 CGTAGCTAACACTCTCGGTATCGGCAACGCTCCAGGCCTCGGTGCTGCCTTATCTCAAGC 431
      |||
Sbjct 721 CGTAGCTAACACTCTCGGTATCGGCAACGCTCCAGGCCTCGGTGCTGCCTTATCTCAAGC 780

Query 432 TGTCTCTTCAGTCGGCGTTGGAGCCAGTCCAGCACGTACGCTAATGTCGTTTCCAATGC 491
      |||
Sbjct 781 TGTCTCTTCAGTCGGCGTTGGAGCCAGTCCAGCACGTACGCTAATGTCGTTTCCAATGC 840

Query 492 AGTTGGACAATTCTTAGCCGGTCAGGGCGTTT 523
      |||
Sbjct 841 AGTTGGACAATTCTTAGCCGGTCAGGGCGTTT 872
```

The 7 bp direct repeats are highlighted in yellow.

Supplementary Material 3(S3): CySp2N amino acid sequence and signal peptide analysis.

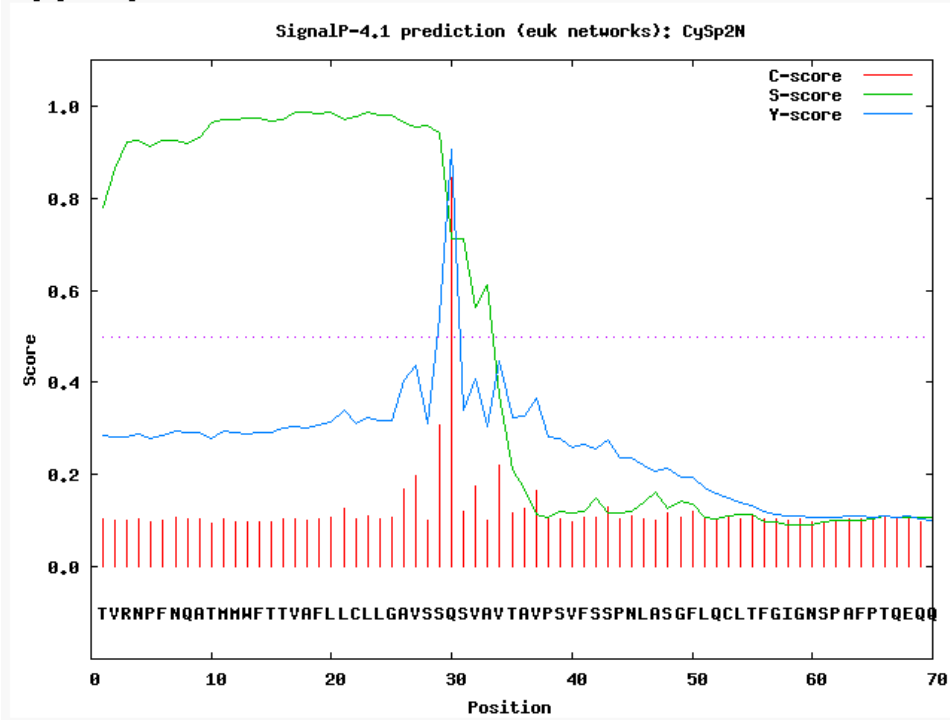
S3a. Amino acid residue sequence was deduced from CySp2N using the 2nd ORF. The initial amino acid sequence (Met) was translated using the start codon highlighted in red.

>CySp2N.

```
1      TVRNPFNQAT MMWFTTVAFLL
21     LCLLGAVSSQ SVAVTAVPSV
41     FSSPNLASGF LQCLTFGIGN
61     SPAFPTQEQQ DLDAIAQVIL
81     NAVSSNTGAT ASARAQALST
101    ALASSLTDLL IAESASSNYS
121    NQLSELTGIL SDCFIQTTGS
141    DNPAFVSRIQ SLISVLSQNA
161    DVNIISSAGI PSVSGRSGAR
181    GGFSNAASQS ASSAASQASA
201    SSFAQASSAS LAASSSFSSA
221    FSSANLSAL  GNVAYQLGFN
241    VANTLGIGNA PGLGAALSQA
261    VSSVGVGASS STYANVVSNA
281    VGQFLAGQGV
```

S3b. Signal peptide analysis for CySp2N using signalP 4.1

```
# SignalP-4.1 euk predictions
>CySp2N by Han
```



# Measure	Position	Value	Cutoff	signal peptide?
max. C	30	0.846		
max. Y	30	0.905		
max. S	17	0.988		
mean S	1-29	0.950		
D	1-29	0.930	0.450	YES

Name=CySp2N SP='YES' Cleavage site between pos. 29 and 30: VSS-QS D=0.930 D-cutoff=0.450 Networks=SignalP-noTM

Supplementary Material 4(S4): A homology comparison between CySp1 and CySp2N

for the non-repetitive N-terminal region.

S4a . The alignment of the nucleotide sequence of N-termini of CySp1 and CySp2N.

```
Query : CySp1 non-repetitive region DNA sequence, Length 537.
Subject: CySp2N non-repetitive region DNA sequence, Length 531.
Program BLASTN 2.2.27+ Citation Score = 794 bits (880), Expect = 0.0
Identities = 497/534 (93%), Gaps = 3/534 (1%)

Strand=Plus/Plus

Query 1   ATGGTTTGGCTCACTAGCATAGCGTTCCTCGTAGGCTTTCTGGGAGCAGTGTCGTCTCAG 60
        ||| | ||| |||||  ||||||||| || || || | ||||||||| |||
Sbjct 1   ATGATGTGGTTCACACAGTAGCGTTCCTCTTATGCCTTTTAGGAGCAGTGTCGTACAA 60

Query 61  TCAGTTGCAGTTACCGCCGTTCCCTCAGTCTTCAGCAGTCCAAATTTGGCCAGCGGCTTC 120
        ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| |||
Sbjct 61  TCAGTTGCAGTTACCGCCGTTCCCTCAGTCTTCAGCAGTCCAAATTTGGCCAGCGGTTTC 120

Query 121 CTTCAATGTCTCACATTTGGAATCGGAAATTCCCCGCATTTCTACTCAAGAACAACAA 180
        ||||| ||||||||| ||||||||| ||||||||| ||||||||| |||
Sbjct 121 CTTCAAGTGTCTCACATTTGGAATCGGAAATTCTCCCGCATTTCTACTCAAGAACAACAA 180

Query 181 GACTTGGATGCCATTGCTCAGGTGATACTCAATGCTGTTTCAAGCAACACTGGCGCTACA 240
        ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| |||
Sbjct 181 GACTTGGATGCCATTGCCAGGTGATACTCAATGCCGTTTCAAGCAACACTGGCGCCACA 240

Query 241 GCATCGGCCAGAGCTCAAGCTTTAAGTACAGCGCTTGCATCTTCTCTGACAGATCTGCTC 300
        ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| |||
Sbjct 241 GCATCGGCCAGAGCTCAAGCTTTAAGTACAGCGCTTGCATCTTCTCTGACAGATCTGCTC 300

Query 301 ATTGCAGAGTCGGCAGAAAGCAATTACAGCAATCAGTTGCTGAACTAACAGGAATTCTC 360
        ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| |||
Sbjct 301 ATTGCAGAGTCGGCAAAAAGCAATTACAGCAATCAGCTGTCTGAACTAACAGGAATCCTC 360

Query 361 TCCGACTGTTTTATCCAAACTACTGGATCGGACAACCCAGCATTGTGTCCAGAATTCAA 420
        ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| |||
Sbjct 361 TCCGACTGTTTTATCCAAACTACTGGATCGGACAATCCAGCATTGTGTCCAGAATTCAA 420

Query 421 TCTCTCATTTTCAGTGCTTTCCAGAAATGCAGATACAAATATAATTTCTCAGCAGGGATA 480
        ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| |||
Sbjct 421 TCTCTCATTTTCAGTGCTCTCCAGAAATGCAGATGTAAATATAATTTCTCAGCAGGGATA 480

Query 481 CCTTCAGTGTGAGGAAGAAGGGGAGCTGGTGGATTAGGATTCGATAACACCGCA 534
        ||||||||| ||||||||| ||| ||| ||| ||| |||
Sbjct 481 CCTTCAGTGTGAGGAAGAAGTGGAGCTAGAGG---AGGATTCTCTAACGCCGCA 531
```

s4b. The non-repetitive N-terminal sequence alignment generated for CySp1 and CySp2N.

Query: CySp1 N-terminal non-repetitive region, (179 letters)

Sbjct: CySp2N N-terminal non-repetitive region, Length = 177

Score = 306 bits (784), Expect = 1e-088

Identities = 164/178 (92%), Positives = 169/178 (94%), Gaps = 1/178 (0%)

Query: 1 MVWLTSIAFLVGFGLGAVSSQSVAVTAVPSVFSSPNLASGFLQCLTFGIGNSPAFPTQEQQ 60

M+W T++AFL+ LGAVSSQSVAVTAVPSVFSSPNLASGFLQCLTFGIGNSPAFPTQEQQ

Sbjct: 1 MMWF'TTVAFLLCLLGAVSSQSVAVTAVPSVFSSPNLASGFLQCLTFGIGNSPAFPTQEQQ 60

Query: 61 DLDAIAQVILNAVSSNTGATASARAQALSTALASSLDLLIAESAESNYSNQLSELTGIL 120

DLDAIAQVILNAVSSNTGATASARAQALSTALASSLDLLIAESA+SNYSNQLSELTGIL

Sbjct: 61 DLDAIAQVILNAVSSNTGATASARAQALSTALASSLDLLIAESAKSNYSNQLSELTGIL 120

Query: 121 SDCFIQTTGSDNPAFVSRIQSLISVLSQNADTNIISSAGIPSVSGRRGAGGLGFDNTA 178

SDCFIQTTGSDNPAFVSRIQSLISVLSQNAD NIISSAGIPSVSGR GA G GF N A

Sbjct: 121 SDCFIQTTGSDNPAFVSRIQSLISVLSQNADVNIISSAGIPSVSGRSGARG-GFSNAA 177