

# Inter-specific sequence conservation and intra-individual sequence variation in a spider silk gene

Pei-Ling Tai, Guang-Yuh Hwang, I.-Min Tso\*

Department of Life Science, Tunghai University, Taichung 407, Taiwan

## Abstract

Currently, studies on major ampullate spidroin 1 (*MaSp1*) genes of non-orb weaving spiders are few, and it is not clear whether genes of these organisms exhibit the same characteristics as those of orb-weavers. In addition, many studies have proposed that *MaSp1* might be a single gene with allelic variants, but supporting evidence is still lacking. In this study, we compared partial DNA and amino acid sequences of *MaSp1* cloned from different spider guilds. We also cloned partial *MaSp1* sequences from genomic DNA and cDNA of the same individuals of spiders using the same primer combination to see if different molecular forms existed. In the repetitive region of partial *MaSp1* sequences obtained, GGX, GA and poly-A motifs were present in all Araneomorphae and Mygalomorphae species examined. An extreme similarity in *MaSp1* non-repetitive portions was found in sequences of ecribellate, cribellate and Mygalomorphae web-builders and such a result suggested that this sequence might exhibit an important function. A comparison of sequences amplified from the same individual showed that substitutions in amino acids occurred in both repetitive and non-repetitive regions, with a much higher variation in the former. These results suggest that the *MaSp1* of Araneomorphae spiders exhibits several forms in an individual spider and it might be either a multiple gene or a single gene with a multiple exon/intron organization.

© 2004 Elsevier B.V. All rights reserved.

**Keywords:** Spider silk; Dragline; Major ampullate silk; *MaSp1* gene

## 1. Introduction

Currently, most relevant studies on spider silks focus on orb web spiders [30,32]. However, orb web spiders only comprise a small portion of the order Araneae and there are many other spider guilds [6,9]. According to web structure and prey-catching behavior, spiders could be categorized into different guilds such as orb web spiders, space web spiders (three-dimensional), wandering sheet weavers and wandering spiders [29]. Besides, according to the presence of a spinning structure called cribellum, Araneomorphae can be divided into cribellataes and ecribellataes [17]. Cribellum is homologous to the anterior median spinneret of the suborder Mesothelae, and is considered a primitive character [17,21,25]. In many spider guilds, both cribellate and ecribellate species coexist and their silk physical properties and web

functions are quite diverse [3,20–23]. Recent studies on a limited number of taxa suggest that dragline silks have evolved into a diverse array of molecular structures and physical properties. Gatesy et al. [10] found that the poly-A, GA and GGX motifs repeatedly appeared in the major ampullate spidroin 1 (*MaSp1*) gene of nine orb web spiders. They also found that some of the motifs (poly-A and GA) were present in the dragline silk of the primitive suborder Mygalomorphae. So far, studies on the dragline silk genes of the space web spiders and Mygalomorphae are few, and there is no study comparing the dragline silk genes of cribellate and ecribellate members of the same spider guilds. In this study we first compared the *MaSp1*-like gene sequences of Mygalomorphae and orb-weavers and space web builders of Araneomorphae. Within each guild we also compared the sequences of the cribellate and ecribellate members.

So far, the molecular architecture and organization of *MaSp1* remain unclear. Craig et al. [4] found the amino acid composition of dragline silk differed when *Argiope*

\* Corresponding author. Tel.: +886 4 2350 1656; fax: +886 4 2359 0296.  
E-mail address: spider@mail.thu.edu.tw (I.-Min Tso).

*argentata* fed on different types of prey. They proposed that such a result might be generated by differential expression of a multiple gene system or a different way of post-transcription editing. Beckwitt et al. [2] found that different forms of partial *MaSp1* sequences could be amplified from cDNA and genomic DNA of *Nephila clavipes*. They suggested that different sequences obtained from different individual spiders represented allelic variants of *MaSp1*, rather than different members of a gene family. Recently, a spider silk gene has been found to exhibit a multiple exon and intron organization. Hayashi and Lewis [13] reported that the flagelliform silk of orb-weaving spiders is expressed by a single locus (the *Flag* gene) with multiple exons and introns. Whether *MaSp1* and *Flag* have similar molecular architecture is not clear because the full length sequence of the former is still unknown [12]. However, if different forms of *MaSp1* sequences are obtained from the genomic DNA of the same individual, then this gene may be a multiple gene system or a single locus with multiple exons. Therefore, in this study we amplified, cloned, sequenced and compared the *MaSp1* partial sequences from the same individual spider for several different individuals to understand the organization features of this silk gene.

## 2. Materials and methods

### 2.1. Taxa examined

*MaSp1*-like genes of species from two suborders of Araneae were examined in this study. In the suborder Araneomorphae, we examined two species of orb-weavers and two of space web builders. All these species inhabit forest understory and the specimens used in this study were collected from low elevation forests in central Taiwan. For the cribellate species we used *N. pilipes* and *Cyrtophora moluccensis*. *N. pilipes* (Tetragnathidae) is the largest orb-weaving spider in Taiwan and it builds a vertical orb web at the edge of forests. The diameter of the orb web can reach 1 m and the height is about 2 m above ground vegetation [14,28]. *C. moluccensis* (Araneidae) constructs a space web composed of a horizontal orb and vertical tangle lines above and below the orb. The space webs constructed by *C. moluccensis* may exist in the field for more than one month [18]. For cribellate spiders we used *Octonoba varians* and *Psechrus sinensis*. *O. varians* (Uloboridae) is a small orb weaving spider and it builds a horizontal orb (diameter about 10 cm) a few cm about ground in the forest under story [31]. *P. sinensis* (Psechridae) builds a funnel-web composed of a tube retreat, a horizontal tangle web and vertical supporting lines. The webs built by this genus have been reported to last in the field for weeks [7]. For the suborder Mygalomorphae, we used *Macrothele holsti* (Hexathelidae), which also inhabits low elevation forest under story. *M. holsti* builds a funnel-web which structure is similar to that of *P. sinensis* and the web may last for months (P.L. Tai personal observation).

### 2.2. Genomic DNA preparation and sequence amplification

Genomic DNA was extracted from the muscle of the cephalothorax or legs using PUREGENE DNA isolation kit (Gentra Systems Inc., Minneapolis, MN, U.S.A.). We used PCR primers reported by Beckwitt et al. [2], which were designed from the sequences reported in Xu and Lewis [33] and Beckwitt and Arcidiacono [1]. Most sequences of this study were amplified from the primer combinations PALA: 5'-GCGCCATCCATGGCAGCAGCAGCAGCAGCT-3' with S1R: 5'-GGCGAATTCACCTAGGGCTTGATAAACTGATTGAC-3' or S1L: 5'-CCCGGATCCGGAGGTGCCGGACAAGGAGGATATGGAGGT-3', with some from the combinations ABR: 5'-GGGAAGCTTGTGCGGCTGGAGTAGTAGGTCCACTA-3' with SALA: 5'-GCNGCNGCNGC-3'. The conditions of PCR were as follows: the denaturing temperature was initially 94 °C for 5 min. The reactions were subjected to 35 cycles of the following temperature regimes: denaturing at 94 °C for 1 min, annealing at 50–60 °C for 3 min, and elongation at 72 °C for 5 min. The final extension step was increased by 10 min to insure a full-length double-stranded product. After these reactions the samples were held at 4 °C awaiting further treatment.

### 2.3. cDNA preparation and sequence amplification

In this part of the study we used *N. pilipes* and *C. moluccensis*, which were large spiders and thus easier to handle. Total RNA preparations followed the protocol by Sambrook et al. [26]. Reverse transcription polymerase chain reaction (RT-PCR) was performed to obtain cDNA of dragline silk gene. The primers of RT-PCR were PALA, S1R and oligo-dT. Conditions of the RT-PCR were initially 50 °C for one hour, then the denaturing temperature was 94 °C for 5 min. The reactions were subjected to 35 cycles of the following temperature regimes: denaturing at 94 °C for 30 s, annealing at 52 °C for 30 s, and elongation at 68 °C for 1 min. The final extension step was 68 °C for 2 min.

### 2.4. Cloning, sequencing and sequence analysis

PCR and RT-PCR products were cloned into Invitrogen pCRII-TOPO cloning vector and sequenced. The liquid culture products were sequenced using the BigDye terminator cycle sequencing kit and analyzed on an ABI 3100 or 3700 automated DNA sequencer (Applied Biosystems, Foster City, CA). Chromatograms and contiguous alignments were edited using Seqman version 4.00 (DNASTar Inc., Madison, U.S.A.). DNA and amino acid sequences were analyzed using the DNASTAR computer programs and the SeqWeb version 2 Web-based Sequences Analysis program in GCG (Genetics Computer Group) package.

### 3. Results

#### 3.1. DNA sequences amplified from genomic DNA

Most of the sequences amplified from genomic DNA were obtained from the primer combinations S1L-S1R and S1R-S1R (Table 1). Nine sequences were obtained from three *N. pilipes* individuals, with lengths ranging from 298 to 505 bp. Three sequences were obtained from three *O. varians* individuals, with lengths ranging from 298 to 346 bp. Four sequences were obtained from three *C. moluccensis* individuals, with lengths ranging from 346 to 381 bp. There were four sequences amplified from two *P. sinensis* individuals, with lengths ranging from 298 to 346 bp. There was only one sequence amplified from *M. holsti* and the length was 357 bp. The length of the aforementioned DNA sequences did not include the primers used. Among the repetitive region of all the sequences amplified, there were very few prolines. Since proline is an important and characteristic component of *MaSp2* [15], all the sequences obtained in this study should be partial sequence of the *MaSp1* gene.

#### 3.2. DNA sequences amplified from cDNA

Ten sequences were amplified from two *N. pilipes* individuals, with lengths ranging from 343 to 819 bp (Table 2). Most of the sequences were amplified from the primer combinations PALA-PALA and PALA-S1R and with lengths greater than 500 bp. Four sequences were obtained from two individuals of *C. moluccensis* using the primer combination S1R-S1R (Table 2). The length of all these sequences was 346 bp. The length of the aforementioned DNA sequences did not include the primers used.

#### 3.3. Non-repetitive C-terminal region of sequences amplified

All the sequences amplified from genomic DNA contained the non-repetitive C-terminal region (Figs. 1–3). There were about 90 amino acids in the non-repetitive C-terminal region, which correspond to the amino acids 648 to 747 in published *N. clavipes MaSp1* [33] (Fig. 1). The non-repetitive C-terminal amino acid sequences obtained from five species were very similar to each other, but with some substitutions between different species or individuals (Figs. 1–3). There were 30 variable sites in which substitutions occurred in the form of glycine switching to serine, histidine to glutamine, leucine to proline, and proline to serine. In all the variable sites switching only occurred between two types of amino acids.

#### 3.4. Repetitive region of sequences amplified

Except those of *N. pilipes*, sequences of the other four species only contained a very short repetitive region with a length of about 30 amino acids (Figs. 2 and 3). The longest

repetitive-region-containing sequences were amplified from cDNA of *N. pilipes* using the primer combination PALA-PALA. The lengths of these three sequences ranged from 242 to 273 amino acids, containing no non-repetitive C-terminal region (Fig. 4). GGX, GA and poly-A motifs repeatedly appeared in the repetitive region of sequences amplified from all species. Although sequences amplified from species other than *N. pilipes* had a very short repeat region, they all contained these three motifs (Figs. 2 and 3). In *N. pilipes*, the repeat unit was usually composed of two or three GGX motifs followed by GA and poly-A motifs (Fig. 4). In the three longest sequences amplified from cDNA of *N. pilipes*, there was a continuous region with a length of 30 peptides which did not contain any repeat motif (Fig. 4).

#### 3.5. Inter- and intra-individual variation of sequences amplified

When sequences amplified from Araneomorphae species were aligned and compared, differences in those amplified from different and same individuals were found. Among these sequences the repetitive region exhibited numerous differences. In *N. pilipes*, NP02-PS02 and NP02-PS08 were two sequences amplified from the same individual by the same primer combination. The non-repetitive region of them was identical but the repetitive region was very different (Fig. 1). For example, compared with NP02-PS02 in the aligned sites 612 to 618, those in NP02-PS08 did not have one GGQ and two GA motifs (Fig. 1). Moreover, in the same region there were four deletions in NP02-PS08. The same phenomenon also occurred in the two sequences amplified from the same *N. pilipes* individual NP11. Compared with NP11-SS08, in the aligned sites 624 to 631, there were eight deletions in the sequence NP11-SS01. The repetitive region of sequences amplified from cDNA of *N. pilipes* also showed intra-individual variations. Among the five sequences obtained, four were identical. However, the other sequence (NP072-RTPS01) in the align sites 596 to 600 had additional GA and GGA motifs (Fig. 1). Also, in aligned sites 559 to 567 the peptide of NP072-RTPS01 was also different from those of the other four sequences. In the non-repetitive C-terminal region, there were also intra-individual differences in sequences of *N. pilipes*, *C. moluccensis*, and *P. sinensis*. For example, in the aligned site 687 in *N. pilipes*, NP02-PS02 has leucine and NP02-PS08 has valine (Fig. 1). In the aligned site 652 in the two sequences obtained from the same individual of *C. moluccensis*, CM05-PS09 has arginine and CM05-PS11 has histidine (Fig. 2). In the three sequences amplified from the same individual of *P. sinensis*, PS02-SS03 and PS02-SS11 had leucine and PS02-SS03 had proline in aligned site 653 (Fig. 3).

When the DNA sequences of the partial *MaSp1*-like gene of four Araneomorphae species were compared, the inter- and intra-individual variations were even greater. In *N. pilipes*, the align sites 1486 to 1885 in the repetitive region were very different from the corresponding region of *N. clavipes MaSp1*.

Table 1

Primers used, nucleotide length and accession numbers of *MaSp1*-like sequences amplified from genomic DNA extracted from different spider species

| Species                       | Number     | Primers  | Length | Accession number |
|-------------------------------|------------|----------|--------|------------------|
| <i>Nephila pilipes</i>        | NP02-PS02  | PALA-S1R | 502    | AY666049         |
|                               | NP02-PS08  | PALA-S1R | 478    | AY666050         |
|                               | NP11-SS01  | S1L-S1R  | 397    | AY666051         |
|                               | NP11-SS05  | S1L-S1R  | 325    | AY666052         |
|                               | NP11-SS08  | S1R-S1R  | 463    | AY666053         |
|                               | NP13-SS01  | S1R-S1R  | 355    | AY666054         |
|                               | NP13-SS04  | S1L-S1R  | 298    | AY666055         |
|                               | NP13-SS07  | S1L-S1R  | 505    | AY666048         |
| <i>Octonoba varians</i>       | OV01-SS06  | S1L-S1R  | 319    | AY666057         |
|                               | OV02-SS01  | S1L-S1R  | 298    | AY666058         |
|                               | OV10-SS16  | S1L-S1R  | 298    | AY666059         |
| <i>Cyrtophora moluccensis</i> | CM06-SSA04 | SALA-S1R | 381    | AY666060         |
|                               | CM05-PS09  | S1R-S1R  | 346    | AY666061         |
|                               | CM05-PS11  | S1R-S1R  | 346    | AY666062         |
|                               | CM07-SS01  | S1R-S1R  | 346    | AY666063         |
| <i>Psechrus sinensis</i>      | PS01-PS02  | PALA-S1R | 298    | AY666065         |
|                               | PS02-SS03  | S1R-S1R  | 346    | AY666064         |
|                               | PS02-SS06  | S1R-S1R  | 346    | AY666066         |
|                               | PS02-SS11  | S1R-S1R  | 346    | AY666067         |
| <i>Macrothele holsti</i>      | HO149-PS06 | PALA-S1R | 357    | AY666068         |

In the non-repetitive C-terminal region, there were also nucleotide substitutions between sequences amplified from the same individual. In *O. varians*, where three sequences were amplified from three different individuals, there were 13 variable sites. In *C. moluccensis*, CM05-PS09 and CM05-PS11 were the sequences amplified from the same individual but 8 variable sites were found between them. In *P. sinensis*, six variable sites were found between three sequences amplified from the same individual PS02.

#### 4. Discussion

Results of this study show that the repetitive regions of sequences are similar between all the species examined. These species build different types of webs to catch prey and they

are not phylogenetically closely-related [3]. An examination of the repetitive region of sequences obtained in this study shows that GGX, GA and poly-A are found in the *MaSp1*-like sequences of both Mygalomorphae and Araneomorphae species examined. Gatesy et al. [10] also examined the repetitive region of the dragline gene of the Mygalomorphae *Eugrurus* (Dipluridae) and reported the presence of poly-A and GA motifs. Results of this and prior studies suggest that GGX, GA and poly-A motifs are plesiomorphic molecular traits and are already present in *MaSp1*-like genes of the primitive sub-order Mygalomorphae.

Results of this study also show that between cribellate and cribellate members of orb and space web builders, the non-repetitive C-terminal region of their *MaSp1*-like gene has very high similarity. Although in cribellate and cribellate web spiders the catching threads are produced in quite

Table 2

Primers used, nucleotide length and accession number of *MaSp1*-like sequences amplified from cDNA of *Nephila pilipes* and *Cyrtophora moluccensis*

| Species                | Number                        | Primers      | Length  | Accession number |
|------------------------|-------------------------------|--------------|---------|------------------|
| <i>Nephila pilipes</i> | NP-RTPD01                     | PALA-PALA    | 728     | AY666069         |
|                        | NP-RTPD03                     | PALA-PALA    | 731     | AY666070         |
|                        | NP-RTPD08                     | PALA-PALA    | 819     | AY666071         |
|                        | NP072-RTPS01                  | PALA-S1R     | 538     | AY666072         |
|                        | NP072-RTPS02                  | PALA-S1R     | 595     | AY666073         |
|                        | NP072-RTPS35                  | PALA-S1R     | 595     | AY666074         |
|                        | NP072-RTPS39                  | PALA-S1R     | 598     | AY666075         |
|                        | NP072-RTPS40                  | PALA-S1R     | 727     | AY666076         |
|                        | NP072-RTPS10                  | S1R-S1R      | 346     | AY666077         |
|                        | <i>Cyrtophora moluccensis</i> | CM091-RTPS02 | S1R-S1R | 346              |
| CM091-RTPS06           |                               | S1R-S1R      | 346     | AY666079         |
| CM091-RTPS07           |                               | S1R-S1R      | 346     | AY666080         |
| CM092-RTPS01           |                               | S1R-S1R      | 346     | AY666081         |

different ways [24], their *MaSp1*-like gene seems to be similar to each other. Beckwitt and Aracdiacono [1] suggested that the evolutionary mechanism generating the characteristic molecular structure of *MaSp1* had occurred before the phylogenetic separation of the two major orb-weaving families, Araneidae and Tetragnathidae. In this study, the C-terminal

|     |  |  |
|-----|--|--|
| 480 | Spidroin-1   | 480  |
|     | GLGSQAGRGR GQGAGAAAAA AVGAGQEGIR GQAGQGGYGG GLGSQGSGRG |  |
|     | NP02-PS02  |  |
|     | NP02-PS08  |  |
|     | NP11-SS01  |  |
|     | NP11-SS05  |  |
|     | NP11-SS08  |  |
|     | NP13-SS01  |  |
|     | NP13-SS04  |  |
|     | NP13-SS05  |  |
|     | NP13-SS07  |  |
|     | NP072-RTPS01   |  |
|     | NP072-RTPS02   | G-A-Q-   |
|     | NP072-RTPS10   |  |
|     | NP072-RTPS35   | G-A-Q-   |
|     | NP072-RTPS39   | G-A-Q-   |
|     | NP072-RTPS40   | -G-G--GYG- LG-Q-SG--- SGAGQGGYGG LG-QGA-AAA AAAAG-A-Q- |
| 530 | Spidroin_1   | 579  |
|     | GLGGQGAGAA AAAAGGAGQG GLGGQGAGQG AGAAAAAGG VRQGGYGGGLG |  |
|     | NP02-PS02  | ----- AG-----*   |
|     | NP02-PS08  | -- AG-----*  |
|     | NP11-SS01  |  |
|     | NP11-SS05  |  |
|     | NP11-SS08  | SI*  |
|     | NP13-SS01  |  |
|     | NP13-SS04  |  |
|     | NP13-SS05  |  |
|     | NP13-SS07  | L*S--- ----- AG-----*                                  |
|     | NP072-RTPS01   | ----- -Y--L*S--- S-*--G--- AG-----*                    |
|     | NP072-RTPS02   | -Y--L-GQG- G--GA----- -Y--L*S--- ----- AG-----*        |
|     | NP072-RTPS10   |  |
|     | NP072-RTPS35   | -Y--L-GQG- G--GA----- -Y--L*S--- ----- AG-----*        |
|     | NP072-RTPS39   | -Y--L-GQG- G--GA----- -Y--L*S--- ----- AG-----*        |
|     | NP072-RTPS40   | -Y--L-GQG- G--GA----- -Y--L*S--- ----- AG-----*        |
| 580 | Spidroin-1   | 629  |
|     | SQGAGRGGQG AGAAAAAGG AGQGGYGGGLG GQGVRGGLG G***QGAGAA  |  |
|     | NP02-PS02  | *****_--- S-----**_* T----- ---A-AA-AA AAAVG---QG      |
|     | NP02-PS08  | *****_--- S-----S**_* T----- ---A-AA-AA AAAVG---QG     |
|     | NP11-SS01  | *****_--- S-----**_* T----- ---A-AA-AA AAAV*****       |
|     | NP11-SS05  |  |
|     | NP11-SS08  | *****SSLS CTSGS**_* T----- ---A-AA-AA AAAVG---QG       |
|     | NP13-SS01  | *****_--- S-----**_* T----- ---A-AA-AA AAAV*****       |
|     | NP13-SS04  | *****_--- S-----**_* T----- ---A-AA-AA AAAV*****       |
|     | NP13-SS05  | *****_--- S-----**_* T----- ---A-AA-AA AAAV*****       |
|     | NP13-SS07  | *****_--- S-----**_* T----- ---A-AA-AA AAAV*****       |
|     | NP072-RTPS01   | *****_--- S-----**_* T----- ---A-AA-AA AAAV*****       |
|     | NP072-RTPS02   | *****_--- S-----**_* T----- ---A-AA-AA AAAV*****       |
|     | NP072-RTPS10   | *****_--- S-----**_* T----- ---A-AA-AA AAAV*****       |
|     | NP072-RTPS35   | *****_--- S-----**_* T----- ---A-AA-AA AAAV*****       |
|     | NP072-RTPS39   | *****_--- S-----**_* T----- ---A-AA-AA AAAV*****       |
|     | NP072-RTPS40   | *****_--- S-----**_* T----- ---A-AA-AA AAAV*****       |

Fig. 1. Aligning the *MaSp1*-like gene partial amino acid sequences amplified from *N. pilipes* with the corresponding portion of that of *N. clavipes*. While sequence from NP072 was obtained from cDNA, all the rest were amplified from genomic DNA. Dash line represents that the amino acid was the same as that of the spidroin-1 protein. Star symbol represents gap.

|     |  |  |
|-----|--|--|
| 630 | Spidroin-1   | 679  |
|     | AAGGAGQGGY GGVGSGASAA SAAASRLSSP QASSRVSSAV SNLV       |  |
|     | NP02-PS02  | GY--V----- **A-- ----- E----- ---S-----    |
|     | NP02-PS08  | GY-DV----- **A-- ----- E----- ---S-----    |
|     | NP11-SS01  | **----- **A-- ----- E----- ---S-----       |
|     | NP11-SS05  | ----- **A-- ----- E----- ---S-----         |
|     | NP11-SS08  | GY--V----- **A-- ----- E----- ---S-----    |
|     | NP13-SS01  | **----- **A-- ----- E----- ---S-----       |
|     | NP13-SS04  | ----- **A-- -S-----P-- E-----P-- ---S----- |
|     | NP13-SS05  | -V----- **A-- ----- E----- ---S-----       |
|     | NP13-SS07  | **----- **A-- ----- E----- ---S-----       |
|     | NP072-RTPS01   | **----- **A-- ----- E----- ---S-----       |
|     | NP072-RTPS02   | **----- **A-- ----- E----- ---S-----       |
|     | NP072-RTPS10   | **----- **A-- ----- EV----- ---S-----      |
|     | NP072-RTPS35   | **----- **A-- ----- E----- ---S-----       |
|     | NP072-RTPS39   | **----- **A-- ----- E----- ---S-----       |
|     | NP072-RTPS40   | **----- **A-- ----- E----- ---S-----       |
| 680 | Spidroin-1   | 729  |
|     | SAALSSTISN VVSQIGASNP GLSGCDVLIQ ALLEVVSALI QILGSSSIGQ |  |
|     | NP02-PS02  | ---N--- L---S--- -----V- ---L--- H-----    |
|     | NP02-PS08  | ---N--- L---SS--- -----V- ---L--- H-----   |
|     | NP11-SS01  | ---N--- L---SS--- -----V- ---L--- H-----   |
|     | NP11-SS05  | ---N--- L---S--- -----V- ---L--- -----     |
|     | NP11-SS08  | ---N--- L---S--- -----V- ---L--- -----     |
|     | NP13-SS01  | ---N--- L---SS--- -----V- ---L--- H-----   |
|     | NP13-SS04  | ---N--- L---S--- -----V- ---L--- -----     |
|     | NP13-SS05  | ---N--- L---S--- -----V- ---L--- -----     |
|     | NP13-SS07  | ---N--- L---SS--- -----V- ---L--- H-----   |
|     | NP072-RTPS01   | ---N--- L---SS--- -----V- ----- H-----     |
|     | NP072-RTPS02   | ---N--- L---SS--- -----V- ----- H-----     |
|     | NP072-RTPS10   | ---PN--- L---SS--- -----V- ----- H-----    |
|     | NP072-RTPS35   | ---N--- L---SS--- -----V- ----- H-----     |
|     | NP072-RTPS39   | ---N--- L---SS--- -----V- ----- H-----     |
|     | NP072-RTPS40   | ---N--- L---SS--- -----V- ----- H-----     |
| 730 | Spidroin-1   | 750  |
|     | VNYGSAGQAT QIVGQSIVYQA L                               |  |
|     | NP02-PS02  | -----                                      |
|     | NP02-PS08  | -----                                      |
|     | NP11-SS01  | -----                                      |
|     | NP11-SS05  | -----                                      |
|     | NP11-SS08  | -----                                      |
|     | NP13-SS01  | -----                                      |
|     | NP13-SS04  | -----                                      |
|     | NP13-SS05  | -----                                      |
|     | NP13-SS07  | -----                                      |
|     | NP072-RTPS01   | -----                                      |
|     | NP072-RTPS02   | -----                                      |
|     | NP072-RTPS10   | -----                                      |
|     | NP072-RTPS35   | -----                                      |
|     | NP072-RTPS39   | ---FCWD** RLLKSF                           |
|     | NP072-RTPS40   | -----                                      |

Fig. 1. (Continued).

region of the *MaSp1*-like gene of the Mygalomorphae funnel-web builder *Macrothelae holsti* is highly similar to that of the Araneomorphae web builders examined. This result shows that such molecular traits might have already been present before the separation of Araneomorphae and Mygalomorphae about 240 million years ago [27]. On the other hand, *Psechrus* is closely-related to *Dolomedes* and both are members of the superfamily Lycosoidae [8]. Despite the close phylogenetic relationship between *Psechrus* and *Dolomedes*,



|              |            |            |                                 |
|--------------|------------|------------|---------------------------------|
|              | 623        |            | 672                             |
| Spidroin-1   | AGAAAAGGAG | QGGYGGVGS  | ASAASAAASR LSSPQASSRV SSAVSNLVA |
| CM05-PS09    | -A--V----  | -----*     | *A-----H ---E-----S-            |
| CM05-PS11    | -A--V----  | -----*     | *A-----H ---E-----S-            |
| CM06-SSA04   | -A--V----  | -----*     | *A-----H ---E-----S-            |
| CM07-SS01    | -A--V----  | -----*     | *A-----H ---E-----S-            |
| CM091-RTPS02 | -A--V----  | -----*     | *A-----H ---E-----S-            |
| CM091-RTPS06 | -A--V----  | -----*     | *A-----H ---E-----D--S-         |
| CM091-RTPS07 | -A--V----  | -----*     | *A-----H ---E-----S-            |
| CM092-RTPS01 | -A--V----  | -----*     | *A-----H ---E-----S-            |
|              | 673        |            | 722                             |
| Spidroin-1   | GPTNSAALSS | TISNVVSQIG | ASNPLSGCD VLIQALLEV SALIQILGSS  |
| CM05-PS09    | ---P--PN   | -----S     | S----- -V----- -H----           |
| CM05-PS11    | -S-----PN  | -----S     | S----- -V----- -H----           |
| CM06-SSA04   | -----PN    | -----S     | S----- -V----- -H----           |
| CM07-SS01    | -----PN    | -----S     | S--S----- -V----- -H----        |
| CM091-RTPS02 | -S-----PN  | -----S     | S----- -V----- -H----           |
| CM091-RTPS06 | -S-----PN  | -----S     | S----- -V----- -H----           |
| CM091-RTPS07 | -S-----PN  | -----S     | S----- -V----- -H----           |
| CM092-RTPS01 | -S-----PN  | -----S     | S----- -V----- -H----           |
|              | 723        |            | 751                             |
| Spidroin-1   | SIGQVNYGSA | GQATQIVGQS | VYQAL                           |
| CM05-PS09    | -----      |            |                                 |
| CM05-PS11    | -----      |            |                                 |
| CM06-SSA04   | -----      |            | -----GEFA                       |
| CM07-SS01    | -----      |            |                                 |
| CM091-RTPS02 | -----      |            |                                 |
| CM091-RTPS06 | -----      |            |                                 |
| CM091-RTPS07 | -----      |            |                                 |
| CM092-RTPS01 | -----      |            |                                 |

Fig. 2. Aligning the *MaSp1*-like gene partial amino acid sequences amplified from *Cyrtophora moluccensis* with the corresponding portion of that of *N. clavipes*. While sequences from individuals CM091 and CM092 were obtained from cDNA, all the rest were amplified from genomic DNA. See Fig. 1 for symbols used.

their sequences in the non-repetitive C-terminal region differ considerably (Fig. 5). These results indicate that the C-terminal sequences of *MaSp1* are similar between distantly-related species exhibiting similar foraging behavior but are different between closely-related taxa having different foraging modes.

The high similarity of *MaSp1*-like gene in the non-repetitive C-terminal region of a phylogenetically diverse array of web spiders suggests that this region might be maintained by strong selection pressures and has important functions. During the spinning process when the water soluble silk protein is transformed into water insoluble fibers many chemical reactions such as disulfide bond formation, cation interaction and glycosylation occur [16,19]. The non-repetitive C-terminal sequence might serve as a signal to stimulate the secretion of certain glands involved in those reactions [1]. Also, the maintenance of a high concentration of soluble protein in the silk gland involves axial alignment and interaction of polypeptides [32]. Beckwitt and Arcidiacono [1] proposed that the C-terminal sequence may be involved in these processes to maintain the soluble or liquid crystalline state of silk while it is still in the gland. So far there is no evidence for these hypotheses and functional studies are needed to deter-

|            |            |            |                                 |
|------------|------------|------------|---------------------------------|
|            | 623        |            | 672                             |
| Spidroin-1 | AGAAAAGGAG | QGGYGGVGS  | ASAASAAASR LSSPQASSRV SSAVSNLVA |
| OV01-SS06  |            | D K-----   | .A----- -E-----S-               |
| OV02-SS01  |            |            | ---.A----- -E-----S-            |
| OV10-SS16  |            |            | ---.A----- -E-----S-            |
| PS01-PS02  |            |            | ---.A----- H--E-----S-          |
| PS02-SS03  | -A--V----  |            | .A----- -E-----S-               |
| PS02-SS06  | -A--V----  |            | .A----- P--E-----S-             |
| PS02-SS11  | -A--V----  |            | .A----- -E-----S-               |
| HO149-PS06 | -A--V----  |            | .A-----H ---E-----SG            |
|            | 673        |            | 722                             |
| Spidroin-1 | GPTNSAALSS | TISNVVSQIG | ASNPLSGCD VLIQALLEV SALIQILGSS  |
| OV01-SS06  | -----N     | -----S     | S----- -V----- -P-H----         |
| OV02-SS01  | -----N     | -----S     | S----- -V----- -H----           |
| OV10-SS16  | -----N     | ---L----   | S----- -V-----L-----            |
| PS01-PS02  | -----N     | -----S     | S----- -V----- -H----           |
| PS02-SS03  | -----PN    | -----S     | S----- -V----- -H----           |
| PS02-SS06  | -----PN    | -----S     | S----- -PV----- -H----          |
| PS02-SS11  | -----PN    | -----S     | S----- -V----- -H----           |
| HO149-PS06 | -S-----PN  | -----S     | S----- -V----- -H----           |
|            | 723        |            | 747                             |
| Spidroin-1 | SIGQVNYGSA | GQATQIVGQS | VYQAL                           |
| OV01-SS06  | -----      |            |                                 |
| OV02-SS01  | -----      |            |                                 |
| OV10-SS16  | -----      |            |                                 |
| PS01-PS02  | -----      |            |                                 |
| PS02-SS03  | -----      |            |                                 |
| PS02-SS06  | -----      |            |                                 |
| PS02-SS11  | -----      |            |                                 |
| HO149-PS06 | ----D----  |            | -----A                          |

Fig. 3. Aligning the *MaSp1*-like gene partial amino acid sequences amplified from *Octonba varians* (OV), *Psechrus sinensis* (PS) and *Macrothele holsti* (MH) with the corresponding portion of that of *N. clavipes*. See Fig. 1 for symbols used.

mine why this portion has been so conservatively maintained during evolution.

In this study, inter- and intra-individual variations in *MaSp1* sequences are found in four Araneomorphae species examined. The variations are especially pronounced in the repetitive regions of *N. pilipes*. The repetitive motifs of *N. pilipes* are the same as those of the published sequences [2,10,33], but the arrangement pattern is very different. In addition, there are some inter- and intra-individual variations in C-terminal region amino acid sequences. Although some of the amino acids were identical between sequences, an examination of their DNA sequences shows that there are many nucleotide substitutions. Such results provide important insight about the genetic architectures of the *MaSp1* gene. In this study, we obtained several forms of *MaSp1*-like gene sequences from genomic DNA of different and same individuals. Since a great majority of sequences contained the conservative C-terminal region, such variation in sequences cannot be interpreted as amplified from different portions of the highly variable repetitive region. Therefore, in Araneomorphae spiders *MaSp1* exhibits several forms in one individual,

|            |            |             |                                   |
|------------|------------|-------------|-----------------------------------|
|            | 1          |             | 50                                |
| Sequence 1 |            | GGAGQG      | GYGSLGGQGR GG1GGTRSLG             |
| Sequence 2 |            | GGAGQG      | GYGSLGGQGR GG1GGTRSLG             |
| Sequence 3 | TAGAGGRGGY | GGLGGQAGA   | AEAAGGAGQG GYGSLGGQGR GG1GGTRSLG  |
|            | 51         |             | 100                               |
| Sequence 1 | ANSGEADAAG | DRGFRVITYEA | TLQDARQGGY GGLGGQGTGA GATAAAAAGG  |
| Sequence 2 | ANSGEADATG | DRGFRVITYEA | TLQDARQGGY GGLGGQGTGA GATAAAAAGG  |
| Sequence 3 | ANSGEADAAG | DRGFRVITYEA | TLQDARQGGY GGLGGQGTGA GATAAAAAGG  |
|            | 101        |             | 150                               |
| Sequence 1 | AGQGGYSGLD | RQRSAAAAAG  | ASVAAQGGYG GLGGQAGAA GAAAAGGAGR   |
| Sequence 2 | AGQGGYSGLD | RQRSAAAAAG  | ASVAAQGGYG GLGGQAGAA GAAAAGGAGR   |
| Sequence 3 | AGQGGYSGLD | RQRSAAAAAG  | ASVAAQGGYG GLGGQAGAA GAAAAGGAGR   |
|            | 151        |             | 200                               |
| Sequence 1 | GGYGGLVGQG | AGAAAAGAGE  | AGQGGYGSGLG GQGTGAGGEE AAATGGAGQG |
| Sequence 2 | GGYGGLVGQG | AGAAAAGAGE  | AGQGGYGSGLG GQGTGAGGEE AAATGGAGQG |
| Sequence 3 | GGYGGLVGQG | AGAAAAGAGE  | AGQGGYGSGLG GQGTGAGGEE AAATGGAGQG |
|            | 201        |             | 250                               |
| Sequence 1 | GYGGLGGQGA | SAAGSAGQGG  | YGGLAGQRAG AAATGGLFDG SRDVGGFSFA  |
| Sequence 2 | GYGGLGGQGA | SAAGSAGQGG  | YGGLAGQRAG AAATGGLFDG SRDVGGFSFA  |
| Sequence 3 | GYGGLGGQGA | SAAGSAGQGG  | YGGLAGQRAG AAATGGLFDG SRDVGGFSFA  |
|            | 251        | 273         |                                   |
| Sequence 1 | VSGSSGGYRR | GGYVGN      |                                   |
| Sequence 2 | VSGSSGGYRR | GGYVGN      |                                   |
| Sequence 3 | VSGSSGGYRR | GGYVGNSTAA  | AAA                               |

Fig. 4. Repetitive region of sequences amplified from cDNA of one *N. pilipes* individual.

indicating that it is either a multiple gene system or has a multiple exon/intron organization.

Currently, we are not able to evaluate these two hypothesized genetic architectures of *MaSp1*. Based on the large size of the known partial sequences, investigators have proposed that the *MaSp1* gene is composed of a single large exon and has allelic variants [1,2,11,15,33]. According to this hypothesis *MaSp1* might then be a multiple gene system and sequences of this study were amplified from different members of this gene family. On the other hand, Hayashi and Lewis [13] reported that the *Flag* gene of orb-weaving spiders has multiple exons and introns. The exons contain both conservative and repetitive regions and the arrangement and number of motifs in the latter are variable. Although the evolutionary relationship between *MaSp1* and *Flag* is unknown [12], we cannot rule out the possibility that both have similar molecular architecture. Moreover, Craig and Riekel [5] compared the molecular structure and organization of *MaSp1* and *Ap-fhc*

|                  |            |            |            |             |            |
|------------------|------------|------------|------------|-------------|------------|
| <i>Nephila</i>   | SRLSSPEASS | RVSSAVSNLV | SSGPTNSAAL | SNTISNVVVSQ | ISSSNPGLSG |
| <i>Psechrus</i>  | --H-----   | -----      | -----      | -----       | -----      |
| <i>Dolomedes</i> | -----Q-A-  | -----S-    | -N-QV-V--- | PSI--SLS-S  | --A-STAA-D |
| <i>Nephila</i>   | CDVLVQALLE | VVSALIHILG | SSSIQGVNYG | SAGQATQIVG  | QSVYQAL    |
| <i>Psechrus</i>  | -----      | -----      | -----      | -----       | -----      |
| <i>Dolomedes</i> | -E---V---  | I---VQ-VS  | -ANVGYI-PE | AS-SLNAVGS  | ALAAAAMG   |

Fig. 5. A comparison of *MaSp1* non-repetitive region amino acid sequences of *Nephila pilipes*, *Psechrus sinensis* and *Dolomedes*. Sequence of *Dolomedes* was published in Gatesy et al. (2001).

genes of the moth *Antheraea pernyi* and suggested that the former might have a genetic architecture similar to the latter, which was made up of two exons. If this is the case, different sequences obtained from same or different individuals of spiders in this study might be amplified from different exons. Full-length gene sequences obtained from genomic DNA are needed to be able to determine the genetic architecture of *MaSp1*.

## Acknowledgements

We would like to thank Y.H. Chang, Y.R. Huang and J. Haupt for their help in specimen collections and C.H. Chang for assistance and advice in molecular operations. Discussions with W.Y. Lo greatly improved our manuscript. This work was supported by a grant from the National Science Council, Taiwan, ROC (NSC-92-2311-B-029-002) to I.M. Tso.

## References

- [1] Beckwitt R, Arcidiacono S. *J Biol Chem* 1994;269:6661.
- [2] Beckwitt R, Arcidiacono S, Stote R. *Insect Biochem Mol Biol* 1998;28:121.
- [3] Coddington JA, Levi HW. *Annu Rev Ecol Syst* 1991;22:565.
- [4] Craig CL, Riekel C, Herberstein ME, Weber RS, Kaplan D, Pierce NE. *Mol Biol Evol* 1994;17.
- [5] Craig CL, Riekel C. *Comp Biochem Physiol B* 2002;133:493.
- [6] Eberhard WG. *Annu Rev Ecol Syst* 1990;21:341.
- [7] Elgar MA. *J Anim Ecol* 1994;63:880.
- [8] Fang K, Yang CC, Lue BW, Chen SH, Lue KY. *Zool Stud* 2000;39:107.
- [9] Foelix RF. *Biology of spiders*. New York; 1996.
- [10] Gatesy J, Hayashi C, Motriuk D, Woods J, Lewis R. *Science* 2001;291:2603.
- [11] Guerette PA, Ginzinger DG, Weber BHF, Gosline JM. *Science* 1996;272:112.
- [12] Hayashi CY. In: DeSalle R, Giribet G, Wheeler W, editors. *Molecular systematics and evolution: theory and practice*, 2002. p. 209.
- [13] Hayashi CY, Lewis RV. *Science* 2000;287:1477.
- [14] Higgins L. *Oikos* 2002;99:377.
- [15] Hinman MB, Lewis RV. *J Biol Chem* 1992;267:19320.
- [16] Kaplan DL, Fossey S, Mello CM. *Mater Res Soc Bull* 1992;10:41.
- [17] Lehtinen PT. *Ann Zool Fennici* 1967;4:199.
- [18] Lubin YD. *Forma et Funct* 1973;6:337.
- [19] Mori H, Tsukada M. *Rev Mol Biotechnol* 2000;74:95.
- [20] Opell BD. *Biol J Linn Soc Lond* 1997;62:443.
- [21] Opell BD. *Funct Ecol* 1998;12:613.
- [22] Opell BD. *Biol J Linn Soc Lond* 1999;68:593.
- [23] Opell BD, Bond JE. *Evol Ecol Res* 2001;3:567.
- [24] Peters H. *Zoomorphology* 1984;104:96.
- [25] Platnick NI. *Am Mus Novit* 1977;2627:1.
- [26] Sambrook J, Fritsch EF, Maniatis T. *Molecular cloning: a laboratory manual*. New York; 1989.
- [27] Selden PA, Gall JC. *Palaeontology* 1992;35:211.
- [28] Tso IM, Severinghaus LL. *Anim Behav* 1998;56:219.
- [29] Uetz GW, Halaj J, Cady AB. *J Arachnol* 1999;27:270.
- [30] Vollrath F. *Int J Biol Macromol* 1999;24:81.
- [31] Watanabe T. *Behav Ecol* 1999;10:607.
- [32] Winkler S, Kaplan DL. *Rev Mol Biotechnol* 2000;74:85.
- [33] Xu M, Lewis RV. *Proc Natl Acad Sci* 1990;87:7120.