Structural Organization and Evolution of a Cluster of Small Serum Protein Genes of *Protobothrops Flavoviridis* Snake

Yasuyoshi Tanaka[1)], Narumi Shioi[1)], Shigeyuki Terada[1)], and Masanobu Deshimaru[1),*]

（Received November 30, 2012）

Abstract

We previously purified and characterized five isoforms of small serum proteins, SSP-1 to -5 from *Protobothrops flavoviridis* (Habu snake) as venom enzyme blockers, and cloned their respective cDNAs. As a result of comparison of cDNA nucleotide sequences, coding sequences were much more diverse than those of untranslated regions and, nonsynonymous substitutions occurred unusually frequently, implying accelerated evolution of *SSPs*. In order to investigate whether such evolutionary charasteristics are observed in whole nucleotide sequences of genes for SSPs including introns, we made an attempt to clone genomic fragment encoding SSPs by PCR using oligonucleotides designed according to gene organization of PSP94, a human homolog of SSP. Consequently, DNA fragments encompassing genes for SSP-1 and -2, which consist of four exons and three introns, were successfully obtained and their respective nucleotide sequences were determined. When their sequences were aligned and compared, unusual diversification of their exonic sequences were demonstrated while the intronic sequences were highly conserved. This finding strongly supports the probability of accelerated evolution of *SSPs*. In addition, we cloned genomic fragment corresponding to an intergenic region between genes for SSP-1 and -2. As analyzed its nucleotide sequence, it was found that genes for SSP-2 and SSP-1 occurred in head-to-tail arrangement with a 3.4-kb intergenic region, implying that *SSP* genes form a cluster in habu snake genome.

Introduction

Crotalinae snake venoms generally contain a wide variety of hydrolytic enzymes enabling efficient harm to the prey of these snakes and the predators. The venom enzymes can be classified into several isozyme families, including the phospholipase A_2 (PLA_2) family,[1] the serine protease family,[2] and the metalloprotease (MP) family.[3] Even members of the same isozyme family can differ with respect to their physiological activities. For example, among the PLA_2 isozymes of the *P. flavoviridis* (Habu snake), the Asp49 PLA_2 shows phospholipase activity, whereas basic proteins I and II with potent myotoxicities show only very weak enzymatic activity.[4,5] Furthermore, the physiological activities of the enzymes of the MP family of snake venom are considerably varied: some induce hypodermic hemorrhage and others trigger

[1)] Department of Chemistry, Faculty of Science, Fukuoka University, 8-19-1 Nanakuma, Jonan-ku, Fukuoka, 814-0180, Japan

* Correspondence should be addressed: Masanobu Deshimaru, Tel: +81-92-871-6631 ext. 6215. Fax: +81-92-865-6030. E-mail: deshi@fukuoka-u.ac.jp.

apoptotic cell death.[3] Comparisons of the nucleotide sequences of genes for such snake venom isozymes revealed the striking feature that nucleotide substitutions have occurred substantially more in the protein-coding sequences than in the noncoding sequences. This observation is in stark contrast to the observations of general isozyme genes. Furthermore, a statistical analysis indicated that this pattern was due to an enhanced substitution rate in the protein-coding sequences rather than the suppression of substitutions in the noncoding sequences. Accordingly, this evolutionary pattern, specifically observed in snake venom isozyme genes, is an example of "accelerated evolution" and is considered to have contributed to the rapid adaptive evolution of snake venoms.[6-9]

Since these venom components are also highly toxic to the snakes themselves when accidentally injected into their bloodstream, venomous snakes possess protective proteins that inhibit venomous enzymes in their blood. For example, several isoforms of PLA_2 inhibitors[10, 11] and a Habu serum factor (HSF) that inhibits venom MPs[12, 13] have been isolated and characterized in the Habu snake. In addition, we isolated a novel 10-kDa protein, small serum protein 1 (SSP1), having MP inhibitory activity from Habu serum.[14] Unlike other antihemorrhagic proteins, however, SSP1 could only inhibit brevilysin H6, a weakly hemorrhagic MP isolated from *Gloydius halys brevicaudus* venom.[15] Analysis of the amino acid sequence indicated that SSP1 belongs to the prostatic secretory protein of 94 amino acids (PSP94) protein family, based on topological similarities of the cysteine residues, although the sequence homology was low. Four additional similar proteins, SSP2 to SSP5, in the same serum from the habu snake were also isolated; SSP2 and SSP5 have binding affinity to triflin, a smooth muscle contraction blocker isolated from habu snake venom whereas SSP3 inhibits brevilysin H6 similarly to SSP1.[16, 17] We subsequently cloned cDNAs encoding these five SSPs and observed an unusually high accumulation of nonsynonymous nucleotide substitutions in the protein-coding region encoding small serum proteins (SSPs), compared to the noncoding region of cDNA sequences, in the Habu snake.[17] This observation indicated that accelerated evolution has also occurred in SSPs, implying that toxic enzymes and their compensatory inhibitors co-evolved via such a unique evolutionary manner. Thus, the accelerated evolution of SSPs was estimated from their cDNA sequences, but it remains unknown whether intensive accumulation of the nucleotide substitutions occurred exclusively in their protein-coding regions as compared to their intronic sequences.

In the present study, for a further understanding of the evolution of *SSPs*, genomic sequence corresponding to their genes and intergenic regions were cloned and analyzed. Consequently, we obtained a result supporting the hypothesis of accelerated evolution of *SSPs*. In addition, unusually dense arrangement and small sizes of their genes were estimated. Involvement of these characteristics in the evolution of *SSPs* will be discussed.

Materials and Methods

Materials

Habu snake livers were collected in Amami Oshima Island. All PCR reactions were performed using PrimeStar GXL DNA polymerase (Takara Bio, Kyoto, Japan). All enzymes used for other genetic manipulations were purchased from Takara Bio Inc. All chemical reagents were obtained from Nakarai Tesque (Kyoto, Japan). Oligonucleotides used for amplification of genomic fragments were synthesized by Genenet (Fukuoka, Japan). Their respective nucleotide sequences are listed in Table 1.

Cloning of partial DNA fragments of *SSP* genes

Genomic DNA was extracted from Habu snake liver according to the methods outlined in Blin and Stafford.[18] DNA fragments corresponding to 5'-half (exon 1 to exon 2), 3'-half (exon 2 to exon 4), and intergenic region of *SSP1* and *SSP2* genes were amplified by PCR from the genomic DNA, respectively. Oligonucleotide pair used in each amplification is summarized in Figure 1A and B. The thermal program used for all the amplifications consisted of 35 cycles of 98°C for 10 s; 55°C for 15 s; 68°C for 8 min. Amplified DNA fragments were analyzed on and purified from 0.8% agarose/TBE gel. They were then ligated to the pBluescript II plasmid vector at the *Eco*RV site, and the resultant bacterial clones were isolated. Nucleotide sequences of the cloned DNA fragments were determined by the dideoxy dye

Table 1. Oligonucleotide primers for amplification of genomic fragments.

Primer	Sequence (5' to 3')	Location[a]	Orientation
gSSP1-e1S	TCGTATAAATTGGAGGAGCAGATTCCTCAG	20-49	sense
gSSP1-e2A	CATCTATATGCTCAGGCCTTGGAGCACAGA	182-211	antisense
gSSP1-e2S	GGAGTCTGTGCTCCAAGGCCTGAGCATATAGAT	178-210	sense
gSSP1-e4A	GATTCAACACTGATTCGCATCCATCCCTGA	356-385	antisense
gSSP2-e1S	GCTTCCTTACATGACTTCACAGAGAGGTGCTG	1-32	sense
gSSP2-e2A	CCATGGCATCATCTGTAGGCGAAGATACAAGGG	159-188	antisense
gSSP2-e2S	CTGCGGTATAGGTCCCCTTGTATCTTCGCCTACA	145-178	sense
gSSP2-e4A	GCATCCTGCCCTTTCAGCAAGACCACCATACCTG	304-337	antisense
gSSP-3'S	GCGCTGTGATGCCTGATCTCTTTACCGGGACCCACTGG	400-437 (SSP2)	sense
gSSP-5'A	GGGAGGAAGGASAARGAGSRRYACCTCTCTSTGAAGTCA	50-88 (SSP1)	antisense

[a]Nucleotide position in the corresponding cDNA sequence.

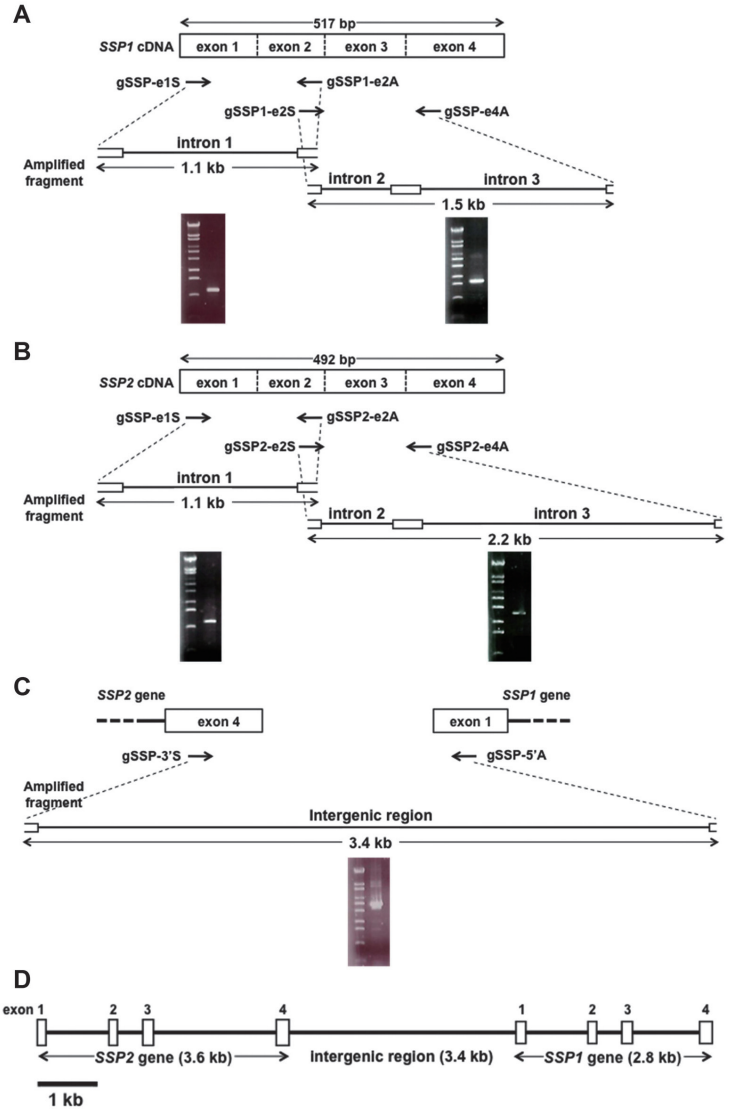

Figure 1. Strategy for cloning of genomic fragment and gene organization of SSPs. (A)(B) Schematic representation of the structure of cDNAs for SSP1 (A) and SSP2 (B) and the locations of oligonucleotides used for amplification of genomic fragments (upper). Broken lines in the cDNA structure indicate the positions of intron insertion predicted from the case of human *PSP94* gene. Structures of DNA fragments resulted from PCR with respective pairs of oligonucleotides are shown with corresponding electrophenograms (lower). (C) Schematic representation of the locations of oligonucleotides used for amplification of intergenic region (upper) and structure of that between *SSP2* and *SSP1* resulted from PCR with corresponding electrophenogram (lower). In all the electrophoreses, PCR product (right lane) was loaded together with λ/Sty I DNA size marker (left lane). (D) Schematic showing of a organization of genomic region containing genes for SSP-2 and SSP-1 completed by assembling the PCR-amplified genomic fragments. Note that *SSP1* gene is located at the downstream of *SSP2* gene and the orientations of two genes are the same.

terminator method using an ABI PRISM3100 genetic analyzer.

Mathematical analyses

Alignment and phylogenetic analysis of nucleotide and amino acid sequences were performed using CLUSTALW software, which is available online at the DNA DataBase Japan (DDBJ) website (http://clustalw.ddbj.nig.ac.jp). Values of nucleotide substitutions per site (K_N) in the noncoding sequences, including introns, were estimated from the aligned sequence data. Numbers of synonymous and nonsynonymous substitutions per site (K_S and K_A, respectively) in the protein-coding sequence were calculated using the Nei-Gojobori method[19] as implemented with the KaKs Calculator Beta, developed by Zhang *et al.*[20] Other analyses, including an homology dot plot, were performed using the DNASIS package sequence analysis software developed by Hitachi Solutions Co. Ltd. (Tokyo, Japan).

Results

Cloning of genomic DNA fragments composing *SSP* genes.

To determine complete DNA sequences of *SSP1* and *SSP2* genes, two pairs of oligonucleotide primers were used for PCR amplification of 5'-half (exon 1 to exon 2) and 3'-half (exon 2 to exon 4) of each gene (Figure 1A). The primers were designed according to the exon/intron organization of gene for PSP94, which is human homolog of SSPs and consists of four exons and three introns (Table 1).[21] As a result of respective PCR reactions, respective DNA fragments were obtained as shown in Figure 1A: the fragment sizes were 1.1 kb and 1.5 kb for 5'- and 3'-halves of *SSP1* and 1.1 kb and 2.2 kb for *SSP2*. Their complete nucleotide sequences were then determined and confirmed to agree to corresponding cDNA sequences. As the nucleotide sequences of 5'- and 3'- halves of genomic sequences and the cDNA sequences were assembled, the *SSP1* and *SSP2* genes consisted of 2,791 bp and 3,571 bp, from transcription start sites to poly(A) addition sites, respectively (Figure 1C). Comparison of the organization of *SSP* genes and human *PSP94* gene shows striking difference in their size resulted simply from the much smaller sizes of introns of *SSP* genes (Figure 2).

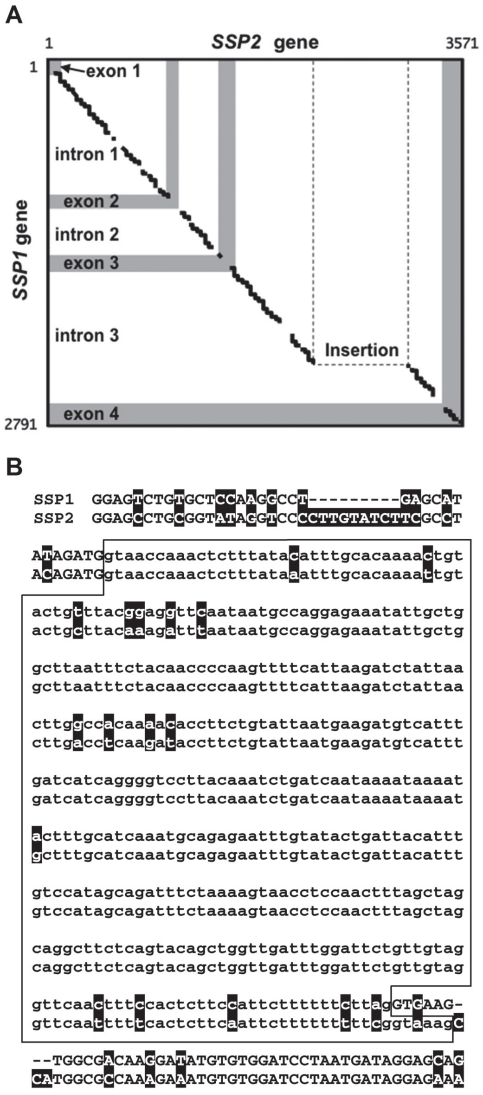

Figure 2. Comparison between genomic sequences encoding SSP1 and SSP2. (A) Homology dot plot resulted from the comparison between nucleotide sequences of *SSP1* gene (2791 bp) and *SSP2* gene (3571 bp). Nucleotide sequences for SSP1 and SSP2 were placed vertically and horizontally, respectively, and a dot was placed at the point where the 22 out of the 23-nt window matched. Regions in which the exonic sequences were compared are shaded. (B) Aligned nucleotide sequences encompassing intron 3 of *SSP1* and *SSP2* genes. Exonic and intronic nucleotides are shown in upper and lower cases, respectively. Nucleotide region enclosed by solid line corresponds to intron 3. Nucleotides in discordance between *SSP1* and *SSP2* are highlighted.

On the other hand, as an attempt to obtain intergenic fragment between *SSP* genes, PCR amplification from habu snake genomic DNA was carried out. Sense and antisense primers used for this amplification was designed according to conserved nucleotide sequence in 3'- and 5'-UTR of SSP cDNAs, respectively, for simultaneous amplification of any combination of neighboring *SSP* genes (Figure 1B and Table 1). An electrophoretic analysis of PCR product showed amplification of multiple DNA fragments with different lengths, implying amplification of multiple intergenic regions as expected. The most effectively amplified DNA fragment with a length of 3.4 kb was isolated and its nucleotide sequence was determined (Figure 1B). As compared to the cDNAs for five SSPs, it was found that a 3,245-bp intergenic region connecting 3'-end of *SSP2* gene and 5'-end of *SSP1* gene. This result also suggested that these two *SSP* genes are located in close proximity in the genome with 3.4-kb interval (Figure 1D). As an additional result from this experiment, a 2.3-kb DNA fragment corresponding to intergenic region between two *SSP* genes was obtained (data not shown). Nucleotide sequence of its one end is completely agree with that of 5'-terminal of *SSP2*, but the another end showed high but incomplete identity to 5'-terminal of *SSP2*. This suggests that more than two *SSP* genes form a cluster in habu snake genome and there is a gene encoding uncharacterized SSP2-like protein or a pseudogene of *SSP* gene family.

Comparison of nucleotide sequences of *SSP1* and *SSP2* genes

In our previous report for cloning of SSP cDNAs, low sequence identity between mature protein-coding sequences despite of high conservation of UTR sequences was demonstrated and this characteristics was regarded as indicating accelerated evolution that had been observed in several snake venom isozyme families.[17] Here, to determine whether such diversity of nucleotide sequences is limited only in the protein-coding regions in the case of SSPs, the nucleotide sequences of *SSP1* and *SSP2* genes were compared. A result from homology dot plot applied to the whole nucleotide sequences of these two genes was shown in Figure 2A. Dots, which indicates an agreement at 22 nt positions within a 23-nt window compared, appeared almost consecutively in comparison between intronic sequences whereas they interrupted at the comparison between coding exons, indicating that nucleotide substitutions between these two genes are accumulated mainly in the coding exons and conservation of intronic sequences seems to be comparable to that of exonic noncoding sequences. Additionally, this plot also suggests that a large insertion, which consists of 790 bp, in intron 3 of *SSP2* gene was the main cause of difference in length of two *SSP* genes.

An alignment of the nucleotide sequences shows more obviously the concentration of nucleotide substitution to the coding regions. Aligned nucleotide sequences of *SSP* genes, ranging from the ending of exon 2 to starting of exon 3, is shown in Figure 2B. In the intronic sequence, relatively small number of substitutions were found: sequence identity of this intron was 98.0%. This tendency is very similar to the case of 5'- and 3'-UTR. In stark contrast, much larger number of substitutions, as well as insertions/deletions, were found in the flanking exonic regions as seen in previous report.[17] These results clearly indicated that nucleotide substitutions between SSP genes occurred intensively in their protein-coding regions while the exonic noncoding sequences and introns have been conserved to a similar extent.

For a detailed characterization of nucleotide substitutions between *SSP1* and *SSP2*, their nucleotide sequences were mathematically analyzed using the Nei-Gojobori method. Comparison of the protein-coding exons revealed a substitution rate per synonymous site (K_S) and non-synonymous site (K_A) of 0.020 to 0.077 and 0.127 to 0.195, respectively (Figure 3). On the other hand, comparison of the exonic noncoding sequences and introns revealed substitution rates per nucleotide site (K_N) of 0.020 to 0.126 (Figure 3). Thus, the K_S value was approximately equal to the K_N value, indicating that substitution in the coding sequences was not suppressed. Furthermore, the K_A values are 2.4-, 9.8-, and 4.1-fold larger than K_S values in exon 2, 3, and 4, respectively, indicating that substitutions occurred predominately at the nonsynonymous amino acid sites and that this tendency occurred evenly over all the coding exons despite of interruption by introns. Although these two observations are in contrast to the patterns of general isozyme genes,[19] they are quite similar to the

patterns observed in snake venom isozymes, implying accelerated, adaptive diversification of SSPs.

Discussion

To date, five isoforms of SSPs, SSP1 to SSP5, have been isolated from habu snake serum and their cDNAs were cloned. In the previous report, we demonstrated that nucleotide substitutions were found intensively in protein-coding regions of cDNAs and they preferably occurred at amino acid nonsynonymous sites.[17] These characteristics were in stark contrast to the case of general isozyme genes but in good accordance with the observation in cDNAs encoding snake venom components, such as PLA2 isozymes, serine proteases, and metalloproteases.[8, 9, 22] In the present study, it was found that intronic sequences of SSP1 and SSP2 are highly homologous just similar to the previously reported case of 5'- and 3'-UTR, indicating that such a high rate of substitution was occurred only in the protein-coding sequence in whole these genes. Thus, this result strongly support the hypothesis that SSPs were also regarded to have evolved via accelerated and adaptive manner. Indeed, different SSPs do not share a physiological activity despite of their homologies to each other: *e.g.* SSP1 and SSP3 inhibit brevilysin H6, an MP isolated from *G. halys brevicaudus* venom, whereas SSP2 and SSP5 exhibit affinity to triflin, a smooth muscle contraction blocker isolated from habu snake venom.[14, 17] Such diversity in SSP activities might have resulted from the selective pressure with which a snake lineage with multiple compensatory mechanisms for their own venom components could thrive better. It is likely that habu snake which lacks SSP1 or SSP2 would become more susceptible to triflin or metalloprotease H6, respectively, when these venom components are injected. This probability could be further discussed given the periods of emergence of respective venom components and of divergence of SSP functions could be estimated.

PCR amplification of intergenic regions were carried out using oligonucleotide primers corresponding to well-conserved sequence of SSPs, since arrangement, and even intergenic distances, of *SSP* genes had been quite unknown. As the result, simultaneous amplification of several DNA fragments was observed and the major product was found to be an intergenic fragment between *SSP1* and *SSP2* after cloning and sequence determination. The nucleotide sequence of this fragment indicated that these two genes were arranged within relatively close distance and in head-to-tail orientation. In this experiment, another DNA fragment with a length of 2.3 kb was additionally analyzed and this also corresponded to an intergenic region of *SSP* genes. Nucleotide sequence of this fragment indicated that this was amplified with a single, gSSP-5' primer, and its one end was completely agree to 5'-end of SSP2 cDNA. Since the nucleotide sequence of the one end of this fragment did not completely agree to that for any known SSPs, it is likely that *SSP2* gene is linked to a gene encoding unknown member of *SSP* or a

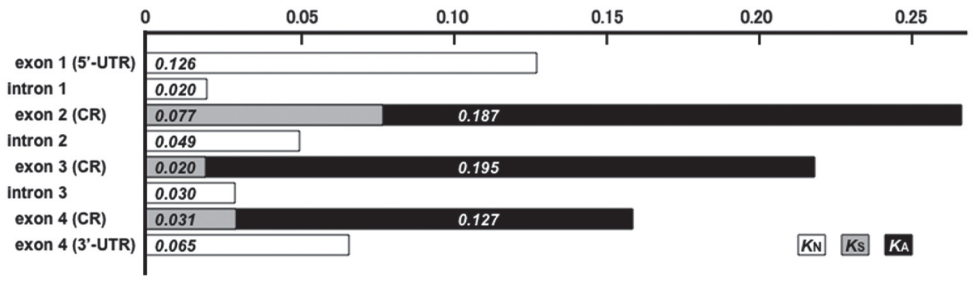

Figure 3. Comparison of substitution rates between different sequence regions of cDNAs and genes for HSF and HLP. K_N values deduced for 5'- and 3'-UTRs and three introns are shown with open bars. K_S and K_A values are shown with shaded and closed bars, respectively for three exons that compose coding regions (CRs) for SSP1 and SSP2. The values are also displayed in the respective bars.

SSP pseudogene in head-to-head orientation with a closer distance. These results from two intergenic fragments demonstrates that more than two *SSP* genes form cluster(s) in the genome and it can contain genes for novel SSPs. If the four genomic segments cloned in this study, corresponding to 2.3-kb intergenic region, SSP1, SSP2, and 3.4-kb intergenic region, respectively, were really linked together in the genome as expected from sequence overlaps between them, it seems that three *SSP* genes are included in an approximately 15-kb genomic region. Such gene arrangement with high density could be achieved by relatively small sizes not only of the intergenic regions but also of the genes. *SSP* genes consist of around 3 kb whereas the size of the gene for human PSP94 is 13.2 kb,[21] despite that both of them belong to a single gene family (Figure 4). As comparing their gene organization, all of three introns of *SSP* genes are strikingly shorter than those of *PSP94* gene. Although it is currently unimaginable what kind of molecular mechanism caused such uniform difference in the length of introns, much smaller sizes of *SSP* genes might be advantageous for cluster formation.

In conclusion, a genomic segment of habu snake was successfully cloned and partial organization of *SSP* genes were estimated. The nucleotide sequences of *SSP1* and *SSP2* genes suggested their accelerated evolution, which might have helped the enrichment of SSPs with different biological activities. On the other hand, cluster formation in genome and relatively small sizes of introns and intergenic regions of *SSP1* and *SSP2* genes were demonstrated. However, genomic arrangement of genes for SSP3, SSP4, and SSP5, as well as whether they also participate in the gene cluster, remains unknown in present study. For further understanding of the cluster formation of *SSP* genes, collective analysis of DNA fragments obtained in this study or new cloning experiment using cosmid library consisting of large genomic fragments of habu snake may be useful.

References

1. Kini RM, *Toxicon*, **42**, 827-840 (2003).
2. Matsui T, Fujimura Y, Titani K, *Biochim. Biophys. Acta*, **1477**, 146-156 (2000).
3. Fox JW and Serrano SM, *Toxicon*, **45**, 969-985 (2005).
4. Iwanaga S and Kihara H, *J. Biochem.*, **107**, 400-408 (1990).
5. Kihara H, Uchikawa R, Hattori S, Ohno M, *Biochem. Int.*, **28**, 895-903 (1992).
6. Ohno M, Ménez R, Ogawa T, Danse JM, Shimohigashi Y, Fromen C, Ducancel F, Zinn-Justin S, Le Du MH, Boulain JC, Tamiya T, Ménez A, *Prog. Nucleic Acid Res. Mol. Biol.*, **59**, 307-364(1998).
7. Betzel C, Genov N, Rajashankar KR, Singh TP, *Cell Mol. Life Sci.*, **56**, 384-897 (1999).
8. Ohno M, Chijiwa T, Oda-Ueda N, Ogawa T, Hattori S, *Toxicon*, **42**, 841-854 (2003).
9. Deshimaru M, Ogawa T, Nakashima K, Nobuhisa I, Chijiwa T, Shimohigashi Y, Fukumaki Y, Niwa M, Yamashina I, Hattori S, Ohno M, *FEBS Lett.*, **397**, 83-88 (1996).

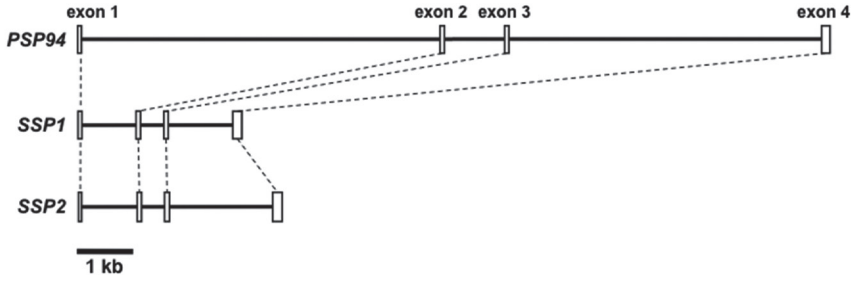

Figure 4. Gene organization of two SSPs and human PSP94. Exonic and Intronic regions are shown with boxes and solid lines, respectively. Corresponding exons are linked each other by broken line.

10. Shimada A, Ohkura N, Hayashi K, Samejima Y, Omori-Satoh T, Inoue S, Ikeda K, *Toxicon*, **51**, 787-796 (2008).
11. Dunn RD and Broady KW, *Biochim. Biophys. Acta*, **1533**, 29-37 (2001).
12. Omori-Satoh T, Sadahiro S, Ohsaka A, Murata R, *Biochim. Biophys. Acta*, **285**, 414-426 (1972).
13. Omori-Satoh T, *Biochim. Biophys. Acta*, **495**, 93-98 (1997).
14. Aoki N, Sakiyama A, Deshimaru M, Terada S, *Biochem. Biophys. Res. Commun.*, **359**, 330-334 (2007).
15. Fujimura S, Oshikawa K, Terada S, Kimoto E, *J. Biochem.*, **128**, 167-173 (2000).
16. Yamazaki Y, Koike H, Sugiyama Y, Motoyoshi K, Wada T, Hishinuma S, Mita M, Morita T, *Eur. J. Biochem.* **269**, 2708-2715 (2002).
17. Aoki N, Matsuo H, Deshimaru M, Terada S, *Gene*, **426**, 7-14 (2008).
18. Blin N and Stafford DW, *Nucleic Acids Res.*, **3**, 2303-2308 (1976).
19. Nei M and Gojobori T, *Mol. Biol. Evol.*, **3**, 418-426 (1986).
20. Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, and Yu J, *Genomics Proteomics Bioinformatics*, **4**, 259-263 (2006).
21. Nolet S, Mbikay M, Chrétien M, *Biochim. Biophys. Acta.*, **1089**, 247-249 (1991).
22. Casewell NR, Wagstaff SC, Harrison RA, Renjifo C, and Wüster W, *Mol. Biol. Evol.*, **28**, 2637-2649 (2011).