Research Article

Unique structure (construction and configuration) and evolution of the array of small serum protein genes of Protobothrops flavoviridis snake *

Takahito Chijiwa1, Kento Inamaru1, Ami Takeuchi1, Marie Maeda1, Kazuaki Yamaguchi1, Hiroki Shibata2, Shosaku Hattori3, Naoko Oda-Ueda4 and Motonori Ohno1

1Department of Applied Life Science, Faculty of Bioscience and Biotechnology, Sojo University, Kumamoto 860-0082, Japan; 2Medical Institute of Bioregulation, Kyushu University, Fukuoka 812-8582, Japan; 3Institute of Medical Science, University of Tokyo, Oshima-gun, Kagoshima 894-1531, Japan; 4Department of Biochemistry, Faculty of Pharmaceutical Sciences, Sojo University, Kumamoto 860-0082, Japan

Correspondence: Takahito Chijiwa (chijiwa@life.sojo-u.ac.jp)

The nucleotide sequence of Protobothrops flavoviridis (Pf) 30534 bp genome segment which contains genes encoding small serum proteins (SSPs) was deciphered. The genome segment contained five SSP genes (PfSSPs), PfSSP-4, PfSSP-5, PfSSP-1, PfSSP-2, and PfSSP-3 in this order and had characteristic configuration and constructions of the particular nucleotide sequences inserted. Comparison between the configurations of the inserted chicken repeat-1 (CR1) fragments of P. flavoviridis and Ophiophagus hannah (Oh) showed that the nucleotide segment encompassing from PfSSP-1 to PfSSP-2 was inverted. The inactive form of PfSSP-1, named PfSSP-1δ (ψ), found in the intergenic region (I-Reg) between PfSSP-5 and PfSSP-1 had also been destroyed by insertions of the plural long interspersed nuclear elements (LINEs) and DNA transposons. The L2 LINE inserted into the third intron or the particular repetitive sequences inserted into the second intron structurally divided five PfSSPs into two subgroups, the Long SSP subgroup of PfSSP-1, PfSSP-2 and PfSSP-5 or the Short SSP subgroup of PfSSP-3 and PfSSP-4. The mathematical analysis also showed that PfSSPs of the Long SSP subgroup evolved alternately in an accelerated and neutral manner, whereas those of the Short SSP subgroup evolved in an accelerated manner. Moreover, the ortholog analysis of SSPs of various snakes showed that the evolutionary emerging order of SSPs was as follows: SSP-5, SSP-4, SSP-2, SSP-1, and SSP-3. The unique interpretation about accelerated evolution and the novel idea that the transposable elements such as LINEs and DNA transposons are involved in maintaining the host genome besides its own transposition natures were proposed.

Introduction

Protobothrops flavoviridis (Pf) (Crotalinae snake) [1] inhabit the southwestern islands of Japan, mainly, Amami-Oshima, Tokunoshima, and Okinawa. The venom of P. flavoviridis contains a huge variety of toxic proteins. The representatives are snake venom metalloproteases (SVMPs) [2], serine proteases [3], phospholipase A2s (PLA2s) [4–7], and triflin [8]. High molecular weight SVMPs, called HR1a and HR1b [9], and middle molecular weight SVMPs, called HR2a and HR2b [10], are isozymes of each other and cause severe hypodermic hemorrhage [11]. Low molecular weight SVMP, called HV1, induces apoptosis of vascular endothelial cells [12]. Serine protease, called Flavoxobin [13,14], is known as coagulant factor. Triflin is a neurotoxin-like protein and blocks muscle contraction [15]. Protobothrops genus snake venom PLA2 isozymes, which hydrolyze phospholipids [16,17], are...
generally classified into four groups, according to the primary structures and physiological activities [18]; that is, neutral [Asp^{49}]PLA_{2}, named PLA2 (pI 7.9, highly lipolytic and myolytic) [19,20]; basic [Asp^{49}]PLA_{2}, named PLA-B (pI 8.6, edema-inducing) [21]; highly basic [Asp^{49}]PLA_{2}, named PLA-N (pI 10.3, neurotoxic) [22]; and two [Lys^{49}]PLA_{2}s, named BPI and BPII (pI 10.2 and 10.3, both weakly lipolytic but strongly myolytic) [19,23,24].

On the other hand, *P. flavoviridis* serum is known to contain inhibitory proteins to neutralize their venomous activities. Such serum proteins are thought to act defensively on occasions of the accidental bites by fellow snakes. PLA_{2} inhibitors (PLIs) which suppress snake venomous PLA_{2} activity [25,26], Habu serum factor (HSF) which inhibits the hemorrhagic induced by the SVMPs [27,28], and small serum proteins (SSPs) [29] are such well-known proteins. SSPs are the low molecular weight (~10 kDa) serum proteins and the five homologs, named PfSSP-1, -2, -3, -4, and PfSSP-5, have been found from *P. flavoviridis* sera [30]. PfSSPs form complexes with HSF in *P. flavoviridis* serum [8] and have various counterparts of *P. flavoviridis* venom proteins; PfSSP-2 and PfSSP-5 show high affinity to triflin [8], PfSSP-1 and PfSSP-4 to HV1 [31], and PfSSP-3 to Flavorase, non-hemorrhagic SVMP [29]. Recently, we found that the homolog of PfSSP-2 binds to BPII [32].

The complementary DNA (cDNA)s encoding five PfSSPs, were isolated and the nucleotide sequences of the cDNAs were also determined [8]. It should be noted that the cDNAs encoding PfSSP-3 or PfSSP-4 are interrupted by nonsense mutations at the same site of the fourth exon and express the truncated mature protein. The genome fragment, in which the genes encoding PfSSP-1 and PfSSP-2, designated as PfSSP-1 and PfSSP-2, respectively, were arranged in tandem, was also cloned. The nucleotide sequence of the genome fragment from PfSSP-1 to PfSSP-2 including the intergenic region (I-Reg) between PfSSP-1 and PfSSP-2, named PfI-Reg12 (in the present paper), was determined [33]. Mathematical analysis of the nucleotide sequences of the two genes showed that they have evolved in an accelerated manner to acquire the different amino acid sequences [8,33].

In the present study, 30534 bp of *P. flavoviridis* genome segment containing five PfSSPs; PfSSP-4, PfSSP-5, PfSSP-1, PfSSP-2, and PfSSP-3 in this order, was deciphered. The particular nucleotide sequences of the long interspersed nuclear elements (LINEs), the DNA transposons, and the repetitive sequences were identified in the introns of PfSSPs and the I-Reg of the array of PfSSPs. The comparison analysis of the configuration of the fragments of chicken repeat-1 (CR1) LINE between *P. flavoviridis* and *Ophiophagus hannah* (Oh) (Elapidae snake) [34] showed that the chromosome inversion of the genome segment encompassing from PfSSP-1 to PfSSP-2 occurred and the inactive form of another PfSSP-1, PfSSP-15(ψ′), was formed at that site. Moreover, according to the types of the nucleotide sequences inserted into the intron of the gene, five PfSSPs were divided into two subgroups. The Long SSP subgroup, which consisted of PfSSP-1, PfSSP-2, and PfSSP-5, contained the large fragment of L2 LINE in the third intron of the gene. The Short SSP subgroup, which consisted of PfSSP-3 and PfSSP-4, contained the particular repetitive sequences in the second intron of the gene and no L2 LINE in the third intron. The mathematical analysis of the nucleotide sequences of the genes also showed that PfSSPs of the Short SSP subgroup evolved in an accelerated manner, whereas those of the Long SSP subgroup evolved alternately in an accelerated and neutral manner. The ortholog analysis of various snake SSPs suggested that the evolutionary emerging order of PfSSPs was reflected in the order of their configuration on the chromosome. In addition to the unique interpretation about accelerated evolution, a novel idea was proposed that the transposable elements such as LINEs and DNA transposons are involved in maintaining the host genome besides their transposition natures.

**Experimental Materials**

*P. flavoviridis* specimen of Amami-Oshima island was provided from the Institute of Medical Sciences of the University of Tokyo. High molecular weight genomic DNAs were prepared from the liver of the snake according to the method of Blin and Stafford [35]. Restriction endonucleases and KOD plus DNA polymerase were purchased from Nippon Gene (Tokyo, Japan) and TOYOBO (Osaka, Japan), respectively. The other reagents and antibiotics were from Nacalai Tesque (Kyoto, Japan) and Takara Bio (Shiga, Japan). Specific oligonucleotide primers were synthesized by GENEQ (Fukuoka, Japan) (Table 1).

**Determination of the nucleotide sequence of *P. flavoviridis* genome segment containing PfSSPs**

From the BLAST analysis against HabAm1 (Habu Amami version 1) [36], we found that PfSSP-5, PfSSP-1, PfSSP-2, and PfSSP-3 are harbored in this order on the scaffold 2858. Furthermore, PfSSP-4, which had not been identified via gene prediction so far, was also found in the 5′ upstream region of this arrangement of four PfSSPs on the scaffold.
Table 1 The primers utilized for acquiring the nucleotide sequence of the genome domain of the array of PfSSPs

| Name   | Positions         | Nucleotide sequence (GC content: %, Tm: °C) |
|--------|-------------------|--------------------------------------------|
| JUS1   | 6973-6993(f)      | 5'-ATT CCT CCC TAC CAA gAg TCT-3'          |
| JUS2   | 11016-11037(f)    | 5'-TCT ATg TgA gAg gAT gAg AAT C-3'        |
| JUS3   | 10961-10983(3)    | 5'-CAT gCC AAC AtG CTA CTT AA-3'          |
| JUS4   | 13451-13473(3)    | 5'-ACC CAC Tgg AAT AAA TTT CTC AT-3'       |
| TOY1   | 1098-1121(f)      | 5'-ggA gTA TCT CTC TAC CgT AAA Tgg-3'      |
| TOY2   | 6964-6985(f)      | 5'-ggA Tgg TgC ACA TCT ggg TgT TTC AA-3'   |
| TOY6   | -220-198(f)       | 5'-ggA gTg CACA TCT ggg TgT TTC AA-3'      |
| TOY7   | 1180-1201(f)      | 5'-TTC CTC CgT gCA gTg TTA gAC C-3'       |
| prs-1  | 29119-29138(3)    | 5'-gAg TgT TCC TCT ACC TAg A3-3'          |
| prs-2  | 26172-26192(3)    | 5'-TTg TCA TCC TAg gAg Tgg-3'             |
| prs-5  | 26342-26364(1)    | 5'-GgC TgT gAg CAC TAg gA gAg AAT gg-3'   |
| prs-6  | 22985-2317(3)     | 5'-AAg gAgC ACC TCT CgT TAa Ag-3'          |
| prs-13 | 30295-30314(3)    | 5'-TTC CTT CgT gCA gTg gAT TC-3'          |
| prs-14 | 29044-29066(f)    | 5'-TTC TCC Tgg CgT TAT TAG ACA-3'         |

The f and r in the parentheses after position numbers indicate the direction of the primers: 'f, forward' or 'r, reverse' means whether the directions of elongations are same or opposite to those of transcriptions, respectively. The nucleotide positions were referenced to the nucleotide sequence reported in the present study (MK574076).

In order to acquire the complete nucleotide sequence of this genome segment, genomic polymerase chain reactions (PCRs) against P. flavoviridis genome were performed with the specific sense and antisense primers referring to the nucleotide sequences of transcripts and genes of PfSSP or HabAm1 (Table 1).

The sense primer, JUS1, 5'-ATT CCT CCC TAC CAA gAg TCT-3', which can anneal specifically to the first exon of PfSSP-5, and the antisense primer, JUS2, 5'-TCT ATg TgA gAg gAT gAg AAT C-3', which can anneal specifically to the fourth exon of PfSSP-5, referring to the nucleotide sequence of the cDNA encoding PfSSP-5 (AB360910), amplified the 4065-bp genome fragment, named Pf-jb-I. The Pf-jb-I fragment was cloned into pCR®-Blunt II-TOPO® vector (Life Technologies, Carlsbad, CA, U.S.A.) and transformed with DH5α competent cells (Takara Bio) and sequenced. The nucleotide sequences were determined with an ABI 3130xl capillary sequencer. The Pf-jb-I was found to encompass from the first exon to the fourth exon of PfSSP-5. In the present study, PfSSP-5 contained extra Val at position 89 encoded by 265GTG and Asp at position 91 encoded by 271GAT was also substituted to Asn encoded by 271AAT.

In order to acquire the nucleotide sequence of the I-Reg between PfSSP-5 and PfSSP-1, named as PfI-Reg51, genomic PCR was carried out against P. flavoviridis genome with the sense primer, JUS5, 5'-CAT gCC AAC ATg AAT CTT ATA gg 3', which can anneal to the fourth exon of PfSSP-5, and the antisense primer, JUS8, 5'-ACC CAC Tgg AAT AAA TTT CTC AT-3', which can anneal to the fourth exon of PfSSP-1, referring to the nucleotide sequence of SSP-1 gene (AB769881), amplified the 2513 bp genome fragment, named Pfjb-II. The Pfjb-II fragment was also cloned and sequenced. The Pfjb-II was found to encompass from the fourth exon of PfSSP-5 to the fourth exon of PfSSP-1 including PfI-Reg51. The 4065 bp Pfjb-I overlapped 77 bp with the 2513 bp Pfjb-II. The physical structure of 6501 bp segment encompassing from the fourth exon of PfSSP-5 to the first exon of PfSSP-1 was deciphered.

As the nucleotide sequences of the first exons of PfSSP-4 and PfSSP-3 are completely identical, it is hard to design the specific primer at the first exon which can amplify each gene differentially. So, the sense primer, TOY6, 5'-ggC TgC ACA TCT ggc TgT TTC AA-3', which can anneal specifically to 220 bp 5' upstream of the first exon of PfSSP-4, and the antisense primer, TOY7, 5'-TTC CTC CTC CgT gCA gTg TTA gAC C-3', which can anneal specifically to the second intron of PfSSP-4, avoiding the nucleotide sequence of the open reading frame (ORF) of PfSSP-4 (AB360909), amplified the 1421 bp genome fragment, named Pfjb-IV. The Pfjb-IV fragment was cloned and sequenced. The Pfjb-IV was found to encompass from 220 bp 5' upstream of the first exon to the second intron of PfSSP-4.

Then, the sense primer, TOY1, 5'-ggA gTA TTT AAC TAC AAG Tgg-3', which can anneal specifically to the second exon of PfSSP-4, and the antisense primer, TOY2, 5'-ggT Agg gAg gAA TTA CgC gAg gA-3', which can anneal specifically to the first exon of PfSSP-5, referring to the nucleotide sequence of the cDNA encoding PfSSP-5 (AB360910), amplified the 5888 bp genome fragment, named Pfjb-III. The Pfjb-III fragment was cloned and sequenced. The Pfjb-III was found to encompass from the second exon of PfSSP-4 to the first exon of PfSSP-5 including the I-Reg between PfSSP-4 and PfSSP-5, named as PfI-Reg45.
The 5888 bp Pfjb-III overlapped 104 bp with the 1421 bp Pfjb-IV. The physical structure of 7205 bp segment encompassing from 220 bp 5′ upstream of the first exon of PfSSP-4 to the first exon of PfSSP-5 was deciphered. The nucleotide sequence of the genome segment containing from PfSSP-1 to PfSSP-2 has been already reported by Tanaka et al. ([33], AB769881). Moreover, the 5888 bp Pfjb-III overlapped 13 bp with the 4065 bp Pfjb-I. The physical structure of 23020 bp segment encompassing from 220 bp 5′ upstream of the first exon of PfSSP-4 to 5′ terminal of the first exon of PfSSP-2, including the nucleotide sequence of the genome segment encompassing from PfSSP-1 to PfSSP-2, was deciphered.

The sense primer, prs-2, 5′-TTg TCA TTC TCT gAg AAg Tgg-3′, which can anneal specifically to 571 bp 3′ downstream of the fourth exon of PfSSP-3, and the antisense primer, prs-1, 5′-gAg TgT TCC TCT ACC TAT Ag-3′, which can anneal specifically to the second exon of PfSSP-3, referring to the nucleotide sequence of the cDNA encoding PfSSP-3 (AB360908), amplified the 2967 bp genome fragment, named Pfjb-V. The Pfjb-V fragment was cloned and sequenced. The Pfjb-V was found to encompass from 571 bp 3′ downstream of the fourth exon to the second exon of PfSSP-3.

Then, the sense primer, prs-6, 5′-AAg AgC AgC ACC TCT CTg TgA Ag-3′, which can anneal specifically to the first exon of SSP-2, and the antisense primer, prs-5, 5′-CgC TTg CAC TgA AgA TgC AAT gg-3′, which can anneal specifically to 378 bp 3′ downstream of the fourth exon of PfSSP-3, avoiding the nucleotide sequence of the ORF of PfSSP-3 (AB360908), amplified the 3470 bp genome fragment, named Pfjb-VI. The Pfjb-VI fragment was cloned and sequenced. The Pfjb-VI was found to encompass from 401 bp 5′ upstream of the fourth exon of PfSSP-3 to the first exon of PfSSP-2. The 2967 bp Pfjb-V fragment overapped 193 bp with the 3470 bp Pfjb-VI.

The sense primer, prs-14, 5′-TTc TCC Tgg GgT TAT AGA Tgg gAT TC-3′, which can anneal specifically to the second intron of PfSSP-3, and the antisense primer, prs-13, 5′-TTc TCC Tgg GgT TAT AGA Tgg gAT TC-3′, which can anneal specifically to 59 bp (the present study) 5′ upstream of the first exon referring to the nucleotide sequence of PfSSP-3 (AB360908), amplified the 1269 bp genome fragment, named Pfjb-VII. The Pfjb-VII fragment was cloned and sequenced. The Pfjb-VII was found to encompass from the second intron to 59 bp 5′ upstream of the first exon of PfSSP-3. The 1269 bp Pfjb-VII overlapped 93 bp with the 2967 bp Pfjb-V. The physical structure of 7420 bp segment encompassing from the first exon of PfSSP-2 to 59 bp 5′ upstream of the first exon of PfSSP-3 was deciphered. The 3470 bp Pfjb-VI overlapped 37 bp with the first exon of PfSSP-2 (AB769881). Finally, the physical structure of 30534 bp segment encompassing from 220 bp 5′ upstream of the first exon of PfSSP-4 to 59 bp 5′ upstream of the first exon of PfSSP-3 was completely deciphered.

The nucleotide sequence and the detailed annotations of the genome domain composed of Pfjb-IV, Pfjb-III, Pfjb-I, Pfjb-II, Pfjb-VI, Pfjb-V and Pfjb-VII, are available in the Genbank/EMBL/DDBJ databases under accession number MK574076.

**Determination of the nucleotide sequences and the chromosomal configurations of the genes encoding the orthologs of PfSSPs of various snakes**

The draft nucleotide sequences of *O. hannah* [34,37], *Python bivittatus* (Pythonidae snake) [38,39], *P. mucrosquamatus* (Viperidae snake) [40,41], and *Thamnophis sirtalis* (Colubridae, Naticeinae snake) [42,43], were downloaded to make personal genome databases. Referring to the nucleotide sequences and the deduced amino acid sequences of PfSSPs via tblastn or blastn, the nucleotide sequences encoding the orthologs of PfSSPs and the flanking regions of them in each snake genome data were deciphered.

**RepeatMasker analysis of the nucleotide sequence of the genome segment harboring SSPs**

The personal database was constructed with the repetitive sequences of the genomes of various organisms collected from the Repbase of the Genetic Information Research Institute [44]. RepeatMasker was carried out the nucleotide sequences of the genome segments containing SSPs of *O. hannah*, *P. bivittatus*, *P. flavoviridis*, *P. mucrosquamatus*, and *T. sirtalis*, against the database via BLAST+, RMBlast (NCBI), and Tandem Repeats Finder (Boston University) [45].

**Mathematical analysis**

Alignment of the amino acid sequences of PfSSPs was performed using ClustalX software. The nucleotide sequences of ORFs encoding the mature proteins of PfSSPs were rearranged, removing the gaps, by PAL2NAL according to the aligned amino acid sequences. The rates of synonymous (\(K_s\)) and nonsynonymous (\(K_a\)) substitutions per site
between the ORFs of the genes were calculated using Nei-Gojobori method as implemented with PAML [46]. The rates of substitutions of the introns ($K_N$) were calculated from the aligned sequence data.

**Results and discussion**

**Peculiar structure of the array of PfSSPs**

The nucleotide sequence of 30534 bp of *P. flavoviridis* genome segment was deciphered as described above and found to contain the array of five PfSSPs, that is, PfSSP-4, PfSSP-5, PfSSP-1, PfSSP-2, and PfSSP-3 in this order (Figure 1A). The precise construction of each of the five PfSSPs including the promoter and four exons was revealed (MK574076) 3733 bp of PfSSP-4, 4198 bp of PfSSP-5, 2796 bp of PfSSP-1, 3619 bp of PfSSP-2, and 3513 bp of PfSSP-3.

The draft genome data of *O. hannah* (Elapidae snake) was also investigated to decipher the nucleotide sequences of the genome segment harboring the orthologs of PfSSPs (see the details in the ‘Determination of the nucleotide sequences and the chromosomal configurations of 221 the genes encoding the orthologs of PfSSPs of various snakes’ subsection of ‘Experimental’ section). Then, four nucleotide sequences of the genes which encode SSP-1, SSP-2, SSP-4, and SSP-5, were identified. As the draft genome data contained many unidentified nucleotides which were described as ‘N’, only the nucleotide sequences of four exons of the gene encoding SSP-4 were identified. The nucleotide sequences of the third exons of the genes encoding SSP-1 and SSP-2 were partly identified, and that of the fourth exon of the gene encoding SSP-5 could not be identified (Figure 1B). The nucleotide sequences encoding SSP-1 and SSP-2 were named as OhSSP-1 and OhSSP-2, respectively. As the nucleotide sequences of the third exons of both genes encoding SSP-4 and SSP-5 contained the insertion of 98 and 8 nucleotides to cause nonsense mutation, they were named as OhSSP-5(ψ) and OhSSP-4(ψ), respectively. It should be noted that the directions of the transcription of OhSSPs were all the same, in contrast with those of PfSSP-4 and PfSSP-5 were opposite to those of the other three PfSSPs, PfSSP-1, PfSSP-2, and PfSSP-3. It is likely that the genome fragment harboring PfSSP-1, PfSSP-2, and PfSSP-3, has been inverted.

**Complicated construction of the I-Reg between PfSSP-5 and PfSSP-1**

Detailed analysis showed that the I-Reg between PfSSP-5 and PfSSP-1, PfI-Reg51, was an interesting structure. First, the nucleotide sequence which encodes another fragmented PfSSP-1 was found in the middle portion of PfI-Reg51. The nucleotide sequence, named PfSSP-16(ψ), consisted of 48 bp of the 3’ portion of the third intron and the fourth exon with five nucleotides of the 3’ untranslated region (UTR) of PfSSP-1 (Figure 1C). Second, the Repeatmasker revealed that the fragments of three types of LINE, L2, R4, and Gypsy, were inserted so as to sandwich the PfSSP-16(ψ) (Figure 1D). The two fragments of L2 and R4 LINEs were located in the 3’ downstream of PfSSP-16(ψ) and that of Gypsy LINE was located in the 5’ upstream of PfSSP-16(ψ). Each of the three fragments encoded most of the reverse transcriptase (RT) domain of each LINE.

Two DNA transposon fragments hAT [47], Mariner [48], and another Gypsy LINE fragment were inserted in the region between PfSSP-5 and the L2 fragment-R4 fragment-PfSSP-16(ψ) arrangement. Both DNA transposons are known to carry out gene conversion via double-strand break [49,50]. In addition, the 30 bp nucleotide sequence predicted to form the stem-loop structure (Figure 1E) was found immediately next to the L2 fragment-R4 fragment-PfSSP-16(ψ) arrangement. The stem-loop structure is also known to be the scaffolding of the gene conversion [51,52]. PfSSP-16(ψ) should be the remnant of the amplified PfSSP-1 which was destroyed by the plural times of insertions of LINEs and DNA transposons after being amplified into PfI-Reg51.

**Chromosome inversion interrupted the array of PfSSPs**

Further investigation of the nucleotide sequences of the arrays of SSPs of *P. flavoviridis* and *O. hannah* showed that there were two pairs of the particular nucleotide sequences. One pair was 140 nucleotides in the 3’ downstream of PfSSP-1 and 140 nucleotides in the 3’ downstream of OhSSP-1, the other pair was 937 nucleotides in the 5’ upstream of PfSSP-2 and 961 nucleotides in the 5’ upstream of OhSSP-2 (Figure 2A,B). The nucleotide sequence of the former pair was designated as ‘α’ and that of the latter pair as ‘β’. The identity between the nucleotide sequences of α or β pairs was 69 or 65%, respectively, but the direction of the nucleotide sequences of the α or β pairs was opposite. These findings showed that the *P. flavoviridis* genome segment encompassing from the α sequence to the β sequence had been inverted, moreover the tandem arrangement of PfSSP-1 and PfSSP-2 had already been formed before the inversion occurred.

Interestingly, RepeatMasker analysis also showed the five fragments of CR1 LINE, which is the most major LINE contained in the reptilian genome, were found in all the PfI-Regs except for PfI-Reg51. They are CR1_45, the CR1 fragment inserted in the PfI-Reg45, and CR1_12_j and CR1_12_i, or CR1_23_j and CR1_23_i, the CR1 fragments
Figure 1. Detailed analysis of the genome segments of *P. flavoviridis* and *O. hannah* containing SSPs

(A) The schematic representation of *P. flavoviridis* 30534 bp genome segment containing the array of six *PfSSPs* including the fragmented *PfSSP-1, PfSSP-1*(ψ)**. Bold and hatched broken arrows indicate the areas and the directions of the transcription of the genes in the segment. Gray bars represent exons.

(B) The schematic representation of *O. hannah* 17678 bp (except for 4745 ‘N’s) genome fragment containing the array of four *OhSSPs*. ‘N’ means the unidentified nucleotides.

(C) The alignment of the nucleotide sequences of *PfSSP-1*(ψ) and the corresponding portion of *PfSSP-1*. The portions of introns are shown in lower case letters and those of exons are shown in upper case letters. The stop codons are underlined.

(D) The schematic configuration of the fragments of LINEs and DNA transposons inserted into the I-Reg *PfI-RegS1*. Three closed ellipses, open ellipse, and specked ellipse represent the inserted fragments of LINEs, Gypsy, L2, and R4, respectively. Closed and open stars represent the inserted DNA transposons, hAT and Mariner, respectively. The position of the stem-loop structure is also shown.

(E) The predicted stem-loop structure of the 30 nucleotides located between the fragment of Gypsy and the L2 fragment-R4 fragment-*PfSSP-1*(ψ) arrangement. The secondary structure is deduced based on the nucleotide sequence by RNA secondary structure prediction of GENETYX ver. 16. The numerals at both termini of the sequence are the position numbers of the corresponding nucleotides reported in the present study (MK574076). Abbreviation: UTR, untranslated region.
Figure 2. Comparison of the genome segments of *P. flavoviridis* and *O. hannah* containing SSPs

Comparison of the schematic configuration of (A) the array of OhSSPs, (B) the array of PfSSPs, and (C) the temporary array of PfSSPs in the case where the chromosome inversion did not occur. White and black daggers represent the positions and the directions of the nucleotide sequences of α and β, which are linked with dashed lines to each other. Harpoons represent the positions and the directions of the transcription of the inserted fragments of CR1 LINE.

contained only the CTCR of the RT domain. Interestingly, the direction of the transcription of the fragments CR1\_12\_i, CR1\_12\_ii, and CR1\_23\_i, was opposite to that of the transcription of the fragments CR\_45 and CR\_23\_ii (Figure 2B). Namely, these findings further showed that the inversion of the genome segment encompassing from the α sequence with CR1\_23\_i to the β sequence at the 3′ terminal of *PfSSP-1* had occurred. If the inversion had not occurred, the direction of the transcription of all five CR1 fragments would be the same and the five CR1 fragments should have been located 3′ downstream of all *PfSSP* except for *PfSSP-3* (Figure 2C). Ikeda et al. (2010) [57] also found that the genes encoding *P. flavoviridis* venom PL2 isozymes are linked to the fragments of CR1 LINE, named PLA2 gene-coupled RT fragment (PcRTF), in their 3′ downstream. Thus, CR1 LINE seems to be involved into the amplified genes in *P. flavoviridis* genome.

**Particular nucleotide sequences inserted in the genes classified *PfSSP* into two subgroups**

Figure 3 showed the schematic configuration of the nucleotide sequences inserted into five *PfSSP*. The nucleotide sequences of the fragments of L1 and CR1 LINEs were inserted at the same sites of the first intron of all five *PfSSP* and those of two fragments of Gypsy, named Gypsy-i and Gypsy-ii, were also inserted at the same sites of the 3′ terminal of the third intron of all five *PfSSP*. The nucleotide sequence of the fragment of Mariner, named Mariner-ii, was inserted at the same sites in the middle portion of the third intron of the four *PfSSP* except for *PfSSP-1*. The identities of the nucleotide sequences of the five inserted fragments were considerably high (Table 2). They must have already been inserted into the gene prior to the amplification of *PfSSP*.

On the other hand, the types of the nucleotide sequences inserted in the second or third intron of the gene classified five *PfSSP* into two subgroups. One subgroup consisted of three *PfSSP*, *PfSSP-1*, *PfSSP-2* and *PfSSP-5*, was characterized by the nucleotide sequence of the fragment inserted into the third intron of the gene, which encoded the RT domain of L2 LINE [56,58]. As the three SSPs belonging to this subgroup encoded the full-length proteins [8], this subgroup was designated as the Long SSP subgroup. Interestingly, the inserted fragments were truncated according to the order of the name of each gene. The lengths and the constructions of the three inserted fragments were as follows. The fragment inserted into *PfSSP-5* was 1011 bp, which encoded nine subdomains from 0 to VIII of RT domain. The fragment inserted into *PfSSP-2* was 431 bp, which encoded four subdomains from 0 to III of RT domain. The fragment inserted into *PfSSP-1* was 320 bp, which encoded three subdomains from 0 to II of RT domain.
Figure 3. Schematic representation of the configurations of the fragments of LINEs and DNA transposons inserted into *PfSSP*s.

Gray bars represent exons. Half closed, hatched, open, and closed ellipses represent the fragments of LINEs, L1, CR1, L2, and Gypsy. Open and closed stars represent the fragments of DNA transposons, Mariners and hAT. The positions of the corresponding fragments are linked with dashed lines to each other. The inserted positions of the repetitive sequences of (TAAAA) and (AATAA) are indicated by the carets and the numbers of repetitions of them are also shown as the subscribed suffixes.

domain. Though LINEs are known to be generally truncated from the 5' terminal region and become inactive, the fragments of L2 LINE inserted into *PfSSP*-1, *PfSSP*-2, and *PfSSP*-5 truncated from the 3' terminal. The first *PfSSP* in this subgroup should have been *PfSSP*-5 with the inserted L2 LINE fragment. As the amplifications occurred from *PfSSP*-5 to *PfSSP*-2 and then from *PfSSP*-2 to *PfSSP*-1, the inserted L2 fragment became truncated every time at each. The nucleotide sequence between L2 fragment and Mariner-ii in the third intron of *PfSSP*-2 and *PfSSP*-5 was considered to be an irrelevant nucleotide sequence brought in from the genome site where L2 LINE had been retro-transposed just before. In *PfSSP*-1, this 'orphan' nucleotide sequence is thought to have disappeared accompanying the transposition of Mariner-ii.

The other subgroup consisted of *PfSSP*-3 and *PfSSP*-4 was characterized by three nucleotide sequences of the fragments of DNA transposons inserted into the second and third introns of the gene. One was the fragment of Mariner, named as Mariner-i, inserted into the same site of the second intron of the gene. The other two were the fragments of hAT and another Mariner, named as Mariner-iii. Two juxtaposed fragments were inserted into the same site between Mariner-ii and Gypsy-i in the third intron of the gene. In addition, the particular repetitive nucleotide sequences were also found at the same site of the second intron of the gene (see the details in the next section). As *PfSSP*-3 and *PfSSP*-4 encoded the truncated proteins [8], the subgroup was designated as the Short SSP subgroup. The positions and the nucleotide sequences of the eight inserted fragments, L1 and CR1 fragments in the first intron, Mariner-i in the second intron, Mariner-ii, hAT, Mariner-iii, Gypsy-i and Gypsy-ii in the third intron were almost the same (Table 2). Namely, the insertion of them should have occurred before the branching of *PfSSP*-3 and *PfSSP*-4 and not much time must have passed since two genes were branched.
Table 2 The identities of the nucleotide sequences between the fragments of transposable elements inserted into the introns of PfSSPs

| PfSSP-1 | PfSSP-2 | PfSSP-5 | PfSSP-3 | PfSSP-4 |
|---------|---------|---------|---------|---------|
| **<The fragments of L1 LINE>** |         |         |         |         |
| PfSSP-1 (125 bp) | 100      | 66      | 65      | 65      |
| PfSSP-2 (125 bp) | 66       | 65      | 65      |         |
| PfSSP-5 (142 bp) |         | 63      | 63      |         |
| PfSSP-3 (152 bp) |         |         | 100     |         |
| PfSSP-4 (152 bp) |         |         |         |         |
| **<The fragments of CR1 LINE>** |         |         |         |         |
| PfSSP-1 (42 bp) | 97       | 83      | 81      | 81      |
| PfSSP-2 (42 bp) | 81       | 79      | 79      |         |
| PfSSP-5 (43 bp) |         | 93      | 93      |         |
| PfSSP-3 (43 bp) |         |         | 100     |         |
| PfSSP-4 (43 bp) |         |         |         |         |
| **<The fragments of L2 LINE>** |         |         |         |         |
| PfSSP-1 (320 bp) | 94       | 87      |         |         |
| PfSSP-2 (431 bp) |         | 86      |         |         |
| PfSSP-5 (1010 bp) |         |         |         |         |
| **<The fragments of Mariner-i>** |         |         |         |         |
| PfSSP-3 (54 bp) |         |         |         | 100     |
| PfSSP-4 (54 bp) |         |         |         |         |
| **<The fragments of Mariner-ii>** |         |         |         |         |
| PfSSP-2 (57 bp) | 92       | 82      | 82      |         |
| PfSSP-5 (56 bp) |         | 85      | 85      |         |
| PfSSP-3 (56 bp) |         |         | 100     |         |
| PfSSP-4 (56 bp) |         |         |         |         |
| **<The fragments of Mariner-iii>** |         |         |         |         |
| PfSSP-3 (55 bp) | 94       |         |         |         |
| PfSSP-4 (55 bp) |         |         |         |         |
| **<The fragments of hAT>** |         |         |         |         |
| PfSSP-3 (57 bp) |         |         |         | 96      |
| PfSSP-4 (57 bp) |         |         |         |         |
| **<The fragments of Gypsy-i>** |         |         |         |         |
| PfSSP-1 (72 bp) | 100      | 76      | 81      | 80      |
| PfSSP-2 (72 bp) | 76       | 81      | 80      |         |
| PfSSP-5 (72 bp) |         | 67      | 68      |         |
| PfSSP-3 (72 bp) |         |         | 94      |         |
| PfSSP-4 (72 bp) |         |         |         |         |
| **<The fragments of Gypsy-ii>** |         |         |         |         |
| PfSSP-1 (25 bp) | 92       | 76      | 68      | 76      |
| PfSSP-2 (25 bp) | 76       | 60      | 68      |         |
| PfSSP-5 (37 bp) |         | 68      | 76      |         |
| PfSSP-3 (38 bp) |         |         | 92      |         |
| PfSSP-4 (39 bp) |         |         |         |         |

The lengths of the fragments, from which the indels (inserted or deleted fragments) are excluded, are described in the parentheses.

Different evolutionary path that PfSSPs of two subgroups followed

The evolutionary process of the Long SSP subgroup was not plain. The mathematical analysis showed that the branching between PfSSP-1 and PfSSP-2 of the Long SSP subgroup had occurred in an accelerated manner (Table 3) [8,33]. In addition, the fact that the rate of $K_N$ between the introns of PfSSP-1 and PfSSP-2 was 0.0649 also suggested that the time passed after branching of PfSSP-1 and PfSSP-2 was very short (Table 4). On the other hand, the rate of $K_K/K_S$ between the ORFs of PfSSP-1 and PfSSP-5 or PfSSP-2 and PfSSP-5, which is the relative ratio of synonymous substitution rate to nonsynonymous substitution rate, was 0.625 or 0.646 (Table 3) and the rate of $K_N$ between the
introns of PfSSP-1 and PfSSP-5 or PfSSP-2 and PfSSP-5 was 0.328 or 0.312, respectively (Table 4). These results suggested that PfSSP-1 or PfSSP-2 and PfSSP-5 had been diverged in a ‘neutral manner’ long time ago and that PfSSP-1 and PfSSP-2 branched in a very short time long time later.

As concerning about the Short SSP subgroup, the nucleotide sequence between PfSSP-3 and PfSSP-4 including the fragments of LINEs and DNA transposons was almost identical except for the number of repetition of the nucleotide sequences in 2nd intron (Figure 3). The repetitive sequences were of two types. One was the repetition of five nucleotides of TAAAA, which was repeated 32 times for PfSSP-3 and 36 times for PfSSP-4. The other was that of five nucleotides of AATAA immediately next to the repeat of TAAAA, which was repeated 42 times only for PfSSP-3. Without consideration of the repeats, the rates of \( K_A/K_S \) and \( K_N \) estimated between the introns of PfSSP-3 and PfSSP-4 were 1.54 (Table 3) and 0.0488 (Table 4), respectively. These results showed that PfSSP-3 and PfSSP-4 has been branched very recently in an accelerated manner.

### Orthologs of SSPs from various snakes

Detailed tblastx analysis against the draft genome databases of four snakes P. bivittatus, T. sirtalis, O. hannah, P. mucrosquamatus, in addition to that of P. flavoviridis has revealed the orthologs for PfSSPs and their chromosomal configurations as follows (Figure 4). The genome of non-venomous Pythonidae snake, P. bivittatus, contained three orthologs of PfSSP-5, named PbSSP-5\( \alpha \), PbSSP-5\( \beta \), and PbSSP-5\( \gamma \), that of Colubridae snake, T. sirtalis, contained the ortholog of PfSSP-4, named TsSSP-4, and two orthologs of PfSSP-5, named TsSSP-5\( \alpha \), and TsSSP-5\( \beta \), that of neurotoxic Elapidae snake, O. hannah, contained the orthologs of PfSSP-4, PfSSP-5, PfSSP-1, and PfSSP-2 on one chromosome in this order, named OhSSP-4\( \gamma \), OhSSP-5\( \gamma \), OhSSP-2, and OhSSP-1, respectively (see the details in the first section of this chapter), that of Viperidae Taiwan Habu snake, P. mucrosquamatus, contained the orthologs of PfSSP-5, PfSSP-1, PfSSP-2 and PfSSP-3 on one chromosome in this order, named PmSSP-5, PmSSP-1, PmSSP-2, and PmSSP-3, respectively (in the present study). In addition, the ortholog of PfSSP-4, named PmSSP-4, was also found in P. mucrosquamatus in another scaffold.

In the present study, the names of four orthologs were renamed. As a result of our detailed investigation based on the deduced amino acid sequences, the nucleotide sequences which have been annotated as PbSSP-2 and TsSSP-2 in the original databases should be renamed as PbSSP-5\( \beta \) and TsSSP-5\( \beta \). In addition, the nucleotide sequence newly found from the genome of P. bivittatus in the present study, which encoded the ortholog of PfSSP-5 but contained the deletions of 34 nucleotides and 7 nucleotides causing frameshifts at the second and third exons, respectively. It was named as PbSSP-5\( \gamma \). Therefore, those already annotated as PbSSP-5 and TsSSP-5 in the original databases should be named as PbSSP-5\( \alpha \) and TsSSP-5\( \alpha \), respectively. The relationship between PbSSP-5\( \alpha \), PbSSP-5\( \beta \) and PbSSP-5\( \gamma \), or TsSSP-5\( \alpha \) and TsSSP-5\( \beta \) was paralog. Furthermore, the nucleotide sequence, which was newly found from the genome of T. sirtalis in the present study, encoded the ortholog of PfSSP-4 then was named as TsSSP-4.
Evolutionary emerging order of SSPs analyzed from the constructions and configurations of SSP genes

The configuration of the paralogs of SSPs in each snake genome (Figure 4) seemed to show their emerging order. Before the branching of Colubridae and Booidae snakes, SSP-5 had already existed in advance. After the branching, the genome of Colubridae snakes acquired SSP-4 derived from the paralog of SSP-5. On the other hand, the formation of the paralog of SSP-5 occurred twice in P. bivittatus genome or once in T. sirtalis genome. They became PbSSP-5β and PbSSP-5g(ψ) or TsSSP-5β, respectively. In the genome of Elapidae and Viperidae snakes, the derivation of the paralogs of SSP-5 occurred twice at least and then they have become SSP-1 and SSP-2. The comparative analysis of the construction of the L2 LINE fragment in the third intron of the gene has already showed that SSP-5, SSP-2, and SSP-1 appeared in this order (see details in the fourth section of the present study). And then, in the genome of Viperidae, the inversion of the genome segment encompassing from SSP-2 to SSP-1 occurred (see details in the third section of the present study). Interestingly, the mathematical analysis between SSP-1 and SSP-2 showed that those of Elapidae snake have been evolved in a neutral manner in contrast with those of Viperidae snakes have been evolved in an accelerated manner. It is only speculation, the nucleotide substitutions dominant at the nonsynonymous sites only occurred immediately after duplication and then random mutations accumulated over time and the selective pressure to preserve the ‘neutral’ mutations at the synonymous sites have reduced the traces of the accelerated evolution. But the inversion of Viperidae genome segment encompassing from PfSSP-1 to PfSSP-2 might have avoided the accumulation of random mutations. The emerging process of the most newcomer, SSP-3, which is structurally highly related to SSP-4, is an issue to be addressed in the next study. It is also interesting that the positions and the nucleotide sequences of the fragments of LINEs and DNA transposons seem to be conserved rather than the nucleotide sequences of the introns and the I-Regs in which those transposable elements are inserted. It is likely that such transposable elements are involved in maintaining the construction of the host genome besides their transposition natures.

Author Contribution

Takahito Chijiwa, Ami Takeuchi and Marie Maeda designed experiments. Takahito Chijiwa, Kento Inamaru, Ami Takeuchi, Marie Maeda, Kazuaki Yamaguchi, Hiroki Shibata, Shousaku Hattori, Naoko Oda-Ueda and Motonori Ohno prepared materials from live specimens. Ami Takeuchi and Marie Maeda performed sequencing and manual assembly of SSP gene array. Hiroki Shibata performed whole genome sequencing and provided HabAm1. Kento Inamaru, Ami Takeuchi, Marie Maeda and Kazuaki Yamaguchi performed bioinformatics analyses. Takahito Chijiwa supervised the project. Takahito Chijiwa and Kento Inamaru wrote the manuscript with contributions from all the other authors.

© 2019 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY).

Figure 4. The relationship between the phylogenetic clade of snakes and the schematic structure of the configuration of the genes encoding SSPs

Gray bars represent exons. Open ellipses represent the fragments of the inserted L2 LINEs. Orthologous SSP-5 genes in each snake genome were linked with dashed lines to each other. The genome segments encompassing from SSP-1 to SSP-2 of O. hannah, P. mucrosquamatus, and P. flavoviridis were linked with dotted lines to each other. Abbreviations: Oh, O. hannah; Pb, P. bivittatus; Pm, P. mucrosquamatus; Ts, T. sirtalis.
Funding
This work was supported in part by the Ministry of Education, Science, Sports and Culture, Grant-Aid for Science Research (C), 2014–2016 of Japan [grant number 26340095 (to Takahito Chijiiwa)].

Competing Interests
The authors declare that there are no competing interests associated with the manuscript.

Abbreviations
cDNA, complementary DNA; CR1, chicken repeat-1; CTCR, carboxy-terminal conserved region; HabAm1, Habu Amami version 1; hAT, hobo-Ac-Tam3; HR1, hemorrhagic factor 1; HR2, hemorrhagic factor 2; I-Reg, intergenic region; LINE, long interspersed nuclear element; Oh, Ophiophagus hannah; ORF, open reading frame; PAML, Phylogenetic Analysis by Maximum Likelihood; PCR, polymerase chain reaction; Pf, Protobothrops flavoviridis; PLA2, phospholipase A2; RT, reverse transcriptase; SSP, small serum protein; SVMP, snake venom metalloprotease; UTR, untranslated region.

References
1 Hallowell, E. (1861) Report upon the Reptilia of the North Pacific Exploring Expedition, under command of Capt. John Rogers, U. S. N. Proc. Acad. Nat. Sci. Philadelphia 12, 480–510
2 Fox, J.W. and Serrano, S.M. (2005) Structural considerations of the snake venom metalloproteinas, key members of the M12 reprolysin family of metalloproteinases. Toxicon 45, 969–985, https://doi.org/10.1016/j.toxicon.2005.02.012
3 Matsui, T., Fujimura, Y. and Titani, K. (2000) Snake venom proteases affecting hemostasis and thrombosis. Biochim. Biophys. Acta 1447, 146–156, https://doi.org/10.1016/S0005-2795(99)00268-X
4 Brunie, S., Bolin, J., Gewirth, D. and Sigler, P.B. (1985) The refined crystal structure of dimeric phospholipase A2 at 2.5 Å Access to a shielded catalytic center. J. Biol. Chem. 260, 9742–9749
5 Holland, D.R., Clancy, L.L., Muchmore, S.W., Ryde, T.J., Einspahr, H.M., Finzel, R.L. et al. (1990) The crystal structure of a lysine 49 phospholipase A2 from the venom of the cottonmouth snake at 2.0 Å resolution. J. Biol. Chem. 265, 17649–17656
6 Renetseder, R., Brunie, S., Dijkstra, B.W., Drenth, J. and Sigler, P.B. (1985) A comparison of the crystal structures of phospholipase A2 from bovine pancreas and Crotalus atrox venom. J. Biol. Chem. 260, 11627–11634
7 Suzuki, A., Matsueda, E., Yamane, T., Ashida, T., Kihara, H. and Ohno, M. (1995) Crystal structure analysis of phospholipase A2 from Trimeresurus flavoviridis (Habu Snake) venom at 1.5 Å resolution. J. Biol. Chem. 267, 730–740, https://doi.org/10.1093/oxfordjournals.jbchem.a124770
8 Aoki, N., Sakiyama, A., Kurki, K., Maenaka, K., Kohda, D., Deshimaru, M. et al. (2008) Serotinin, a CRISP family protein with binding affinity for small serum protein-2 in snake serum. Biochim. Biophys. Acta 1784, 621–628, https://doi.org/10.1016/j.bbagrm.2007.12.010
9 Omori-Saitoh, T. and Sadahiro, S. (1979) Resolution of the major hemorrhagic component of Trimeresurus flavoviridis venom into two parts. Biochim. Biophys. Acta 580, 392–404, https://doi.org/10.1016/0005-2795(79)90151-X
10 Takahashi, T. and Ohsaka, A. (1970) Purification and some properties of two hemorrhagic principles (HR2a and HR2b) in the venom of Trimeresurus flavoviridis; complete separation of the principles from proteolytic activity. Biochim. Biophys. Acta 207, 65–75, https://doi.org/10.1016/0005-2795(70)90137-6
11 Kini, R.M. (2003) Excitement ahead: structure, function and mechanism of snake venom phospholipase A2 enzymes. Toxicon 42, 827–840, https://doi.org/10.1016/S0041-0101(03)001022
12 Matsuda, S., Hayashi, T., Atobe, H., Morita, T. and Araki, S. (2001) Purification, cDNA cloning and characterization of the vascular apoptosis-inducing protein, HV1, from Trimeresurus flavoviridis. J. Biochem. 130, 3339–3345
13 Shieh, T.C., Tanaka, S., Kihara, H., Ohno, M. and Makisumi, S. (1985) Purification and characterization of a coagulant enzyme from Trimeresurus flavoviridis venom. J. Biochem. 98, 713–721, https://doi.org/10.1093/oxfordjournals.jbchem.a135329
14 Shieh, T.C., Kawabata, S., Kihara, H., Ohno, M. and Iwanaga, S. (1988) Amino acid sequence of a coagulant enzyme, flavoxobin, from Trimeresurus flavoviridis venom. J. Biochem. 103, 596–605, https://doi.org/10.1093/oxfordjournals.jbchem.a122313
15 Yamazaki, Y., Koke, H., Sugiyama, Y., Motoyoshi, K., Wada, T., Hishinuma, S. et al. (2002) Cloning and characterization of novel snake venom proteins that block smooth muscle contraction. Eur. J. Biochem. 269, 2708–2715, https://doi.org/10.1046/j.1432-1033.2002.02940.x
16 Dijkstra, B.W., Renetseder, R., Kalk, K.H., Hol, W.G., Drenth, J. and Huber, R. (1983) Structure of porcine pancreatic phospholipase A2 at 2.6 Å resolution and comparison with bovine phospholipase A2. J. Mol. Biol. 168, 163–179, https://doi.org/10.1016/0022-2836(83)80326-3
17 Dijkstra, B.W., Kalk, K.H., Hol, W.G. and Drenth, J. (1981) Structure of bovine pancreatic phospholipase A2 at 1.7 Å resolution. J. Mol. Biol. 147, 97–123, https://doi.org/10.1016/0022-2836(81)90081-4
18 Chijiiwa, T., Abe, K., Ogawa, T., Nikandrov, N.N., Hattori, S., Oda-Ueda, N. et al. (2005) Amino acid sequence of a basic aspartate-49-phospholipase A2 from Trimeresurus flavoviridis venom and phylogenetic analysis of Crotalinae venom phospholipases A2. Toxicon 46, 185–195, https://doi.org/10.1016/j.toxicon.2004.05.004
19 Kihara, H., Uchikawa, R., Hattori, S. and Ohno, M. (1992) Myotoxicity and physiological effects of three Trimeresurus flavoviridis phospholipases A2. Biochem. Int. 28, 895–903
20 Oda, N., Ogawa, T., Ohno, M., Sasaki, H., Sakaki, Y. and Kihara, H. (1990) Cloning and sequence analysis of cDNA for Trimeresurus flavoviridis phospholipase A2, and consequent revision of the amino acid sequence. J. Biochem. 108, 816–821, https://doi.org/10.1093/oxfordjournals.jbchem.a123286
21 Yamaguchi, Y., Shimohigashi, Y., Chijiwata, T., Nakai, M., Ogawa, T., Hattori, S. et al. (2001) Characterization, amino acid sequence and evolution of edema-inducing, basic phospholipase A2 from Trimeresurus flavoviridis venom. Toxicon 39, 1069–1076, https://doi.org/10.1016/S0041-0101(00)00250-6

22 Chijiwata, T., Hamai, S., Tsubouchi, S., Ogawa, T., Deshimaru, M., Oda-Ueda, N. et al. (2003) Interisland mutation of a novel phospholipase A2 from Trimeresurus flavoviridis venom and evolution of crotalidene group II phospholipases A2. J. Mol. Evol. 57, 546–554, https://doi.org/10.1007/s00239-003-2508-4

23 Liu, S.Y., Yoshizumi, K., Oda, N., Ohno, M., Tokunaga, F., Iwanaga, S. et al. (1999) Purification and amino acid sequence of basic protein II, a Lysine-49-Phospholipase A2 with low activity, from Trimeresurus flavoviridis venom. J. Biochem. 107, 400–408, https://doi.org/10.1093/oxfordjournals.jbchem.a123057

24 Yoshizumi, K., Liu, S.Y., Miyata, T., Saita, S., Ohno, M., Iwanaga, S. et al. (1990) Purification and amino acid sequence of basic protein I, a Lysine-49-phospholipase A2 with low activity, from the venom of Trimeresurus flavoviridis (Habu snake). Toxicon 28, 43–54, https://doi.org/10.1016/0041-0101(90)90005-R

25 Nobuhisa, I., Deshimaru, M., Chijiwata, T., Nakashima, K., Ogawa, T., Shimohigashi, Y. et al. (1997) Structures of genes encoding phospholipase A2 inhibitors from the venom of Trimeresurus flavoviridis. Gene 191, 31–37, https://doi.org/10.1016/S0378-1119(97)00204-3

26 So, S., Chijiwata, T., Ikeda, N., Nobuhisa, I., Oda-Ueda, N., Hattori, S. et al. (2008) Identification of the B subtype of γ-Phospholipase A2 inhibitor from Protobothrops flavoviridis venom and molecular evolution of snake venom phospholipase A2 inhibitors. J. Mol. Evol. 66, 298–307, https://doi.org/10.1007/s00239-008-9089-1

27 Yamakawa, Y. and Omori-Satoh, T. (1992) Primary structure of the anehemorrhagic factor in serum of the Japanese Habu: a snake venom metalloproteinase inhibitor with a double-headed cystatin domain. J. Biochem. 112, 583–589, https://doi.org/10.1093/oxfordjournals.jbchem.a123944

28 Deshimaru, M., Tanaka, C., Tokunaga, A., Goto, M. and Terada, S. (2003) Efficient purification of an anehemorrhagic factor (HSV) in serum of Japanese Habu (Trimeresurus flavoviridis). Fukuoka Univ. Sci. Rep. 33, 45–53

29 Shioi, N., Nishijima, A. and Terada, S. (2015) Flavorase, a novel non-haemorrhagic metalloproteinase in Protobothrops flavoviridis venom, is a target molecule of small protein-3. J. Biochem. 158, 37–48, https://doi.org/10.1093/jb/mnv017

30 Aoki, N., Matsuo, H., Deshimaru, M. and Terada, S. (2008) Accelerated evolution of small serum proteins (SSPs)–the PSP94 family proteins in a Japanese viper. Gene 426, 7–14, https://doi.org/10.1016/j.gene.2008.08.021

31 Shioi, N., Ogawa, E., Mizukami, Y., Abe, S., Hayashi, R. and Terada, S. (2013) Small serum protein-1 changes the susceptibility of an apoptosis-inducing metalloproteinase HV1 to a metalloproteinase inhibitor in habu snake (Trimeresurus flavoviridis). J. Biochem. 153, 121–129, https://doi.org/10.1093/jb/mptw127

32 Chijiwata, T., So, S., Hattori, S., Yoshida, A., Oda-Ueda, N. and Ohno, M. (2013) Suppression of severe lesions, myonecrosis and hemorrhage, caused by Protobothrops flavoviridis venom with its serum proteins. Toxicon 76, 197–205, https://doi.org/10.1016/j.toxicon.2013.10.007

33 Tanaka, Y., Shioi, N., Terada, S. and Deshimaru, M. (2013) Structural organization and evolution of a cluster of small serum protein genes of Protobothrops flavoviridis snake. Fukuoka Univ. Sci. Rep. 43, 59–66

34 Cantor, T.E. (1836) Sketch of undescribed hooded serpent with fangs and maxillar teeth. Asiat. Res. Calculta 19, 87–94

35 B hiding, N. and Stafford, D.W. (1976) A general method for isolation of high molecular weight DNA from eukaryotes. Nucleic Acids Res. 3, 2303–2308, https://doi.org/10.1093/nar/3.9.2303

36 Shibata, H., Chijiwata, T., Oda-Ueda, N., Nakamura, H., Yamaguchi, K., Hattori, S. et al. (2018) The habu genome reveals accelerated evolution of venom protein genes. Sci. Rep. 8, 11300, https://doi.org/10.1038/s41598-018-28749-4

37 Vonk, F.J., Casewell, N.R., Henkel, C.V., Heimberg, A.M., Janssen, H.J., McCleary, R.J. et al. (2013) The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. Proc. Natl. Acad. Sci. U.S.A. 110, 20651–20656, https://doi.org/10.1073/pnas.1314702110

38 Kuhl, H. (1820) Beiträge zur Zoologie und vergleichenden Anatomie, Hermannsche Buchhandlung, Frankfurt

39 Castoe, T.A., de Koning, A.P., Hall, K.T., Card, D.C., Schield, D.R., Fujita, M.K. et al. (2013) The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. Proc. Natl. Acad. Sci. U.S.A. 110, 20645–20650, https://doi.org/10.1073/pnas.1314745110

40 Cantor, T.E. (1839) Spicilegium serpentium indicorum. Proc. Zool. Soc. Lond. 7, 31–34

41 Aird, S.D., Arora, J., Barua, A., Qiu, L., Terada, K. and Mikeyev, A.S. (2017) Population genomic analysis of a pitviper reveals microevolutionary forces underlying venom chemistry. Genome Biol. Evol. 9, 2640–2649, https://doi.org/10.1093/gbe/evx199

42 Linnaeus, C. (1758) Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis

43 Warren, W.C. and Wilson, R.K. (2015) Direct Submission, Genome Institute. Washington. Univ. Sch. Med., 4444 For. Park. St. Louis, MO 63108, U.S.A.

44 Smit, A., Hubley, R. and Green, P. (1996) RepeatMasker-open-3.0 [WWW Document] RepeatMasker-open3.0., http://repeatmasker.org

45 Benon, G. (1999) Tandem reverse finder: a program to analyze DNA sequences. Nucleic Acids Res. 27, 573–580, https://doi.org/10.1093/nar/27.2.573

46 Yang, Z. (2007) PAML 4. Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol. 24, 1586–1591, https://doi.org/10.1093/molbev/msm088

47 Warren, W.D., Atkinson, P.W. and O’Brochta, D.A. (1994) The Hermes transposable element from the house fly, Musca domestica, is a short inverted repeat-type element of the hobo, Ac, and Tam3 (hAT) element family. Genet. Res. 64, 87–97, https://doi.org/10.1017/S0016672300032699

48 Robertson, H.M. (1993) The mariner transposable element is widespread in insects. Nature 362, 241–245, https://doi.org/10.1103/362241a10

49 Vo, J.C., De Baere, I. and Plasterk, R.H. (1996) Transposase is the only nuclease protein required for in vitro transposition of Tc1. Genes Dev. 10, 755–761, https://doi.org/10.1101/gad.10.6.755

50 Zhou, L., Mitra, R., Atkinson, P.W., Hickman, A.B., Dyda, F. and Craig, N.L. (2004) Transposition of hAT elements links transposable elements and V(D)J recombination. Nature 432, 995–1001, https://doi.org/10.1038/nature03157
51 Lemoine, F.J., Degtyareva, N.P., Lobachev, K. and Petes, T.D. (2005) Chromosomal translocations in yeast induced by low levels of DNA polymerase: a model for chromosome fragile sites. *Cell* **120**, 587–598, https://doi.org/10.1016/j.cell.2004.12.039

52 Koszul, R. and Fischer, G. (2009) A prominent role for segmental duplications in modeling eukaryotic genomes. *C. R. Biol.* **332**, 254–266, https://doi.org/10.1016/j.crvi.2008.07.005

53 Haas, N.B., Grabowski, J.M., Sivitz, A.B. and Burch, J.B. (1997) Chicken repeat 1 (CR1) elements, which define an ancient family of vertebrate non-LTR retrotransposons, contain two closely spaced open reading frames. *Gene* **197**, 305–309, https://doi.org/10.1016/S0378-1119(97)00276-X

54 Malik, H.S., Burke, W.D. and Eickbush, T.H. (1999) The age and evolution of non-LTR retrotransposable elements. *Mol. Biol. Evol.* **16**, 793–805, https://doi.org/10.1093/oxfordjournals.molbev.a026164

55 Drew, A.C. and Brindley, P.J. (1997) A retrotransposon of the non-long terminal repeat class from the human blood fluke *Schistosoma mansoni*. Similarities to the chicken-repeat-1-like elements of vertebrates. *Mol. Biol. Evol.* **14**, 602–610, https://doi.org/10.1093/molbev/msi054

56 Kajikawa, M., Ichiyanagi, K., Tanaka, N. and Okada, N. (2005) Isolation and Characterization of Active LINE and SINEs from the Eel. *Mol. Biol. Evol.* **22**, 673–682, https://doi.org/10.1093/molbev/msi054

57 Ikeda, N., Chijiwa, T., Matsubara, K., Oda-Ueda, N., Hattori, S., Matsuda, Y. et al. (2010) Unique structural characteristics and evolution of a cluster of venom phospholipase A2 isozyme genes of *Protobothrops flavoviridis* snake. *Gene* **461**, 15–25, https://doi.org/10.1016/j.gene.2010.04.001

58 Kajikawa, M, Ohshima, K and Okada, N (1997) Determination of the entire sequence of turtle CR1: the first open reading frame of the turtle CR1 element encodes a protein with a novel zinc finger motif. *Mol. Biol. Evol.* **14**, 1206–1217, https://doi.org/10.1093/oxfordjournals.molbev.a025730